CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 124, No. 2 August 2008

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

(Continued)

CONTENTS—*Continued from preceding page*

Harbor porpoise clicks do not have conditionally minimum time bandwidth product

Kristian Beedholm[a)]
Institute of Biology, University of Southern Denmark, Campusvej 55, DK 5230 Odense M, Denmark
beedholm@mail.dk

Abstract: The hypothesis that odontocete clicks have minimal time frequency product given their delay and center frequency values is tested by using an in-phase averaged porpoise click compared with a pure tone weighted with the same envelope. These signals have the same delay and the same center frequency values but the time bandwidth product of the artificial click is only 0.76 that of the original. Therefore signals with the same parameters exist that have a lower time bandwidth product. The observation that porpoise clicks are in fact minimum phase is confirmed for porpoise clicks and this property is argued to be incompatible with optimal reception, if auditory filters are also minimum phase.

PACS numbers: 43.80.Ka, 43.80.Lb [CM]
Date Received: December 20, 2007 **Date Accepted:** April 13, 2008

1. Introduction

Odontocete whales use echolocation for prey finding and navigation. The transmitter part of the biological sonar system of porpoises and dolphins emits a short pulse, a few tens to about hundred μs long (Au, 1993; Madsen *et al.*, 2004; Madsen *et al.*, 2005; Villadsgaard *et al.*, 2007). Wiersma (1988) notes that these animals' signals, given their very short durations, have a low bandwidth. This does not necessarily mean that they are all narrow band signals as such, but that there is a close coupling between the duration and bandwidth, so that the narrow band high frequency (NBHF) (Morisaka and Connor, 2007) signals, such as those employed by members of the porpoise family, have longer durations than the very broadband signals used by dolphins like Tursiops (Au, 1993). A low time-bandwidth product (TBP) is an advantage in signal detection, because a simple receiver can process it optimally, i.e., function as a matched filter, equivalent to a cross-correlation operation. A signal with low TBP occupies a small rectangular area in a time frequency representation. A filter with an impulse response that is contained within the same area might constitute a matched filter for that signal. The signals with the absolute minimum TBP value of 0.08 belong to a class of signals termed Gabors (Gabor, 1946; Venkatesh *et al.*, 2005) and consist of a Gaussian multiplied by a sinusoid. The notion that odontocete clicks might have a conditionally minimum TBP may have arisen because Gabors start in principle at time minus infinity and are therefore not practically realizable.

Wiersma (1988) used an iterative method to search in artificially generated signals for the one with the lowest TBP. The search was limited to signals with the same delay (see below) and frequency centroid values as real odontocete clicks. He found that the artificial signals that fit these criteria looked like the odontocete clicks from which the delay and frequency centroid values had been obtained. He therefore concluded that these animals have optimum waveforms under these constraints. Here, the TBP of a signal, given these constraining parameters of delay and frequency centroid, is termed the conditionally minimum TBP.

One problem with the analysis carried out by Wiersma (1988) is that the delay value is hard to determine in any noise at all. The frequency centroid is determined as the value that divides the linear amplitude spectrum in two equal halves. In that case the origin is simply zero

---

[a)]Current address: Zoophysiology, Department of Biological Sciences, University of Aarhus, Denmark.

Hz. The delay is analogously determined as the time delay value that divides the signal envelope in two equal halves, but where is time zero? For a recorded sound this is never obvious. Considering that this value is critical to the analysis, extrapolation seems imprudent. One may also wonder if the use of an optimum outgoing signal is important, since the returning echoes are influenced by the objects that reflected them and thus no longer optimal.

The reasoning given in the following describes a test of Wiersma's assertion of conditionally minimum TBP by investigating another artificial signal whose values of delay and frequency centroid are identical to the signal under scrutiny. If the artificial signal has a lower TBP than the real one, then the conditional minimum was not met in the real signal, which then serve as a falsification of the Wiersma's (1988) assertion.

Harbor porpoise signals were chosen because the spectrum of the signal envelope stays well below the frequency centroid. It is not completely impossible to carry out a somewhat similar analysis when this is not the case as were found for most other odontocete species tested, but it complicates things somewhat. It is also crucial to have a very high signal to noise ratio, since otherwise the effective envelope duration gets too high. Porpoises produce clicks that are so similar from one click to the next that it is possible to perform in-phase averaging over a number of consecutively emitted clicks to reduce the noise without noticeably affecting the waveform.

Below the results of this test are explored and it is argued that a minimum TBP signal is not necessarily optimal given the hearing system of the animals.

2. Analysis

We seek a signal with the same delay [*sensu* Wiersma (1988)] and frequency centroid as the porpoise click, $s(t)$. It is possible to describe the porpoise click signal as

$$s(t) = a(t)\cos(\psi(t)),$$

where $a(t)$ is the envelope and $\psi(t)$ is the instantaneous phase function. Regardless of the shape of $\psi(t)$, $a(t)$ will per definition have the same delay value as $s(t)$.

We can therefore construct a new signal, $\hat{s}(t)$, with the same frequency centroid as $s(t)$ by setting $\psi(t)=2\pi f_c t$, where f_c is the frequency centroid of $s(t)$ but keeping $a(t)$. Note that the spectrum of $\hat{s}(t)$ can be found as the convolution between the spectrum of the envelope $a(t)$ and a frequency shift operator, $\delta(f-f_c)$ moving the center of gravity of the envelope spectrum up to the frequency centroid, f_c, of $s(t)$. The amplitude spectrum of the envelope is symmetrical around zero, so if f_c is higher than the highest frequency in the envelope, then $\hat{s}(t)$ has the same frequency centroid as $s(t)$. If not, the spectrum of $\hat{s}(t)$ was not completely moved up to frequencies above the DC border and the resulting frequency centroid is then not necessarily equal to f_c, and the argumentation used here is then not completely valid (see below).

It is not given that $\hat{s}(t)$ has the minimum attainable TBP under the constraining parameters, but if it has a value that is lower than the TBP of $s(t)$ then the assertion that odontocetes use conditionally TBP signals is falsified, since the constructed signal $\hat{s}(t)$ has both the same delay and the same frequency centroid, which were the measured constraints under which TBP should be minimized.

3. Methods

The analysis was carried out with harbor porpoise (*Phocoena phocoena*) clicks recorded at the Fjord & Belt Center in Kerteminde, Denmark. Signals were sampled at 800 kHz in a star shaped array of four hydrophones (Schotten *et al.*, 2004) during a prey finding session with dead fish. The clicks from a sequence of 51 consecutively emitted clicks (Fig. 1) with similar amplitudes (contained within a range below 3 dB, STD=0.8 dB) were averaged in phase (Fig. 2). The selected signals were more powerful at the center hydrophone so they were assumed to have been recorded close to on axis (Rasmussen *et al.*, 2004).

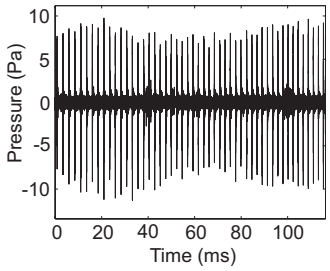


Fig. 1. The click sequence used in the analysis. This is part of a buzz sequence. The rate of click production is approx. 400 Hz.

In performing the averaging, the signals were aligned based simply on the position of the peak pressure value after interpolation to an effective sample rate of 6.4 MHz to improve the quality of the time alignment. The averaged signal was windowed as shown in Fig. 2(a).

4. Results

It is seen from Fig. 2(a) that the aligned signals are remarkably similar. This is confirmed by the relatively small variation in the shape of the amplitude spectra, shown in Fig. 2(b). The thick black line in Fig. 2(b) shows the spectrum of the signal resulting from the averaging in the time domain. Had the alignment been unjustified in that the signals were not similar in every detail, including phase, the result would have been a low-pass filtering effect relative to the spectra of the raw, not-averaged clicks. The spectra are all very similar to the spectrum of the averaged signal in the range from 100 to ~180 kHz. Only at the frequencies outside this range, where porpoise clicks are without appreciable energy, does the spectrum of the averaged signal differ from the spectra of the clicks making it up. The in-phase averaging of the time domain signals would therefore seem to be justified.

The averaged signal has a TBP of 0.33 and an f_c of 136.9 kHz. Multiplying the envelope of the averaged signal with a sinusoid with the f_c value results in a new signal with a TBP of 0.26, which is 1.35 times lower than the original signal. The sinusoid-envelope product, with the phase that gave the best match with the averaged signal, is shown in green behind the average signal in Fig. 2(a). The correlation coefficient between these signals was 0.83. The spectrum of the artificial signal is shown together with the spectrum of the average signal in Fig. 2(b), and is clearly considerably narrower band than are the real signals.

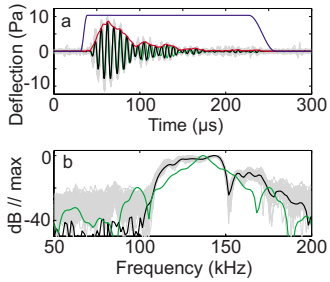


Fig. 2. (Color online) Time aligned echolocation clicks and average click. (a) The *gray lines* are the individual clicks, whereas the *black line* is the time aligned average. *Red trace* immediately above and following the signal is the envelope and the *blue line* shows the apodization function used to isolate the average click for analysis (here, the windowing function is scaled to the peak pressure value of the clicks for better visualization). In *green*, behind the average signal is shown the artificial click made by multiplying the envelope by sinusoid with the best matching phase. (b) *Gray lines* are normalized amplitude spectra of individual clicks. The *black trace* is the spectrum of the averaged click. *Green line* is the normalized spectrum of the artificial click.

The individual signals in the sequence all have higher TBPs than the artificial signal and that result is highly significant (t-test, DF=49, $P < 10^{-24}$). It is, however, not a fair test, since the superimposed noise, which was mostly gotten rid of by in-phase averaging, invariably lifts both the duration and the bandwidth to higher values. A more conservative test should be between the TBP values of the averaged signal and the artificial signal. But to make that comparison we need a measure of the variance. If, for the sake of argument, we assume that the variation in TBP of the individual clicks is due to a dominating influence on the observed variation of the underlying (unknown) noise free signals, we certainly do not underestimate that underlying variation, since the noise is bound to have some influence and that influence will result in a variation that is larger than what would have been observed, had there been no noise. Therefore, by adopting the variance figure from the population of individual TBPs and using that as the variance estimate parameter in the t-test of whether the TBP of the averaged signal is indeed smaller than the TBP obtained from the artificial signal, we produce a highly conservative test. The result of this test is that the difference is indeed also significant (t-test, DF=49, $p < 0.005$).

Several controls were made using other window lengths for the isolation of the average click, other click sequences, and single clicks. On the basis of this exploration of different analysis parameters the measured discrepancy is believed to be conservatively estimated.

The TBP of the averaged signal is four times higher than the theoretical minimum value of 0.08 held by the Gabor functions mentioned above.

5. Discussion

For harbor porpoises it seems that the hypothesis of conditional minimum TBP has therefore been falsified. The relatively high TBP value also shows that the signal is quite different from a Gabor signal.

It might well be argued that a ratio of 1.3 between model and observation is not such a bad fit, but the Wiersma study (1988) implicitly claims to explain even the absolute phase of the clicks in that the clicks portrayed have the same phase as the clicks from which the constraining values of bandwidth and delay came from. Varying the phase in the case of a porpoise click does not change the TBP, and since the claim is based mostly in these visual similarities between waveforms, that result should be cited with some caution.

It should be noted here that using waveforms from other odontocetes, similar results are found, but that in this case the argumentation above is not completely valid, since their bandwidths are above their frequency centroid, which causes the frequency centroid of the artificial signal, $\hat{s}(t)$, to be slightly different from f_c when the envelope is multiplied by $\cos(2\pi f_c t)$. It becomes dependent on the phase. Since it is difficult to imagine that a mammalian hearing system can optimally receiving a signal as broadband as the clicks from, e.g., Tursiops, it is also for that reason more relevant to consider Wiersma's (1988) hypothesis for NBHF sounds alone.

The porpoise clicks do have a low TBP, regardless of the fact that they are by no means optimal in this respect. As mentioned in the introduction, by having a low TBP one achieves that the expected signal in question is contained within a small area in a time-frequency representation. But this is only desirable if the receiver is tuned to detect signals in that same small area. Other receivers require other signals for optimal reception. Since it has been shown here that the signals do not have conditionally minimum TBP and since the receiving system is probably not tuned to receive minimum TBP signals optimally, might not the hearing system be tuned to detect the actual odontocete clicks optimally?

Again, this question is especially relevant to porpoises, where the NBHF echolocation signals might well be imagined to have evolved to be a match for the transfer function of a mammalian auditory filter. There is a single abstract (Olivieri, 2002) that reports odontocete clicks to be minimum phase (MP), and that result is confirmed here for porpoise clicks (Fig. 3). This finding is relevant in this context, because unless the signals are Gabor functions (see above) then MP signals have phase spectra that are different from the straight line that would constitute a simple delay (Biering and Pedersen, 1983). Optimal reception can be achieved only when the amplitude spectrum of the expected signal is identical to that of the transfer function of the receiving filter, and when the derivative of the phase spectrum (group delay) of the re-

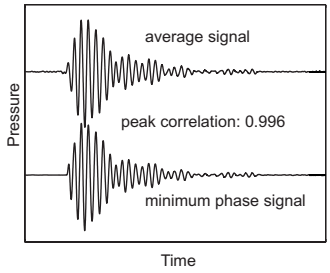


Fig. 3. Comparison between the average porpoise echolocation click and the minimum phase signal that has the same amplitude spectrum as the average click.

ceiving filter is the negation of the same quantity of the expected signal. It follows then that the matched receiving filter cannot be MP, because this would mean that the phase spectrum of both the receiver and the signal would have to be minimum, which again implies a Gabor function (Gabor, 1946; Venkatesh *et al.*, 2005), which is clearly not the case given our measured TBP value. But auditory filters are in fact also reported to be minimum phase (de Boer, 1997) and if that somewhat contended (Recio *et al.*, 1997) result holds, the echolocation clicks then cannot be matched perfectly to the impulse response of auditory filters.

It appears then that the minimum phase property in this case is merely a natural quality of a certain ubiquitous class of transient signals and that it does not represent any optimization in itself. But again, that the match between signal and receiver cannot be optimal does not mean that it is not good enough to work quite well, which the successful use of biosonar and continued existence of the animals proves.

In conclusion it may be restated that at least not all odontocetes appear to have signals that are optimized with respect to the minimal TBP under the constraining parameters of frequency centroid and signal envelope, and therefore signal delay. Consideration of MP properties of both the echolocation signals and typical auditory filters suggests that odontocete clicks in this respect do not have receivers matched to the expected signals.

Acknowledgments

The harbor porpoise is maintained by Fjord & Bælt, Kerteminde, Denmark, under permit No. J.nr. SN 343/FY-0014 and 1996-3446-0021 from the Danish Forest and Nature Agency, Danish Ministry of Environment. The author is indebted to Bertel Møhl, Magnus Wahlberg, Peter T. Madsen, Lee Miller, and two anonymous reviewers for valuable discussions. This work was funded by Carlsberg Research Foundation.

References and links

Au, W. W. L. (**1993**). *Sonar of Dolphins* (Springer-Verlag, New York).
Biering, H., and Pedersen, O. Z. (**1983**). "System analysis and time delay spectrometry," Brüel and Kjær Tech. Rev. **32**, 3–52.
de Boer, E. (**1997**). "Cochlear models and minimum phase," J. Acoust. Soc. Am. **102**, 3810–3813.
Gabor, D. (**1946**). "Theory of communication," J. Inst. Electr. Eng., Part 3 **93**, 429–457.
Madsen, P. T., Kerr, I., and Payne, R. S. (**2004**). "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales Pseudorca crassidens and Risso's dolphins *Grampus griseus*," J. Exp. Biol. **207**, 1811–1823.
Madsen, P. T., Carder, D. A., Bedholm, K., and Ridgway, S. H. (**2005**). "Porpoise clicks from a sperm whale nose—convergent evolution of 130 kHz pulses in toothed whale sonars?" Bioacoustics **15**, 195–206.
Morisaka, T., and Connor, R. C. (**2007**). "Predation by killer whales (*Orcinus orca*) and the evolution of whistle loss and narrow-band high frequency clicks in odontocetes," J. Evol. Biol. **20**, 1439–1458.
Olivieri, M. P. (**2002**). "What can be learned from one of nature's most advanced biosonars: Discussion on Bottlenose dolphins echolocation waveforms with respect to echolocation tasks in shallow water," J. Acoust. Soc. Am. **111**, 2371.
Rasmussen, M. R., Wahlberg, M., and Miller, L. A. M. (**2004**). "Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*)," J. Acoust. Soc. Am. **15**, 1826–1831.
Recio, A., Narayan, S. S., and Ruggero, M. A. (**1997**). "Wiener-kernel analysis of basilar-membrane responses to

white noise," in *Diversity in Auditory Mechanisms*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325–331.
Schotten, M., Au, W. W. L., Lammers, M. O., and Aubauer, R. (**2004**). "Echolocation recordings and localizations of wild spinner dolphins *Stenella longirostris* and pantropical spotted dolphins *Stenella attenuate* using a four hydrophone array," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 393–400.
Venkatesh, Y. V., Raja, S. K. and Sagar, G. V. (**2005**). "On band limited signals with minimal product of effective spatial and spectral widths," Int. J. Math. Math. Sci. **10**, 1589–1599.
Villadsgaard, A., Wahlberg, M., and Tougaard, J. (**2007**). "Echolocation signals of wild harbour porpoises, *Phocoena phocoena*," J. Exp. Biol. **210**, 56–64.
Wiersma, H. (**1988**). "The short-time-duration narrow-bandwidth character of odontocete echolocation signals." in *Animal Sonar*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 129–145.

Effects of modulation wave shape on modulation frequency discrimination with electrical hearing

David M. Landsberger
Department of Auditory Implants and Perception, House Ear Institute, 2100 West 3rd Street, Los Angeles, California 90057,
Department of Otolaryngology, The University of Melbourne, 384-388 Albert St. East Melbourne 3002 Australia, and School of Life and Health Sciences,
Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom
dlandsberger@hei.org

Abstract: Amplitude modulations of pulsitile stimulation can be used to convey pitch information to cochlear implant users. One variable in designing cochlear implant speech processors is the choice of modulation waveform used to convey pitch information. Modulation frequency discrimination thresholds were measured for 100 Hz modulations with four waveforms (sine, sawtooth, a sharpened sawtooth, and square). Just-noticeable differences (JNDs) were similar for all but the square waveform, which often produced larger JNDs. The results suggest that a sine, sawtooth, and sharpened sawtooth waveforms are likely to provide similar pitch discrimination within a speech processing strategy.

PACS numbers: 43.66.Ts, 43.66.Fe, 43.66.Hg, 43.66.Nm [QJF]
Date Received: February 1, 2008 **Date Accepted:** April 17, 2008

1. Introduction

Cochlear implantees using commercial speech processing strategies are often able to understand speech well enough to conduct conversations in quiet situations. These high levels of performance are obtained even though most clinical speech processing strategies do not provide clear fundamental frequency (F0) information. For both CIS (Continuously Interleaved Sampling; Wilson *et al.*, 1991) and ACE (Advanced Combination Encoder; Vandali *et al.*, 2000) processing strategies, F0 cues are primarily provided by amplitude modulations within a channel, which can modulate at F0 if the corresponding filter bandwidth is sufficiently broad (Shamma and Klevin, 2000). It has been shown (Shannon *et al.*, 1995; Faulkner *et al.*, 2000) that F0 is not required for recognition of vowels or consonants, which may partly explain the relatively high speech understanding in quiet despite the limited F0 coding in these strategies.

Implant patients have difficulty with tasks that are facilitated by F0 information. Patients have difficulty discriminating between questions and statements (Green *et al.*, 2005), identifying a speaker (Fu *et al.*, 2004), recognizing tones in tonal languages (Fu *et al.*, 2004), and understanding a speaker in the presence of a speech masker (Stickney *et al.*, 2004). To improve performance in F0 related tasks, novel speech processing strategies have been developed to specifically encode F0 in the modulations of the outputs of the electrodes. Geurts and Wouters (2001) created a speech processing strategy (F0 CIS) in which modulation depths were increased at F0. However, F0 discrimination for synthesized vowels was similar for standard CIS and F0 CIS. Laneau *et al.* (2006) tested a strategy (F0mod) that was a modification of ACE in which the envelopes of each channel were modulated by F0 at 100% modulation depth. Performance with F0mod was better than that with ACE for familiar melody recognition and musical note discrimination when F0 was below 250 Hz. However, F0mod was not tested on speech. Green *et al.* (2005) implemented a strategy where outputs of electrodes were modulated at F0 using a sharpened sawtooth (see illustration in Fig. 1). Compared to standard CIS, listeners were better able to discriminate between rising and falling pitch, and between questions and

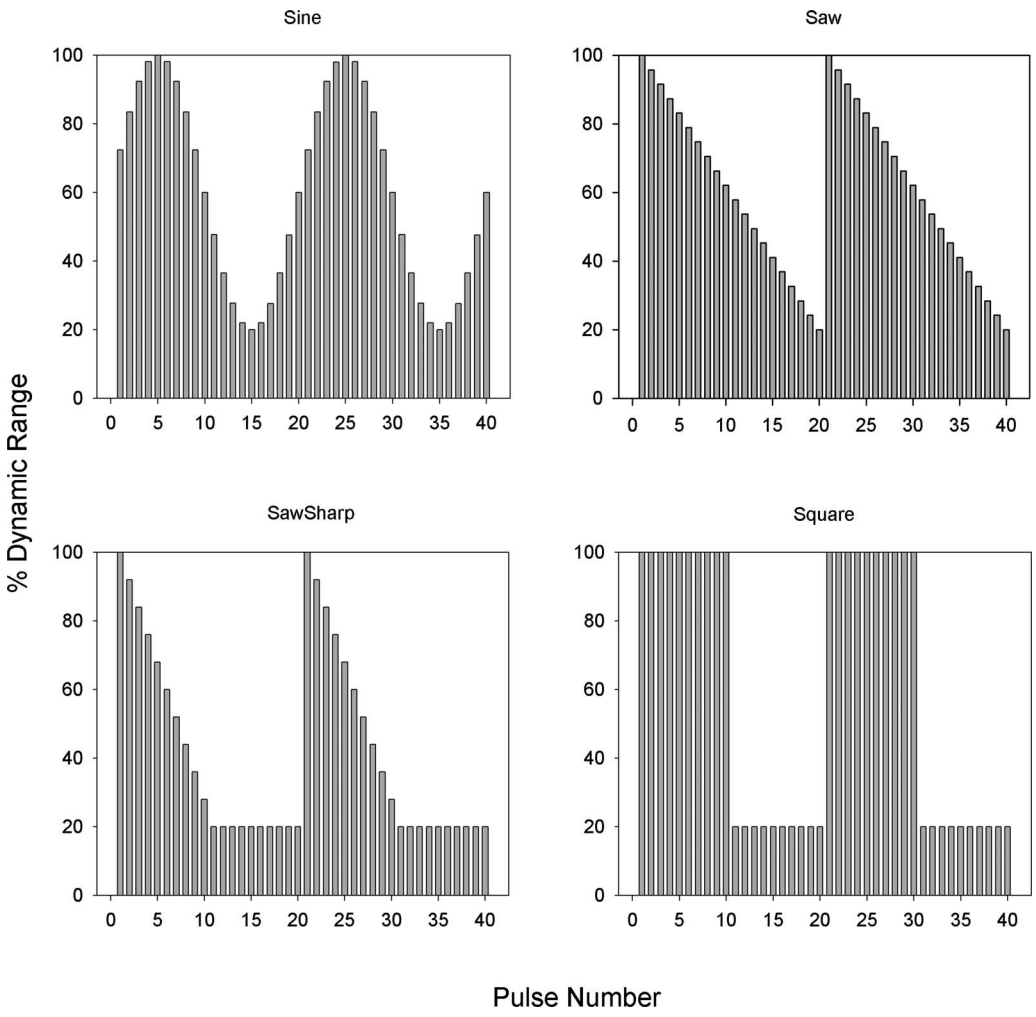


Fig. 1. Illustration of the four experimental modulation waveforms.

statements with the experimental strategy; however, vowel recognition was significantly poorer with the experimental strategy. These results are consistent with Chatterjee and Peng (2008) who showed that modulation frequency discrimination on a single electrode was strongly correlated with speech intonation. Vandali *et al.* (2005) tested several speech processing strategies (both experimental and commercial) and found that the experimental strategies that modulated outputs of the electrodes with F0 information enabled better pitch discrimination in a sung vowel task compared to performance with ACE. However, with the exception of Multi-Channel Envelope Modulation (MEM), all of the experimental strategies resulted in speech comprehension that was poorer than that with ACE. Only the MEM strategy, which modulated the outputs of all electrodes by the envelope of the input signal (and thereby had inherent modulations at F0) provided an improvement in the sung vowel pitch task without detriment to speech performance. Although commercial speech processing strategies do not provide optimal F0 information via channel modulations, the limited modulation pitch information is useful. For example, Laneau *et al.* (2004) showed that removing amplitude modulations above 10 Hz in a speech processing strategy significantly reduced F0 discrimination.

Each of these experimental strategies was developed to improve F0 pitch perception by modulating the current output from the electrodes. However, when designing a speech pro-

cessing strategy, there are many parameters to evaluate. Some of these parameters include which (or how many) electrodes to modulate, what is the appropriate modulation depth, and what waveform should be used for the modulations. Geurts and Wouters (2001) showed that presenting amplitude modulations on multiple electrodes improved listeners' ability to detect and discriminate modulations. Assumptions about the appropriate waveform have been made, but not adequately tested. Green *et al.* (2004) compared the ability of cochlear implant subjects to perform a pitch glide discrimination with either standard CIS, or modified CIS strategies which provided modulations on all electrodes at F0. The F0 modulations were presented either as sawtooth or sharpened sawtooth waveforms. While the modified strategies improved pitch glide discrimination relative to standard CIS, no differences in speech performance were detected between the sawtooth and sharpened sawtooth waveforms. Further experiments with cochlear implantees (Green *et al.*, 2004, 2005) used only the sawtooth waveform for modulating F0.

In the current study, just-noticeable differences (JNDs) for modulation frequency discrimination were measured for different modulation waveforms, namely: sine, sawtooth (saw), sharpened sawtooth (sawsharp), and square waveforms (see Fig. 1). Most previous studies regarding amplitude modulation JNDs with cochlear implants have used sinusoidal modulation (e.g., Geurts and Wouters, 2001; Chatterjee and Peng, 2008); thus, the data from the present study can be readily compared to that from previous studies. Naturally occurring sounds modulate sinusoidally; sinusoidal modulation therefore may be the optimal waveform for cochlear stimulation, and may produce the smallest modulation frequency JNDs. However, if modulation frequency pitch is determined by the interpulse interval in the neural firing pattern, then the sharp onset of the sawtooth and sharpened sawtooth may produce the smallest JNDs. Alternatively, if modulation frequency pitch is determined by neurons that detect the transitions between the onset and offset of a stimulus, the square waveform may produce the smallest JNDs.

2. Methods

2.1 Subjects

Eight subjects with the Nucleus CI24 implant participated in the present study. Subjects used either the SPEAK (Seligman and McDermott, 1995; Whitford *et al.*, 1995) or ACE strategy in their clinical speech processors. The data for five of the subjects was collected at the University of Melbourne in Australia, while the data for the remaining subjects were collected at Aston University in England. The Australian subjects had all participated in previous psychophysical experiments while the English subjects had no previous research experience.

2.2 Stimuli

All stimuli consisted of monopolar (MP1+2) biphasic pulse trains presented on a single electrode using a SPEAR speech processor (HearWorks Ptv. Ltd., 2003) that was controlled by custom-written software. Stimulation consisted of pulses presented to electrodes 20, 12, or 6, which are located in the apical, medial, and basal regions of the cochlea, respectively. Stimuli had a pulse phase duration of 26 μs, an interphase gap of 8.4 μs, and a duration of 500 ms.

The stimuli contained modulations that were either sinusoidal (sine), sawtooth (saw), a sharpened sawtooth (sawsharp), or square (square) waveforms. The modulation depth was fixed at 80% of the dynamic range (DR). The peak of the modulation corresponded to 100% DR and minimum of the modulation corresponded to 20% DR. The carrier rate of the modulated stimuli was fixed at 20 times the modulation rate. As a result, every modulation cycle was encoded with exactly 20 samples. This ensured that, for the saw and sawsharp waveforms, the sharp onset of the waveform was sampled accurately every cycle. In a pilot study, subjects were easily able to discriminate between 100 and 100.5 Hz modulated stimuli (saw waveform) when the carrier was fixed at 2000 pps, most likely because of aliasing artifacts present in the 100.5 Hz stimulus. Thus, in the present study, 100 Hz modulation was coded with a 2000 pps carrier and 100.5 Hz modulation was coded with a 2010 pps carrier.

2.3 Procedure

On each experimental electrode, the DRs were estimated for the 2000 pps (unmodulated) carrier. To measure the loudest comfortable level (*C* level), subjects listened to the stimulus at increasing amplitudes until they reported that the sound was at the loudest comfortable level. Threshold levels (*T* levels) were measured using a one-up, one-down 4 interval forced choice (IFC) task. This procedure was repeated until ten reversals were obtained. The last six reversals were averaged to estimate the 50% detection threshold (Levitt, 1971). The DR was defined as the difference between *C* and *T* levels, in Nucleus current levels.

The loudness of two stimuli were balanced using a one-up, one-down, 2IFC task. One of the stimuli had a fixed amplitude and the other stimulus was adjusted according to subject response by 1 Nucleus current step (0.18 dB). Ten reversals were measured, and the last six reversals were averaged to estimate the level of equal loudness. The loudness balancing procedure was then repeated, with the previous reference as the probe and the previous probe as the reference. The loudness balancing procedure was repeated twice and all of the loudness estimates were averaged. The loudness balance procedure was the same used for subjects MM, BK, and DC in a previous study (Landsberger and McKay, 2005). For each electrode, 100 Hz modulated stimuli with a 2000 pps carrier were loudness balanced to 100 pps unmodulated stimuli at *C* level.

For each of the four waveforms, a 4IFC adaptive procedure (one-up, one-down) was used to measure the modulation frequency JND at each of the three electrode locations. During the procedure, three intervals contained stimuli that were modulated at 100 Hz (the reference stimuli) and one contained a stimulus that was modulated at a frequency above 100 Hz (the target stimulus). To prevent the use of loudness differences as a cue for discrimination, the amplitudes of the target and the three reference stimuli were jittered up or down by a maximum of 4 Nucleus current levels (0.7 dB). Subjects were instructed to select the interval that was different from the others in any way other than loudness. The modulation frequency of the target stimulus was adjusted according to subject response by 10%. A total of ten reversals were measured, and the last six reversals were averaged to estimate the JND corresponding to 50% discrimination. Within each experimental block, the JND was measured once for all four waveforms and three electrodes. The order of stimulus presentation was randomized for each block. A total of five blocks were tested per subject.

To ensure that the ±4 Nucleus current level (0.7 dB) jitter was sufficient to mask any loudness cues generated by the change in carrier rate, the loudness of each of the reference stimuli was balanced to a target stimulus modulated at the frequency minimally discriminable from 100 Hz (i.e., the JND). If an adjustment of more than 1 Nucleus current level (0.18 dB) was required to make the two stimuli equally loud, the modulation frequency JNDs were remeasured with the level-adjusted target stimulus.

A test was performed to verify that the detected changes for each stimulus were not based on the difference in carrier rate. Subjects were asked to discriminate an unmodulated pulse train at the reference stimulus carrier rate (2000 pps) from an unmodulated pulse train at the carrier rate of the target stimulus representing the largest measured difference limen on the same electrode, regardless of waveform. Thus, if the maximum modulation frequency JND was 10%, carrier rate discrimination was measured between 2000 and 2200 pps stimuli. Current levels were randomly jittered by ±4 Nucleus current levels (0.7 dB). Subjects were asked to choose which sound was different, ignoring differences in loudness. Carrier rate discrimination was measured using a 4IFC task, and each comparison was repeated ten times for each electrode. While it was unlikely that there were preceptual differences between the carrier rates (Landsberger and McKay, 2005), it was important to verify this assumption to reduce concerns of carrier rate as a confounding variable.

3. Results

Figure 2 displays the modulation frequency JNDs for each of the four waveforms. There was no obvious advantage observed for any of the experimental waveforms, as performance across the

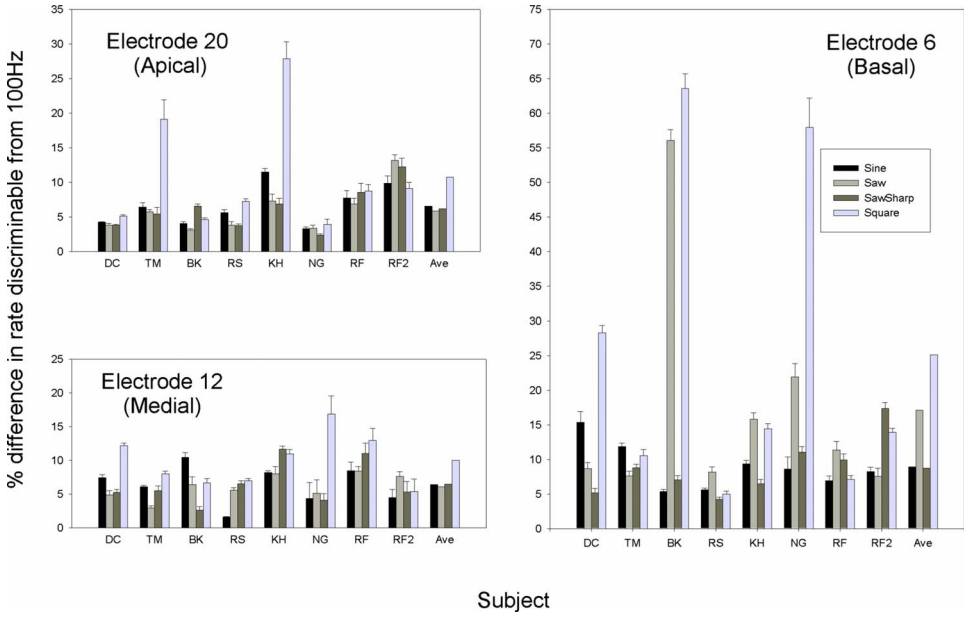


Fig. 2. (Color online) Modulation frequency JNDs (in percent difference re: 100 Hz) for individual subjects. The three panels show data for the three experimental electrode locations. The error bars show one standard error of the mean.

waveforms was relatively similar. The size of the JNDs was generally between 5 and 10%, consistent with previous sinusoidal modulation frequency discrimination data in cochlear implant users (Geurts and Wouters, 2001; Chatterjee and Peng, 2008). The optimal waveform for individual subjects often varied with the electrode. For example, for subject BK, the smallest JNDs were produced by the sawtooth on electrode 20, the sawsharp on electrode 12, and the sine on electrode 6. In general, the square waveform produced JNDs similar to those with the sine, saw, and sawsharp waveforms; about half of the JNDs with the square waveform were below 10%. However, for most subjects, on at least one of the electrodes, the JNDs measured with the square wave were much larger than those with the other waveforms; 25% of the JNDs with the square waveform were above 20%. These outlier JNDs with the square waveform varied inconsistently across subjects, in terms of electrode location. Note the outlier for subject BK with the saw waveform on electrode 6.

A two-way repeated measures analysis of variance found significant effects for both electrode location [$F(2,42)=4.025$, $p=0.042$; power of analysis: 0.486) and modulation waveform [$F(3,42)=5.862$, $p=0.005$; power of analysis: 0.849). However, no significant interaction was observed between electrode location and waveform [$F(6,42)=1.613$, $p=0.168$; power of analysis: 0.214). Presumably, the main effect for waveform was due to the sometimes poorer and more variable performance with the square wave. A pairwise *post-hoc* test (Holm-Sidak) was performed to compare all waveforms. Significant differences were found between the square and the sine waveforms, and between the square and the sawsharp waveforms.

A binomial test was unable to detect a perceptual difference in the carrier discrimination task. Subjects were able to identify the different carrier rate between two and four out of ten times. Given a chance level of 0.25, a subject would have to identify the different carrier rate six or more times to obtain a significantly different result ($\alpha = 0.05$).

4. Discussion

In the present study, modulation frequency JNDs were similar for all experimental waveforms. The size of the JNDs was similar to that found in previous cochlear implant studies in which

sinusoidal modulation was presented to a single electrode (Geurts and Wouters, 2001; Chatterjee and Peng, 2008), or presented to multiple electrodes within a speech processing strategy (Geurts and Wouters, 2001). The data are consistent with the finding of Green *et al.* (2004) that glide discrimination was not significantly different for sawtooth and sharpened sawtooth waveforms. Note that in the present study, the minimum of the amplitude modulations were presented at 20% DR, and not at threshold as in the Green study.

In most cases, the size of the JNDs measured with the square waveform was similar to the JNDs measured with the other waveforms. However, there were a number of outliers with the square waveform across subjects and electrodes. It is unclear why performance with the square waveform was more prone to outliers than were the other waveforms. The outliers were unlikely to be a result of noisy data because the variability for the outlying conditions was relatively small. The outliers also did not seem to follow an obvious pattern, as they were observed for most subjects, but at different electrode locations. Thus, it is unlikely that the outliers were due to individual subject differences or differences in electrode location. Fu (2002) has shown that modulation detection thresholds are strongly correlated with speech recognition in cochlear implants. Pfingst *et al.* (2007) have argued that stimulating at sites within the cochlea which are poor at modulation detection might actually make speech perceptual tasks more difficult. It is possible that, within a speech processing strategy, modulating with a square wave may reduce speech comprehension.

The data suggest that when designing a speech processing strategy in which modulations are used to convey F0, a sine, saw, and sawsharp waveform are interchangeable. Note that in the present study, modulation frequency JNDs were only measured for 100 Hz stimuli with 80% modulation depth. It is unclear whether modulation frequency JNDs may differ among the waveforms at other base frequencies or modulation depths. The choice in waveform may be driven by concerns other than frequency discrimination. For example, it is unclear whether the waveform shapes may interact differently when presented on multiple electrodes within a speech processing strategy. Speech comprehension has often been reduced when modifying a speech processing strategy to include modulations at F0 across electrodes (Geurts and Wouters, 2001; Green *et al.*, 2005; Vandali *et al.*, 2005). It is possible that different modulation waveforms may change the magnitude of effect on speech comprehension. If there is no perceptual difference between the waveforms, the choice of waveform may depend on engineering issues. Using a waveform with a sharp attack would require additional algorithmic steps to ensure that there are no aliasing artifacts. Additionally, the amount of current required to achieve a fixed loudness would likely be higher for narrower waveforms (e.g., sawsharp) than for sine waveforms. As such, the choice of waveform might affect battery life in the speech processor. These engineering issues aside, the present data suggests that sine, saw, or sharpened waveforms may provide similar benefits for speech processing strategies that encode pitch via amplitude modulations.

Acknowledgments

The author would like to thank the subjects for giving their time and effort to this project. Additional gratitude is extended to John Galvin and three anonymous reviewers for their helpful comments. Funding for this project was provided by an NIDCD fellowship.

References and links

Chatterjee, M., and Peng, S. C. (**2008**). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," Hear. Res. **235**, 143–156.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," J. Acoust. Soc. Am. **108**, 1877–1887.

Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.

Fu, Q. J., Chinchilla, S., Nogaki, G., and Galvin, J. J., 3rd (**2005**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.

Fu, Q. J., Hsu, C. J., and Horng, M. J. (**2004**). "Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant users," Ear Hear. **25**, 501–508.

Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.
Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.
Green, T., Faulkner, A., Rosen, S., and Macherey, O. (**2005**). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," J. Acoust. Soc. Am. **118**, 375–385.
HearWorks Pty. Ltd. (**2003**). "Spear 3 Research System" http://www.hearworks.com.au/spear, Last viewed 6/20/08.
Landsberger, D. M., and McKay, C. M. (**2005**). "Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implantees," J. Acoust. Soc. Am. **117**, 319–327.
Laneau, J., Wouters, J., and Moonen, M. (**2004**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **116**, 3606–3619.
Laneau, J., Wouters, J., and Moonen, M. (**2006**). "Improved music perception with explicit pitch coding in cochlear implants," Audiol. Neuro-Otol. **11**, 38–52.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Pfingst, B. E., Xu, L., and Thompson, C. S. (**2007**). "Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants," J. Acoust. Soc. Am. **121**, 2236–2246.
Seligman, P., and McDermott, H. (**1995**). "Architecture of the Spectra 22 speech processor," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 139–141.
Shamma, S., and Klein, D. (**2000**). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," J. Acoust. Soc. Am. **107**, 2631–2644.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
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.
Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., and McDermott, H. J. (**2005**). "Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies," J. Acoust. Soc. Am. **117**(5), 3126–3138.
Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., and Clark, G. M. (**1995**). "Evaluation of the Nucleus Spectra 22 processor and new speech processing strategy (SPEAK) in postlinguistically deafened adults," Acta Oto-Laryngol. **115**, 629–637.
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.

Quantifying the through-thickness asymmetry of sound absorbing porous materials

Y. Salissou and R. Panneton
GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada
yacoubou.salissou@usherbrooke.ca, raymond.panneton@usherbrooke.ca

Abstract: A method to quantify the through-thickness asymmetry of a sound absorbing porous material is proposed and discussed. Its calculation only requires impedance tube measurements of the acoustical surface impedance performed on both sides of the tested material. The method may be used for quality control or to assess the level of asymmetry of the material in terms of its acoustic properties. As a first validation, a two-layered porous system seen as an equivalent asymmetrical single porous layer with a sudden change in its physical properties is studied. From this study, a criterion of asymmetry is suggested and experimentally tested.

PACS numbers: 43.20.Jr, 43.20.Hq, 43.55.Ev, 43.58.Bh [AN]
Date Received: February 25, 2008 **Date Accepted:** April 17, 2008

1. Introduction

Sound absorbing porous materials are widely used in noise-control applications. When the material behaves as an equivalent fluid (i.e., rigid motionless frame or limp frame),[1,2] the acoustic behavior of the material is completely defined by two intrinsic dynamic properties: its equivalent acoustical characteristic impedance ($\tilde{Z}_{eq}$) and complex wave number ($\tilde{k}$), or alternatively by its equivalent dynamic density ($\tilde{\rho}_{eq}=\tilde{Z}_{eq}\tilde{k}/\omega$) and bulk modulus ($\tilde{K}_{eq}=\tilde{Z}_{eq}\omega/\tilde{k}$), where ω is the angular frequency. Acoustical impedance tube methods have been developed for measuring these dynamic properties. The traditional standing wave method[3] was the first proposed. Nowadays, the two-cavity method by Utsuno *et al.*,[4] the three-microphone method by Iwase *et al.*,[5] and the transfer matrix method by Song and Bolton[6] are commonly used. All the aforementioned methods, used to retrieve the intrinsic dynamic properties, assume the porous material to be a single layer with symmetrical acoustic properties. However, common porous materials (e.g., foams, fibrous media, and soils) are usually subjected to some variations in their microstructure during manufacturing. This may lead to macroscopic asymmetrical behaviors.[7] Unfortunately, there is no mean of simply quantifying the level of asymmetry (or symmetry) of a single porous layer.

In this paper, a method to quantify the through-thickness asymmetry of a sound absorbing porous material is proposed and discussed. The method only requires impedance tube measurements (ASTM E1050, ISO 10534) of the acoustical surface impedance performed on both sides of the tested material.

2. Why quantifying asymmetry?

Let P be a given bulk property (e.g., absorption coefficient, acoustical surface impedance, acoustical characteristic impedance, or complex wave number) measured with an impedance tube. For the two-layered porous material shown in Fig. 1, let P_{AB} be the value of P when the impedance tube measurement is performed when side A of the material is facing the incident sound wave (normal configuration), and P_{BA} the value of P when side B is now facing the incident sound wave (inverted configuration). For a perfectly symmetrical material, the measured properties do not vary with the side of the material facing the incident sound wave; therefore P_{AB} is equal to P_{BA}. In practice, sound absorbing materials are generally not perfectly symmetrical. Consequently, P_{AB} is not strictly equal to P_{BA}. The two questions arising from this

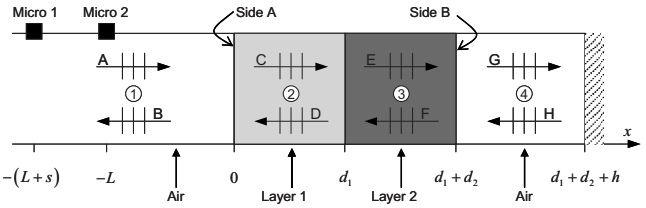


Fig. 1. A schematic view of the impedance tube configuration with the two-layered porous material backed by a plenum of air and a hard termination.

are (1) *where does the symmetry stop*, and (2) *when do the aforementioned impedance tube characterization methods apply?*

3. Index of asymmetry

With a view to answer the previous questions, one needs first to quantify the asymmetry of a material. For each angular frequency, this asymmetry can be quantified by the relative difference between P_{AB} and P_{BA} expressed as

$$RD(\omega) = \frac{|P_{AB}(\omega) - P_{BA}(\omega)|}{\max(|P_{AB}(\omega)|, |P_{BA}(\omega)|)}. \tag{1}$$

From Eq. (1), the average relative difference (ARD) of the bulk property P is defined as

$$ARD = \frac{1}{n}\sum_{i=1}^{n} RD(\omega_i), \tag{2}$$

where n is the number of discrete frequencies in the considered frequency range.

To verify if RD is a good basis for quantifying the asymmetry of a porous material, let us consider a porous material having oblique circular cylindrical pores. Let us assume there are some irregularities in the fabrication process leading to a slight variation in the inclination and radius of the pores through the thickness of the material. To simplify the complexity of the problem, this material will be viewed as a two-layered porous system. The first layer is made up from identical oblique circular cylindrical pores of radius $r=0.1$ mm. Their inclination with respect to the surface normal is $\theta=10°$. The geometrical properties of the second layer are $\theta' = (1+x)\theta$ and $r' = (1+x)r$, where x is their relative variation compared to the first layer. Each layer has a thickness $H=12.5$ mm, and a surface pore density of $N=30\times10^6$ pores/m^2.

Simulated measurements of the normal incidence impedance tube problem shown in Fig. 1 are conducted to obtain the sound absorption coefficient (α), acoustical surface impedance ($\tilde{Z}_s$), equivalent acoustical characteristic impedance ($\tilde{Z}_{eq}$), and complex wave number ($\tilde{k}$) on the two-layered porous system for different values of x. The simulations are made for both normal and inverted configurations and over the frequency range (300–4000 Hz). Also, the effect of random noise is included in the simulations in order to better reflect real measurements. The simulation method is explained in the Appendix. It is worth mentioning that the two-cavity method[4] is the method simulated to obtain ($\tilde{Z}_{eq}$) and ($\tilde{k}$). Also, the porous material is modeled using the model described in Ref. [8].

Figure 2 shows the ARD values for α, $\tilde{Z}_s$, $\tilde{Z}_{eq}$, and $\tilde{k}$ as a function of the variation x. This result indicates that ARD (hence, RD) is consistent with the composition of each sample. Indeed, the ARD increases as the change in the geometrical properties increases. Also, it indicates that the complex wavenumber and sound absorption coefficient are less sensitive to the asymmetry of the material compared to the acoustical characteristic impedance and acoustical surface impedance. Therefore, ARD (or RD) of $\tilde{Z}_{eq}$ and ARD (or RD) of $\tilde{Z}_s$ are both good candidates for studying the asymmetry of the material.

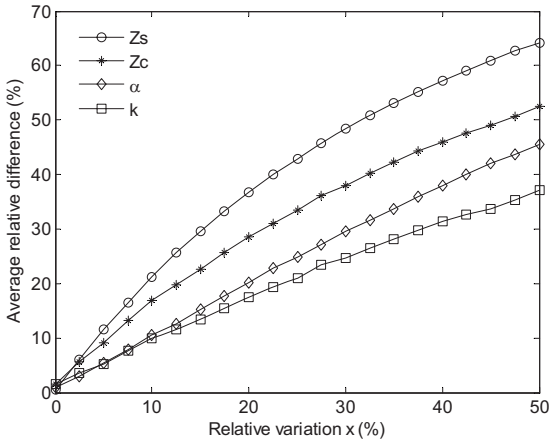


Fig. 2. Average relative difference (ARD) of the acoustical surface impedance, characteristic impedance, sound absorption coefficient, and complex wavenumber plotted as a function of the relative variation x.

Contrary to the acoustical characteristic impedance, the acoustical surface impedance is not an intrinsic acoustical property of the sample (i.e., it depends on the thickness, rear boundary condition, and excited side). It has always a physical meaning even when the material is asymmetric—its determination does not require the symmetry property. This is not the case with the acoustical characteristic impedance. In fact, all the methods[3–6] used to measure the acoustical characteristic impedance assume the porous material to be symmetrical in terms of its acoustical properties. Because of that, the acoustical characteristic impedance loses its physical meaning when the material is asymmetric. For these reasons, it is not likely to quantify the asymmetry of a porous material from the acoustical characteristic impedance; the most suitable indicator of the asymmetry should be defined from the acoustical surface impedance (i.e., from RD of $\tilde{Z}_s$).

4. Criterion for asymmetry

From the previous analysis, a material can be considered acoustically symmetric if a hard backed sample of the material yields exactly the same surface impedance when measured in its direct and inverted configurations. Since such a situation is unlikely to happen due to some random noise and imperfect fabrication process, one needs to define a threshold to asymmetry based on the proposed index of asymmetry (i.e., RD of $\tilde{Z}_s$).

To define the threshold, let us analyze how RD of $\tilde{Z}_s$ varies with the reduced frequency and with the change of the geometrical properties of the pores for the porous material with oblique cylindrical pores described in the previous section. The reduced frequency is defined as $\varpi = \omega/\omega_v$ with the viscous transition frequency $\omega_v = 8\eta \cos\theta/\rho_0 r^2$, where ρ_0 and η are the density and viscosity of air. From the results of the simulations shown in Fig. 3(a), one can observe the following:

- the typical evolution of RD as a function of x and ϖ from the low frequency ($\varpi < 1$) to high frequency ($\varpi > 1$) regime;
- the inflection point of the curves is very close to the viscous transition frequency (i.e., $\varpi = 1$) of the material (the inflection can be observed only if the material is sufficiently asymmetric and if the transition frequency is in the frequency range of the analysis);
- the maximum of a RD curve occurs approximately at $1.5 \cdot \omega_v$ (i.e., $\varpi = 1.5$);
- at $\varpi \to 0$, and for RD lower than 10%, RD is a good estimate of the overall variation of the geometrical properties (i.e., $x \approx$ RD at $\varpi \to 0$);

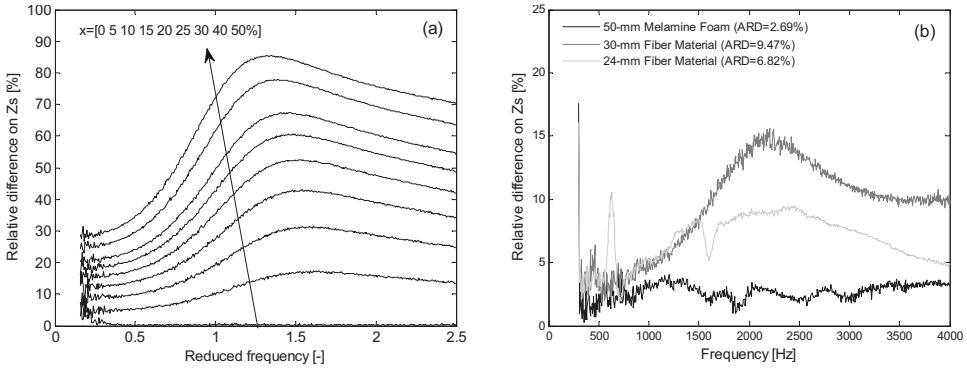


Fig. 3. Relative difference (RD) curves computed (a) from simulations on the two-layered porous system having oblique circular cylindrical pores with different values of the relative variation x, (b) from experimental measurements on melamine foam, and two fiber materials.

- at $\varpi = 1$, and for RD lower than 30%, RD/2 is a good estimate of the overall variation of the geometrical properties (i.e, $x \approx \mathrm{RD}/2$ at $\varpi = 1$);
- at $\varpi = 1.5$, and for RD lower than 45%, RD/3 is an estimate of the overall variation of the geometrical properties (i.e, $x \approx \mathrm{RD}/3$ at $\varpi = 1.5$);

From these observations, assuming that 5% variation in the geometrical properties of the material is tolerable, and assuming that the same behavior would be observed for any porous materials, one can suggest a material to be acoustically asymmetric if one of the following criteria is verified:

$$\begin{cases} \mathrm{RD} \geqslant 10\,\%\,; \text{ at the inflection.} \\ \max(\mathrm{RD}) \geqslant 10\,\%\,; \text{ when only the low frequency regime is observable.} \\ \max(\mathrm{RD}) \geqslant 15\,\%\,; \text{ when only the high frequency regime is observable.} \end{cases} \tag{3}$$

To simplify, one could also use the following criterion of asymmetry over the frequency range of analysis:

$$\max(RD(\omega)) \geqslant 10\,\%\,. \tag{4}$$

5. Experimental tests

Here, the surface impedance of three real materials in their direct and inverted configurations are experimentally tested with a 44.5-mm-diam impedance tube. The validity range of the tube is from 300 to 4000 Hz. The first material is a 50-mm-thick melamine foam known to be homogeneous. The second material is a 30-mm-thick fiber material. The third material is another 24-mm-thick fiber material. For each material, the surface impedance measurements are performed following standard ASTM E1050 on each side of the material sample backed by the rigid termination of the tube. From these measurements, the RD curves are computed and plotted in Fig. 3(b) in function of the frequency. One can observe that the typical behavior previously discussed is observed for the real materials. From these curves, it is clear that the melamine foam is acoustically symmetric (RD always smaller than 5%). Also, the 24-mm-thick fiber material can be considered acoustically symmetric $[\max(\mathrm{RD}(\omega)) < 10\,\%]$. However, the 30-mm-thick fiber material seems to suffer from a slight asymmetry. In fact, the inflection point is observable and occurs close to 1700 Hz. At this inflection point, RD=10%. From the first criterion given in Eq. (3), this material is at the limit of asymmetry.

6. Conclusion

In this work, a simple method has been proposed in order to quantify the asymmetry of sound absorbing porous materials. The method is based on a relative difference (RD) curve. This RD curve is obtained from surface impedance measurements performed on both sides of the tested material. From the RD curve, a threshold of asymmetry has been proposed and experimentally tested on three real materials. Above this threshold, the evaluation of the intrinsic dynamic properties of the material should not be done using impedance tube methods.[3–6] To prevent any erroneous interpretations or acoustic designs, it is suggested to compute the RD curve before concluding on the results obtained with an impedance tube. Eventually, the RD curve could be systematically given for each material tested with an impedance tube following international standards (e.g., ASTM E1050 and ISO 10534).

Acknowledgments

This work was supported in part by grants-in-aid from Alcan and N.S.E.R.C. A part of the research presented in this paper was also supported by the Fonds québécois de la recherche sur la nature et les technologies (F.Q.R.N.T.) by the intermediary of the Aluminum Research Centre – REGAL. The authors wish to thank Dr. Camille Perrot for his useful comments.

Appendix — Simulated measurements

The schematic view of the impedance tube configuration with hard termination is shown in Fig. 1. Assuming a normal incident plane wave of unit amplitude, the wave decomposition approach yields the following sound pressures in the four sections of the tube: $p_1(z)=e^{-jk_0z}+Ae^{jk_0z}$, $p_2(z)=Be^{-j\tilde{k}_1z}+Ce^{j\tilde{k}_1z}$, $p_3(z)=De^{-j\tilde{k}_2z}+Ee^{j\tilde{k}_2z}$, $p_4(z)=2F\cos(k_0z)$, where k_0 is the wavenumber in air, and $\tilde{k}_1$, and $\tilde{k}_2$ are the complex wavenumbers in porous layers 1 and 2, respectively. Here, each porous layer is modeled as an equivalent fluid,[2] and since it is made from circular cylindrical pores, its dynamic properties ($\tilde{k}$ and $\tilde{\rho}_{\mathrm{eq}}$) can be computed analytically[8] from the knowledge of its geometrical parameters (θ,r,n). Coefficients $A-F$ appearing in the previous equations are found by applying six interface conditions: Pressure continuity ($p_i=p_{i+1}$) and velocity continuity ($u_i=u_{i+1}$), with $i=1,2,3$, on interfaces at $z=0$, d_1, d_1+d_2, respectively. For each equivalent fluid, the velocity to use is the macroscopic value given by[1]: $u=-(1/j\omega\tilde{\rho}_{\mathrm{eq}})(dp/dx)$. Once the coefficients are found, one can evaluate the sound pressures p_i at microphones i (with $i=1,2$) and add a random noise as follows: $p_{\mathrm{ni}}(\omega)=p_i(\omega)+W\cdot(rand(\omega)+j\cdot rand(\omega))|_{i=1,2}$, where *rand* is a normally distributed random number between 0 and 1, and W is the level of noise. For the current analysis, $W=0.01$ is used. Next, as in standard ASTM E1050, the transfer function $H_{12}=p_{1n}/p_{2n}$ is computed. An averaging on 100 simulated transfer functions is made to replicate a typical impedance tube measurement. From this average, the acoustical surface impedance is computed $\tilde{Z}_s=jZ_0(\sin(k_0(L+s))-H_{12}\sin(k_0L))/(H_{12}\cos(k_0L)-\cos(k_0(L+s)))$, where Z_0 is the characteristic impedance of air. From this calculation method, one can fix the thickness of the rear air plenum to zero (i.e., $h_1=0$) to obtain $\tilde{Z}_s$ of the material on hard backing, and evaluate its sound absorption by $\alpha=1-|(\tilde{Z}_s-Z_0)/(\tilde{Z}_s+Z_0)|^2$. Finally, by using two different thicknesses for the air plenum, one obtains $\tilde{Z}_{s1}$ and $\tilde{Z}_{s2}$ for the tested material backed by the respective air plenum. From these two surface impedances; one can deduce the intrinsic dynamic properties $\tilde{Z}_{\mathrm{eq}}$ and $\tilde{k}$ of the material as they would be experimentally obtained from the Utsuno *et al.* two-cavity impedance tube method.[4]

References and Links

[1] J.-F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier, New York, 1993).
[2] R. Panneton, "Comments on the limp frame equivalent fluid model for porous media," J. Acoust. Soc. Am. **122**, EL217–EL222 (2007).
[3] R. A. Scott, "The absorption of sound in a homogeneous porous medium," Proc. Phys. Soc. London **58**, 358–368 (1946).

[4]H. Utsuno, T. Tanaka, T. Fujikawa, and A. F. Seybert, "Transfer function method for measuring characteristic impedance and propagation constant of porous materials," J. Acoust. Soc. Am. **86**, 637–643 (1989).
[5]T. Iwase, Y. Izumi, and R. Kawabata, "A new measuring method for sound propagation constant by using sound tube without any air spaces back of the test material," *Proc. Inter-Noise*, Christchurch, New Zealand (1998).
[6]B. H. Song and J. S. Bolton, "A transfer-matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials," J. Acoust. Soc. Am. **107**, 1131–1152 (2000).
[7]L. De Ryck, W. Lauriks, Z. E. A. Fellah, A. Wirgin, J. P. Groby, P. Leclaire, and C. Depollier, "Acoustic wave propagation and internal fields in rigid frame macroscopically inhomogeneous porous media," J. Appl. Phys. **102**, 024910 (2007).
[8]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).

Low-frequency Fourier analysis of speech rhythm

Sam Tilsen and Keith Johnson
Department of Linguistics, University of California, Berkeley, 1203 Dwinelle Hall, Berkeley, California 94720
tilsen@berkeley.edu, keithjohnson@berkeley.edu

Abstract: A method for studying speech rhythm is presented, using Fourier analysis of the amplitude envelope of bandpass-filtered speech. Rather than quantifying rhythm with time-domain measurements of interval durations, a frequency-domain representation is used—the *rhythm spectrum*. This paper describes the method in detail, and discusses approaches to characterizing rhythm with low-frequency spectral information.

PACS numbers: 43.70.Jt, 43.70.Kv, 43.70.Fq [AL]
Date Received: January 30, 2008 **Date Accepted:** April 23, 2008

1. Introduction

The most successful methods of characterizing cross-linguistic differences in speech rhythm (syllable timing versus stress timing) use interval durations to describe the temporal patterns of speech (Cummins and Port, 1998; Dauer, 1983; Lehiste, 1977; Port *et al.*, 1987; Ramus *et al.*, 1999; Roach, 1982). In Pike's (1945) and Abercrombie's (1967) approach, speech rhythm is defined in terms of the intervals between the onsets of linguistic units—syllables, moras, or feet. The failure to find regularity in these interval durations (Bolinger, 1968; Lehiste, 1977) led to a reconsideration of speech rhythm in terms of temporal properties of consonantal and vocalic intervals (Dauer, 1983; Ramus *et al.*, 1999).

Investigations of the beat of a syllable (Allen, 1972, 1975) and its perceptual moment of occurrence (Morton *et al.*, 1976; Howell, 1988; Pompino-Marschall, 1989) have revealed that speech rhythm (defined as the perceived interval between beats) is influenced by characteristics of the amplitude envelope of energy between beat locations. This observation leads one to consider whether the acoustically defined intervals used in prior tests of the isochrony hypothesis (Nakatani *et al.*, 1981) are perceptually relevant. This concern is heightened given that pitch accent placement in ordinary English discourse (e.g., Ladd, 1996) does not give intonational prominence to every stressed syllable. That is, if rhythm in a "stress-timed" language is sometimes governed by timing between intonationally prominent stressed syllables, leaving out lexically "stressed" but nonaccented syllables, then attempts to find isochrony may have failed because they made a false assumption about the units of timing.

This paper describes the use of a spectro-temporal method of rhythmic analysis that makes no prior assumptions about the rhythms that should be found or the linguistic units that might define beats for any particular stretch of speech. Our method finds that while some utterances in English do exhibit stress-based rhythm, others have a clear syllable-based rhythm, and still others exhibit more regular intervals on a phrasal time scale, i.e., between pitch-accented syllables.

2. Method

Duration measurements represent an interval of speech with a single number, thereby neglecting information about the amplitude envelope of the speech signal. From a naive perspective, this omission might seem odd, but it is so common that it is almost never explicitly noted in methodological appraisals. One culprit for this may be the metaphor in which linguistic units are containers. This metaphor structures our theoretical constructs of the syllable and metric foot, encouraging us to reason about them in some of the same ways we reason about containers. Specifically, in some circumstances, the contents of containers are irrelevant and it is their *sizes* which are important. In many approaches to characterizing rhythm, the duration of a syllable or foot (or intersyllabic or interstress interval) is analogous to the size of a container, and

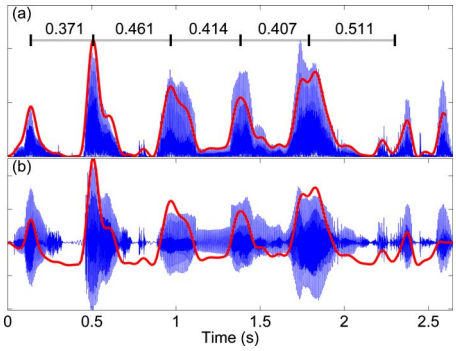


Fig. 1. (Color online) (a) Amplitude envelope superimposed over magnitude of bandpass-filtered signal; intervals between the several most prominent peaks in the amplitude envelope are shown. (b) windowed, mean-subtracted amplitude envelope over original acoustic signal.

its contents are either considered irrelevant, highly abstracted (e.g., labeled *vocalic* or *consonantal*), or thought to consist of other containers (i.e. syllables "within" feet, moras within syllables). Our approach here is to give much less attention to where intervals begin and end, and more attention to the acoustic contents of those intervals. We do this by analyzing the power spectrum of the slowly undulating amplitude envelope of speech.

To illustrate, we use a 2.6 s stretch of speech, in which a male speaker says "at least based on money raised it looks like…" (to listen, click on the link to Mm1 below). First, to capture mainly vocalic energy and filter out glottal energy and obstruent noise, we apply a first-order Butterworth filter with a passband of 700–1300 Hz. Note that this filter has been used to detect *p*-centers, which are salient moments near the onsets of vowels (Cummins and Port 1998). Next we lowpass filter the magnitude of the signal using a fourth-order Butterworth filter with a 10 Hz cutoff, downsample to 80 Hz, and apply a correction for the phase delays of the filters (45 ms, i.e., the sum of the mean phase delays of the filters in their passbands). The resulting signal represents slow changes in vocalic energy. Figure 1(a) shows the lowpass-filtered magnitude of vocalic energy (henceforth *amplitude envelope*) superimposed over the magnitude of the bandpass-filtered waveform. Next we window the amplitude envelope using a Tukey window ($r=0.1$) and subtract the mean, as shown in Fig. 1(b) superimposed over the original waveform. Before performing the spectral analysis, we zero-pad the amplitude envelope to produce a 2048-sample window, and then we normalize to unit variance.

To derive a frequency-domain representation from the time-domain amplitude envelope, we apply a Fourier transform, which partitions the variance of the time series into components of differing amplitude at each of N Fourier analysis frequencies, where N is the number of samples in the zero-padded amplitude envelope. The normalization to unit variance imposed upon the envelope is retained in the sum of the magnitude of the Fourier coefficients, a fact which follows from Parseval's Theorem (cf. Chatfield, 1975; Jenkins and Watts, 1968). We then analyze the power spectrum (the squared magnitude of the complex Fourier coefficients), which shows the contribution of each frequency component to the amplitude envelope.

The power spectrum of the amplitude envelope is arguably more appropriate for measuring rhythm than interval durations are. The spectral representation derives from a sort of wisdom of the crowd: each otherwise insignificant datapoint within all of the intervals in the entire signal contributes to the spectral representation of the signal—as if polling a bunch of people has given us a more accurate idea of the overall inclinations across the population. Indeed, in profound contrast to interval-based approaches, here no intervals whatsoever need be defined, only frequency components with corresponding phases and amplitudes.

Mm. 1. A stretch of speech in which a female speaker says "at least based on money raised it looks like…." This is a "wav" file (83 Kb).

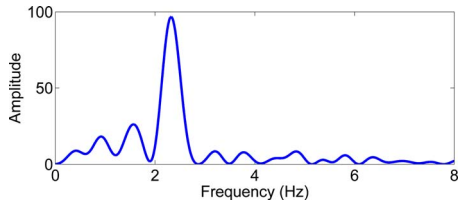


Fig. 2. (Color online) Power spectrum of the amplitude envelope that was shown in Fig. 1(b).

To relate the spectrum in Fig. 2 to the amplitude envelope in Fig. 1, observe that the average duration of the several most prominent peak-to-peak intervals in the amplitude envelope is about 430 ms, and as would be expected, there is a corresponding peak in the spectrum at approximately 2.3 Hz. Note that the duration of a chunk of speech defines a minimal frequency corresponding to the lowest frequency sinusoid that can fit within that duration. Any spectral peaks occurring below twice the minimal frequency do not reflect the presence of a periodicity within the signal; rather, these peaks indicate an imbalance in the distribution of energy in the signal, which can be manifested as substantially louder speech in one part of the utterance, perhaps arising from focus accent, lengthened fillers, etc.

3. Corpus analysis of rhythm in conversational speech

For current purposes, we are using speech from the Buckeye corpus (Pitt *et al.*, 2005), which is a collection of approximately 300 000 words of conversational speech between interviewers and 40 native central Ohio English speakers from a balanced set of ages and genders. The corpus was phonetically transcribed and segmented by transcribers trained to use acoustic and spectrographic information, following a number of conventions to ensure consistency. To analyze the corpus, we first extract chunks of speech with no interruption or nonspeech vocalization. Basic variables associated with each chunk include: chunk duration, syllable count, and speech rate (syllables per second).

For illustrative purposes in this report, we have analyzed chunks in a duration range of $\tau=[2,3]$ s, because we suspect that this range is useful for studying syllable- and foot-timed rhythms. In general, the choice of subset duration range depends upon the time scales of the rhythms being investigated; longer chunk durations are more appropriate for studying rhythms on phrasal time scales. For the present analysis, we divided chunks longer than 3 s into smaller chunks in the desired range, randomly perturbing their durations to provide a more uniform distribution of durations over the $[2,3]$ s range. Figure 3(a) shows the waveform and amplitude envelope of a chunk of speech with a high-amplitude periodicity near 4 Hz in the rhythm spectrum; also shown are citation, transcription, and deleted phones. In this chunk the speaker says "…category of Forrest Gump because Forrest Gump was great guy" (to listen click on the link below). Figure 3(b) shows the power spectrum of this chunk compared to the mean and 2.5 standard deviation region (shaded) for all 2–3 s chunk spectra. The vertical line represents twice the lowest frequency corresponding to the duration of the chunk—all frequencies lower than this correspond to less than two cycles in the amplitude envelope. Larger chunk durations should be used for analyses of lower-frequency, phrasal rhythms; however, larger chunks introduce more variability on syllabic time scales and thus tend to blur the rhythm spectrum at higher frequencies.

Table 1 gives an indication of how often rhythmic speech occurs in the dataset by showing the percentages of chunks with a spectral peak exceeding 50 amplitude units for each of several frequency ranges. These data indicate that (>1 Hz) high-amplitude periodicity occurs in approximately 23.2% of the 2–3 s chunks in the corpus. The presence of periodicity in a variety of frequency ranges shows that speech is rhythmic on stress and syllabic time scales. Analyses conducted with longer duration chunks (not shown) have revealed phrasal (0.33–1 Hz) rhythms as well.

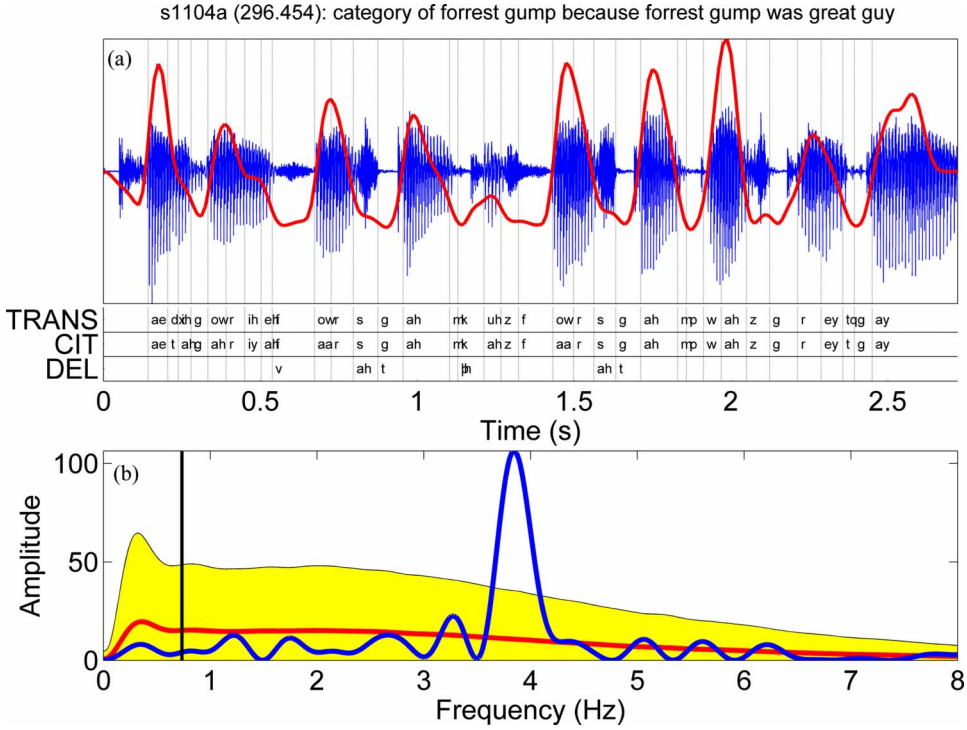


Fig. 3. (Color online) (a) Waveform and amplitude envelope of a chunk of speech, along with citation, transcription, and deleted phones; (b) power spectrum (peaked line) compared to average power spectrum (relatively flat line) and ±2.5 s.d. region for the set of 2–3 s chunks (shaded). Vertical line at twice the minimal frequency corresponding to the duration of the chunk.

Mm. 2. A stretch of speech in which a male speakers says "…category of Forrest Gump because Forrest Gump was great guy." This is a "wav" file (86 Kb).

We visualize the variability in a set of spectra by examining the distribution of peak frequencies and amplitudes, as in Fig. 4(a). For each spectrum, we locate one or more (but in this case one) of the highest peaks within a range of frequencies and then construct a two-dimensional Gaussian kernel density plot. To illustrate an appropriate level of detail, we use an amplitude range from the 0.1 percentile to the 99.9 percentile of amplitude values, an amplitude bandwidth of 5% of this range, and a frequency kernel bandwidth of 0.25 Hz. The most common low-frequency peak in this dataset is at about 1.6 Hz (i.e., a period of 625 ms).

Density plots also offer a useful way to compare datasets by inspecting the difference between density matrices. Figures 4(b) and 4(c) show peak frequency/amplitude density differences between subsets of data consisting of chunks with and without consonant and vowel deletions, where speech rate has been controlled by excluding chunks further than 1 s.d. from the mean speech rate. Deletions are identified by comparing the phonetic transcriptions with citation forms; deletions that occurred less than 80% of the time in their respective words were excluded in order to avoid artifacts due to overly specified citation forms. Rhythms tending to

Table 1. Counts of rhythmic chunks in several frequency ranges.

Rhythmic chunks	0–1 Hz	1–2 Hz	2–3 Hz	3–4 Hz	4–5 Hz	5–6 Hz	Total (>1 Hz)
Count	1354	897	871	396	109	27	2303
Percent	13.7%	9.1%	8.8%	4.0%	1.1%	<0.1%	23.2%

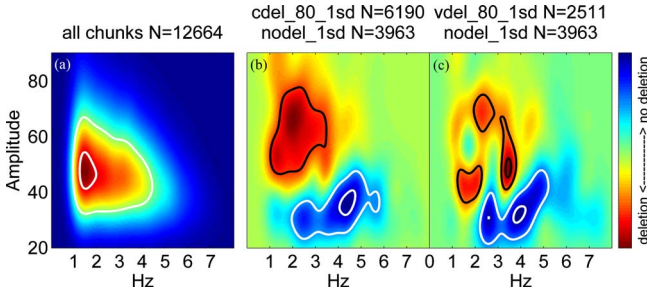


Fig. 4. (Color online) (a) Peak frequency/amplitude densities of 2–3 s chunk dataset where two highest peaks above twice the minimal frequency were taken from each spectrum; darkness corresponds to density and 50% and 90% contours are shown. (b, c) Density difference plots comparing rate-controlled subsets with and without consonant and vowel deletions; 90% and 50% positive and negative density contours are shown. (The information in this figure may not be properly conveyed in black and white.)

occur with more deletions are circled with dark 50% and 90% contour lines, and rhythms tending to occur without deletions are encircled with light contour lines. These figures indicate that very high-amplitude rhythms around 1–2 Hz are associated with consonant deletions, while rhythms around 3–4 Hz are more associated with vowel deletions. The predominance of the absence of deletion at lower amplitude periodicities indicates that when speech is less rhythmic, especially in the 2–3 Hz and 3.5–5 Hz ranges, deletion is less likely. Hence the data show a positive correlation between deletion and speech rhythmicity. Further, consonant and vowel deletions are most strongly correlated with highly rhythmic speech at different frequencies.

4. Conclusion and future directions

This report has presented a method for the quantitative analysis of rhythm that does not rely on interval durations, but rather, uses spectral analysis of the amplitude envelope of vocalic energy in speech. We believe that this "rhythm spectrum" analysis has the potential to augment studies of speech rhythm in a variety of ways. It offers a new approach to cross-linguistic rhythmic typology that involves statistical comparisons between large corpora of conversational speech. It can offer insights into rhythmic styles and characterizations of fluency from sociolinguistic and clinical perspectives. It may also shed light on relations between speech rhythm and intergestural timing, providing a deeper understanding of variation in conversational speech.

References and links

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Aldine, Chicago).
Allen, G. D. (**1972**). "The location of rhythmic stress beats in English: An experimental study, parts I and II." Lang Speech **15**, 72–100,179–195.
Allen, G. D. (**1975**). "Speech rhythm: Its relation to performance and articulatory timing," J. Phonetics **3**, 75–86.
Bolinger, D. (**1968**). *Aspects of Language* (Harcourt, Brace, and World, New York.)
Chatfield, C. (**1975**). *The Analysis of Time Series* (Chapman and Hall, London).
Cummins, F., and Port, R. (**1998**). "Rhythmic constraints on stress timing in English," J. Phonetics **26**, 145–171.
Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," J. Phonetics **11**, 51–62.
Howell, P. (**1988**). "Prediction of P-center location from the distribution of energy in the amplitude envelope," Percept. Psychophys. **43**(1), 90–93.
Jenkins, G. M., and Watts, D. G. (**1968**). *Spectral Analysis and Its Applications* (Holden-Day, San Francisco).
Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge Studies in Linguistics 79) (Cambridge University Press, Cambridge).
Lehiste, I. (**1977**). "Isochrony reconsidered," J. Phonetics **5**(3), 253–263.
Morton, J., Marcus, S., and Frankish, C. (**1976**). "Perceptual Centers (P-centers)," Psychol. Rev. **83**(5), 405–408.
Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (**1981**). "Prosodic aspects of American English speech rhythm," Phonetica **38**(1–3), 84–106.
Pike, K. L. (**1945**). *The Intonation of American English* (University of Michigan Press, Ann Arbor).
Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (**2005**). "The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability," Speech Commun. **45**(1), 89–95.
Pompino-Marschall, B. (**1989**). "On the psychoacoustic nature of the P-center phenomenon," J. Phonetics **17**, 175–192.

Port, R. F., Dalby, J., and O'Dell, M. (**1987**). "Evidence for mora-timing in Japanese," J. Acoust. Soc. Am. **81**(5), 1574–1585.
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **73**, 265–292.
Roach, P. (**1982**). "On the distinction between 'stress-timed' and 'syllable-timed' languages," in D. Crystal, *Linguistic Controversies* (Arnold, London).

Laboratory studies of near-grazing impulsive sound propagating over rough water

Qin Qin, Sergei Lukaschuk, and Keith Attenborough
Department of Engineering, The University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom
q.qin@hull.ac.uk, s.lukaschuk@hull.ac.uk, k.attenborough@hull.ac.uk

Abstract: Acoustic impulses due to an electrical spark source (main acoustic energy near 15 kHz) have been measured after propagating near to the water surface in a shallow container resting on a vibrating platform. Control of the platform vibration enabled control of water wave amplitudes. Analysis of the results reveals systematic variations in the received acoustic waveforms as the mean trough-to-crest water wave amplitude is increased up to 7 mm. The amplitudes of the peaks corresponding to specular reflections are reduced and the variability in the tails of the waveforms is increased.

PACS numbers: 43.50.Vt, 43.28.En, 43.28.Lv, 43.28.Mw [MS]
Date Received: March 3, 2008 **Date Accepted:** April 30, 2008

1. Introduction

If part of the propagation path is over water, then an improved understanding of impulsive sound propagation over water is important for the prediction of noise from explosions and sonic booms. Sound propagation over water is relevant also to noise from aircraft with landing approaches and takeoff trajectories over the sea, ships, offshore wind turbines, and recreational vessels such as power boats and jet skis. Since the specific impedance of water is greater than that of air by four orders of magnitude, water surfaces could be considered to be acoustically hard. On the other hand, the presence of water waves is likely to modify near-grazing sound propagation compared with that over a flat acoustically hard surface. The effects will depend on the relative magnitudes of the water wave amplitudes and sound wavelengths. Studies of the effects of small scale roughness on sound propagation near to the ground surface have shown that it alters the ground effect.[1,2] Roughness elements small compared with the incident sound wavelengths may be considered to change the effective impedance of the ground.[2–5]

Several attempts to predict the acoustical effects of water waves on sound propagation above them have been made assuming that the temporal variation in wave structures during the sound propagation can be ignored. The Generalized Terrain Parabolic Equation method has been used to predict propagation of impulses over gravitational waves on the surface of shallow water.[6] The transmission loss over an idealized water surface formed by intersecting circular segments was predicted to be substantially greater than over a smooth surface, particularly under downward refraction conditions, and it was suggested that this is the result of upward scattering. A wide-angle parabolic equation method has been used to predict sound propagation over a rough sea surface under various meteorological conditions.[7] The sea surface roughness was taken into account through an effective impedance. In the context of predicting sonic booms, the complex excess attenuation spectrum due to a line source above a boundary consisting of intersecting parabolas, which are representative of wind-driven deep water waves, has been predicted by a boundary element method (BEM) and has been used to deduce effective impedance as a function of sea state corresponding to mean wave heights between 0.25 and 7.5 m and for five incidence angles at each height.[8] The resulting predictions suggest that sea surface roughness could influence sonic boom profiles and rise times to an extent comparable to turbulence and molecular relaxation effects.

Despite these predictions, there are few data concerning sound propagation over the sea.[9] Field trials are relatively expensive and, given the impracticability of controlling the sea state, it is difficult to make systematic observations of relationships between water wave char-

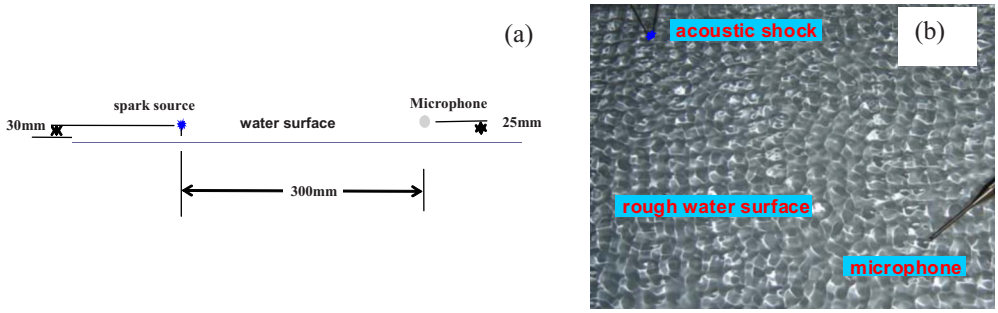


Fig. 1. (Color online) (a) Measurement geometry (b) snapshot of water surface during a test showing source and receiver locations.

acteristics and sound propagation effects. Laboratory measurements are an attractive alternative for exploring propagation phenomena associated with water surface roughness. Clearly it is not possible to reproduce wave heights on the order of meters in laboratory experiments. On the other hand, if the statistical properties of sea wave characteristics can be reproduced at laboratory scale using smaller amplitudes and wavelengths, then a scaling can be applied to the acoustic frequencies of interest. For example, use of water wave amplitudes between 1 and 5 mm and water wavelengths between 10 and 100 mm in the laboratory will require use of acoustic frequencies on the order of 10^3-10^4 times the actual frequencies of interest. However, this requires a controllable means of generating wave characteristics representative of the sea surface.

This paper reports on the results of laboratory measurements made using the acoustic impulses from an electric spark source (main acoustic energy near 15 kHz), over the surface of a water-filled container mounted on a vibrating platform. The water wave amplitude was controlled through the vibration of the platform. Analysis of received acoustic waveforms reveals systematic variations with water roughness amplitude.

2. Measurement arrangements

A 500 mm $\times$ 500 mm $\times$ 50 mm (deep) rectangular transparent-Perspex-walled container (subsequently called a "cell") was filled with water to a depth of 30 mm and mounted on a platform driven vertically by an electromagnetic shaker (V300, Gearing and Watson Ltd.). The amplitude and frequency content of the excitation signal were set by a programmable waveform generator. The vertical vibrations produced parametric instability and hence excited surface disturbances of more chaotic form than traveling waves. The water wave amplitude was measured by laser beam refraction and a PIV light sheet. The platform was vibrated at 30 Hz. Platform vibration amplitudes of between 1 and 3 mm were used to generate water waves with mean trough-to-crest amplitudes (subsequently referred to as mean amplitudes) of between 3 and 7 mm. Figure 1(a) shows a schematic elevation of the measurement arrangement and Fig. 1(b) shows a snapshot of the source, receiver, and water surface during one of the tests.

The spark source electrode gap was close to the water surface (30 mm) and the acoustic signals from electrically generated sparks were received by an 1/8 in. microphone at a height of 25 mm and a horizontal distance of 300 mm from the source. Given that the peak spark source energy is near 15 kHz and, assuming for convenience a peak airborne source impulse energy near 15 Hz, this would correspond to a "real" source being at 25 m height and 300 m range. The analogue microphone signals were fed to an NI data acquisition card and the resulting digitized information was captured and saved using LabView. Two different methods were used to trigger the data acquisition; synchronization with the spark ignition and a received amplitude threshold trigger. The amplitude trigger was found to give superior results and has been used for the data reported here.

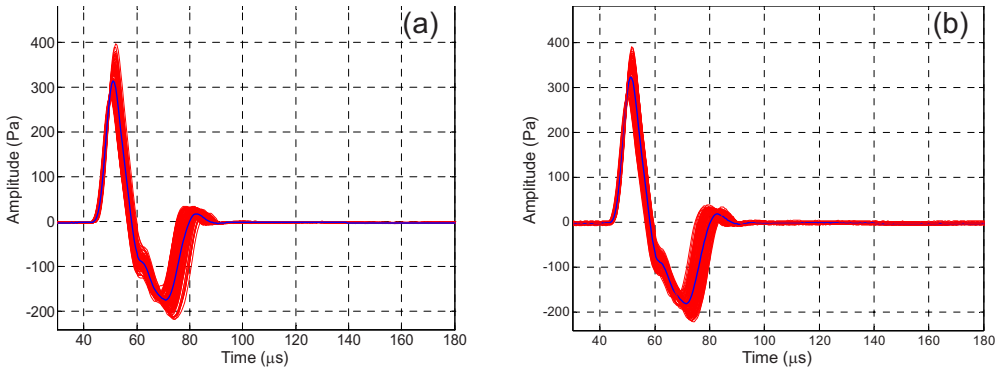


Fig. 2. (Color online) 100 free-field waveforms and their means measured 300 mm from an electrically generated spark source (a) without and (b) with noise from the electromagnetic shaker.

3. Results

Figure 2 shows 100 "free-field" acoustical pulse waveforms and their means measured at a distance of 300 mm from the electric spark source without and with background noise originating predominantly from the electromagnetic shaker. At 300 mm, the peak pulse sound pressure level (SPL) is approximately 140 dB re 20 μPa. Without the platform shaker in operation the background overall SPL in the laboratory was 48 dB. When the shaker was operating the back-

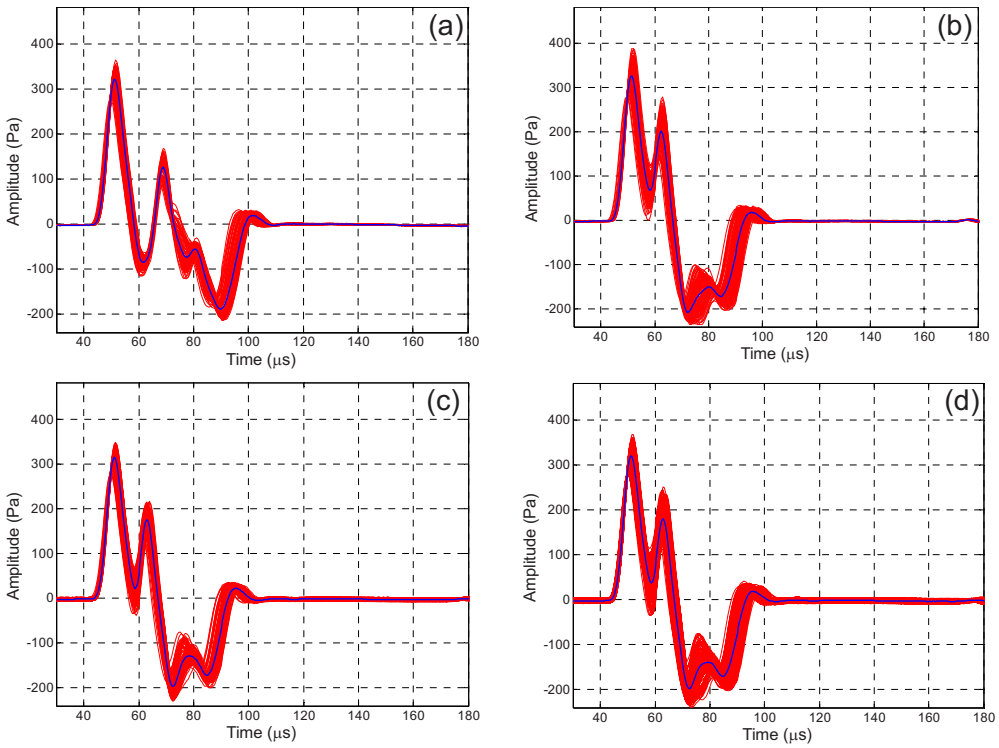


Fig. 3. (Color online) Individual and mean waveforms measured 300 mm from an electrically generated spark source over (a) a stationary flat water surface (b) a stationary smooth plastic plate at the same location (c) a vibrating plastic plate with vibration amplitude 2 mm, and (d) vibrating plastic plate with vibration amplitude 3 mm.

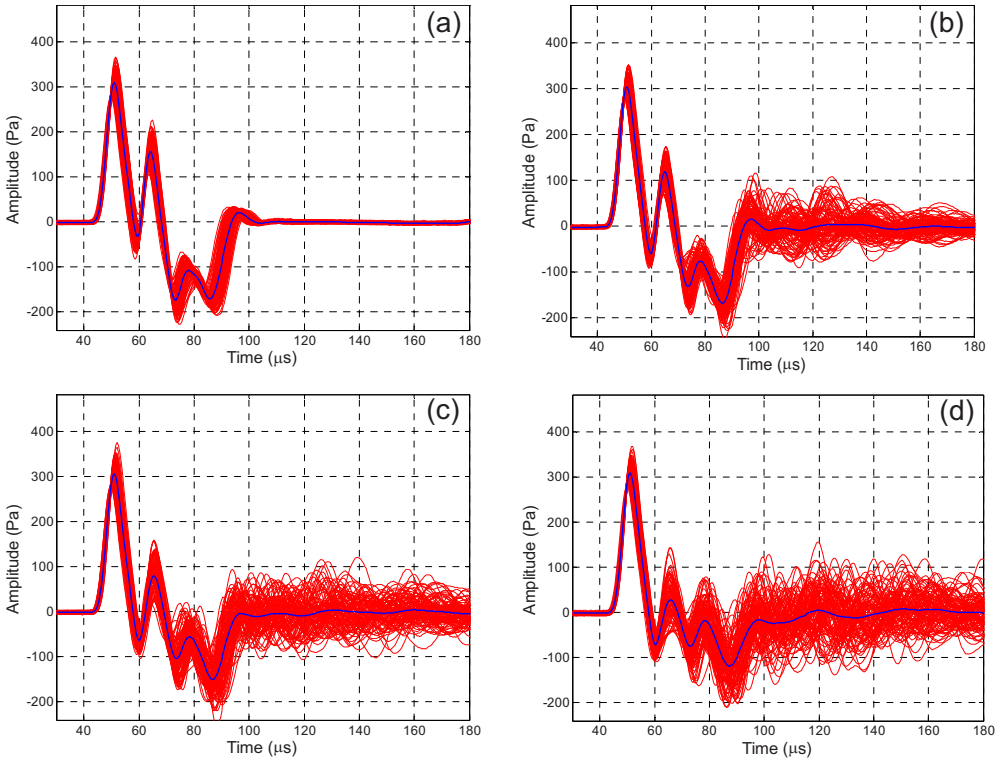


Fig. 4. (Color online) Individual and mean waveforms measured 300 mm from an electrically generated spark source over (a) a stationary flat water surface (b) a rough water surface with mean wave amplitude 3 mm (c) a rough water surface with mean wave amplitude 5 mm and (d) a rough water surface with mean wave amplitude 7 mm.

ground SPL was 69 dB. The pulse waveforms are more or less identical confirming the reproducibility of the waveforms and that an adequate signal-to-noise ratio was achieved when the shaker was operating.

Figure 3 shows received waveforms in the presence of a flat stationary water surface and when the cell was replaced by a smooth plastic sheet without and with vibration amplitudes of 2 and 3 mm. All of these waveforms show second peaks that correspond to specular reflections.

The source and receiver were slightly closer to the plastic sheet surface than to the water surface so the specular reflections occur earlier. In comparison with waveforms obtained over the stationary water surface [Fig. 3(a)], the tails of the waveforms obtained over the "stationary" plastic sheet [Fig. 3(b)] have a slightly wider dispersion. The appearance of a third minor peak indicates an additional reflection from the bottom of the plate. Vibration of the plastic sheet [Figs. 3(c) and 3(d)] appears to have relatively little effect on the received waveforms.

Figure 4 shows acoustical waveforms received over water waves with mean amplitudes of between 0 and 7 mm. Peaks in the waveforms corresponding to direct and specularly reflected arrivals may be observed. A third relatively minor peak corresponding to a reflection from the bottom of the cell appears to be slightly enhanced when the water surface is rough. Two significant features are the systematic decrease in the amplitude of the specularly reflected component (second peak) and a systematic increase of the oscillations in the "tails" of the waveforms. These features are quantified by the data in Table 1.

The received pulse waveform variability has been analyzed in two intervals; between 40 and 80 μs and between 80 and 180 μs. The former interval includes the direct and reflected

Table 1. Variation of acoustic waveform characteristics with mean water wave amplitude.

Mean wave amplitude (mm)	0	3	5	7
2nd peak amplitude (Pa)	155	118	79	27
SD of 100 waveforms (40–80 μs) (Pa)	26	24	27	28
SD of 100 waveforms (80–180 μs) (Pa)	7	25	34	41

arrivals, whereas the second interval is assumed to capture the tails of the waveforms. The variability has been expressed in terms of the mean of the standard deviations in the waveform amplitude observed at each time interval step in the analysis. With increasing water wave amplitude, the mean standard deviation of 100 waveforms between 40 and 80 μs is fairly constant with wave height, whereas there is a systematic increase in the mean standard deviation of the waveforms between 80 and 180 μs.

4. Conclusions

Laboratory experiments using an electrical spark source have shown two systematic effects of increasing mean water wave amplitude on the acoustic waveforms during near-grazing propagation. These are a decrease in the specularly reflected component and an increase in the variability of the tails of the waveforms. Further work will investigate the feasibility of attributing effective surface impedance spectra to the rough water surfaces and the possibility that the increased variability is associated with the generation of instantaneous acoustic surface waves.

Acknowledgment

The work was supported in part by EPSRC (UK) Grant No. EP/E027121/1.

References and links

[1]J. P. Chambers and J. M. Sabatier, "Recent advances in utilizing acoustics to study surface roughness in agricultural surfaces," Appl. Acoust. **63**, 795–812 (2002).
[2]K. Attenborough, T. Waters-Fuller, K. M. Li, and J. A. Lines, "Acoustical properties of farmland," J. Agric. Eng. Res. **76**, 183–195 (2000).
[3]K. Attenborough and S. Taherzadeh, "Propagation from a point source over a rough finite impedance boundary," J. Acoust. Soc. Am. **98**(3), 1717–1722 (1995).
[4]P. Boulanger, K. Attenborough, S. Taherzadeh, T. Waters-Fuller, and K. M. Li, "Ground effect over hard rough surfaces," J. Acoust. Soc. Am. **104**, 1474–1482 (1998).
[5]K. Attenborough and T. Waters-Fuller, "Effective impedance of rough porous ground surfaces," J. Acoust. Soc. Am. **108**(3), 949–956 (2000).
[6]E. Salomons, "Computational study of sound propagation over undulating water," (2007) (http://www.sea-acustica.es/WEBICA07/fchrs/papers/noi-05-008.pdf). Last viewed 6/19/2008.
[7]L. Johansson, "Sound propagation around offshore wind turbines," *Proceedings of the Tenth International Congress on Sound and Vibration*, Stockholm, Sweden, 1481–1488 (2003).
[8]P. M. Boulanger and K. Attenborough, "Effective impedance spectra for rough sea effects on atmospheric impulsive sounds," J. Acoust. Soc. Am. **117**(2), 751–762 (2005).
[9]K. Konishi and Z. Maekawa, "Interpretation of long term data measured continuously on long range sound propagation over sea surfaces," Appl. Acoust. **62**(10), 1183–2010 (2001).

Particle filtering for dispersion curve tracking in ocean acoustics

Ivan Zorych and Zoi-Heleni Michalopoulou
Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102
michalop@njit.edu

Abstract: A particle filtering method is developed for dispersion curve extraction from spectrograms of broadband acoustic signals propagating in underwater media. The goal is to obtain accurate representation of modal dispersion which can be employed for source localization and geoacoustic inversion. Results are presented from the application of the method to synthetic data, demonstrating the potential of the approach for accurate estimation of waveguide dispersion characteristics. The method outperforms simple time-frequency analysis providing estimates that are very close to numerically calculated dispersion curves. The method also provides uncertainty information on modal arrival time estimates, typically unavailable when traditional methods are used.

PACS numbers: 43.60.Hj, 43.60.Jn, 43.30.Pc [JC]
Date Received: March 25, 2008 **Date Accepted:** April 30, 2008

1. Introduction

The evolution of the frequency content of an acoustic signal with time often acts as a "signature" of the propagation medium. This is often the case with broadband signals with frequencies of a few hundred Hz propagating long distances in underwater environments. The variation pattern of frequency content with time reveals dispersion characteristics of the waveguide, which facilitates estimation of modal arrival times and amplitudes for various modes and frequencies. Estimates of these quantities can be employed in conjunction with a global or local optimization technique for source localization and environmental parameter estimation.[1–4] Dispersion estimation for inversion in underwater acoustics has been typically pursued with simple short time Fourier transforms (STFTs) and wavelet analysis.[2–4]

Following the concept of mode identification with state-space models,[5,6] we propose an approach involving particle filtering to accurately extract arrival time information from time-frequency representations of signals and to quantify the uncertainty in arrival time estimation. Our work builds on particle filtering approaches for tracking time-varying spectral features mostly in speech, music, and electroencephalogram data.[7–12] These techniques combine an observation equation, that is used in likelihood calculation based on observed data, and a state equation that determines how frequency evolves at each time step. In essence, these methods extend particle filtering methods that have been commonly applied to multiple source tracking. Based on principles of Kalman Filtering,[13] particle filtering techniques are flexible, can deal with varying numbers of trajectories and types of noise, and do not require linearity in the observation and state equations.[14–17]

The paper is organized as follows: Section 2 briefly discusses particle filtering approaches to target tracking. Section 3 describes how particle filtering methods for motion can be employed in modal trajectory tracking from STFT data and presents results from the application of the proposed approach to synthetic data. Conclusions follow in Sec. 4.

2. Particle filtering for source tracking

Let state vector $\mathbf{x_k}$ represent the location coordinates of a source emitting the measured signal at time k, where $k=1,\ldots,n$; let $\mathbf{y_k}$ be data observations (acoustic signal measurements, for example) also at time k. Our goal is to sequentially estimate the state vector $\mathbf{x_k}$ at times k

$=1,2,\ldots,n$, as, at each k, a new observation vector $\mathbf{y_k}$ becomes available. Two equations are used to model this problem. The state equation describes the evolution of the state vector $\mathbf{x_k}$ with time: $\mathbf{x_k}=q_k(\mathbf{x_{k-1}},\xi_\mathbf{k})$, where q_k is a function relating the state vector at time k to that at time $k-1$ and $\xi_\mathbf{k}$ is noise with a known probability distribution. The observation equation relates measurements $\mathbf{y_k}$ to state vector $\mathbf{x_k}$: $\mathbf{y_k}=h_k(\mathbf{x_k},\gamma_\mathbf{k})$, where h_k is a function relating data $\mathbf{y_k}$ to unknown parameters $\mathbf{x_k}$ and $\gamma_\mathbf{k}$ is noise with a known probability distribution. The likelihood function for $\mathbf{x_k}$ after observing data $\mathbf{y_k}$ is formed with the help of h_k and the statistical model for $\gamma_\mathbf{k}$.

Examining the problem within a Bayesian framework,[14–16,18,19] we are interested in deriving full posterior probability distributions for $\mathbf{x_k}$. The initial distribution of the state vector, $p(\mathbf{x_0})$, is assumed to be known. Let $D_k=[\mathbf{y_1},\mathbf{y_2},\ldots,\mathbf{y_k}]$ be the set of the first k observations. The aim is to estimate $p(\mathbf{x_k}|D_k)$, the posterior distribution of the state vector for time k given all observations.

If the posterior $p(\mathbf{x_{k-1}}|D_{k-1})$ is known at moment $k-1$, then $p(\mathbf{x_k}|D_{k-1})=\int p(\mathbf{x_k}|\mathbf{x_{k-1}})p(\mathbf{x_{k-1}}|D_{k-1})d\mathbf{x_{k-1}}$. When a new observation $\mathbf{y_k}$ becomes available, it can be incorporated in $p(\mathbf{x_k}|D_{k-1})$ as follows:[17]

$$p(\mathbf{x_k}|D_k)=\frac{p(\mathbf{y_k}|\mathbf{x_k})p(\mathbf{x_k}|D_{k-1})}{p(\mathbf{y_k}|D_{k-1})}. \tag{1}$$

The denominator in Equation (1) is the normalizing constant of $p(\mathbf{x_k}|D_k)$ and may be expressed as: $p(\mathbf{y_k}|D_{k-1})=\int p(\mathbf{y_k}|\mathbf{x_k})p(\mathbf{x_k}|D_{k-1})d\mathbf{x_k}$. The posterior distribution $p(\mathbf{x_k}|D_k)$ contains all information provided from the data, the observation equation, and the noise model about target state $\mathbf{x_k}$ at time k.

If state and observation equations are linear and noise components ξ and γ are normally distributed, the described process propagates the Gaussian posterior distribution $p(\mathbf{x_k}|D_k)$ via its mean and covariance matrix, forming a Kalman Filter. Explicit calculation of the probability distribution of Eq. (1) is, however, possible only in limited cases. When this is not feasible, the probability distribution of Eq. (1) may be sequentially approximated.

When a distribution $p(\mathbf{x_k}|D_k)$ is unknown, it can be approximated as: $p(\mathbf{x_k}|D_k)\approx\sum_{i=1}^{M}w_k^i\delta(\mathbf{x}_k-\mathbf{x_k^i})$, where $\mathbf{x_k^i}$, $i=1\ldots,M$, are sample vectors from a known, simpler distribution than p, and w_k^i are suitable weights; $\delta(\cdot)$ is the Dirac function. Vectors $\mathbf{x_k^i}$ are referred to as particles and sets $\{\mathbf{x_k^1},\ldots,\mathbf{x_k^M}\}$ are known as clouds. Various approaches, commonly referred to as particle filters, have been proposed for efficient and accurate estimation of probability distributions based on the above principle. Sequential importance sampling (SIS), or Bayesian bootstrap filtering,[14,15,20] is such a process. SIS with an additional resampling step to avoid degeneracy,[14] SIR, is the approach followed here.

In summary, SIR works as follows: Suppose that at time $k-1$ there is a particle cloud $\{\mathbf{x_{k-1}}^1,\mathbf{x_{k-1}}^2,\ldots,\mathbf{x_{k-1}^M}\}$ of size M that, with associated weights, approximates the posterior distribution $p(\mathbf{x_{k-1}}|D_{k-1})$. Cloud $\{\mathbf{x_{k-1}^1},\mathbf{x_{k-1}^2},\ldots,\mathbf{x_{k-1}^M}\}$ is then propagated through the state equation. This transforms the cloud into cloud $\{\mathbf{x_k^{1*}},\mathbf{x_k^{2*}},\ldots,\mathbf{x_k^{M*}}\}$; each particle in the latter cloud has weight $1/M$. Having measured $\mathbf{y_k}$, we reevaluate the weight of each particle: $w_k^i=p(\mathbf{y_k}|\mathbf{x_k^{i*}})/\sum_{j=1}^{M}p(\mathbf{y_k}|\mathbf{x_k^{j*}})$, where distribution $p(\mathbf{y_k}|\mathbf{x_k^{j*}})$ is defined by the observation equation and knowledge of the statistical behavior of errors in data measurements. These weights are used for the estimation of $p(\mathbf{x_k}|D_k)$. Particles are then resampled from $p(\mathbf{x_k}|D_k)$.

To implement the filtering process for $k=1$, an initial probability distribution for state vector $\mathbf{x_0}$ has to be selected; $p(\mathbf{x_0})$ is often considered to be a uniform distribution or can be more informative if prior information is available. One can sample from $p(\mathbf{x_0})$ to generate clouds and can then apply the algorithm outlined above to estimate subsequent states.

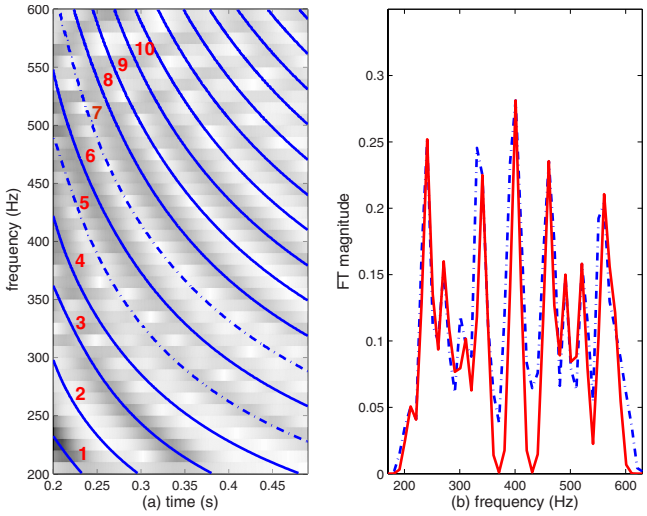


Fig. 1. (Color online). (a) Spectrogram of the acoustic signal; numerically calculated dispersion curves (in blue) are superimposed (trajectory numbers are indicated and dot-dashed lines are used for trajectories 5 and 7) and (b) plot of one spectrogram column for a selected time k (blue, dot-dashed) and corresponding model (red, solid).

3. Particle filtering for dispersion estimation

We assume that a source transmits a broadband signal with frequency content between 200 and 600 Hz which propagates in the ocean and is received at a hydrophone located 20 km away from the source; the ocean depth is assumed to be approximately 200 m. The spectrogram of the received signal, obtained via the application of STFT to the data, is shown in Fig. 1(a). We can identify modal trajectories in the spectrogram, manifested as darker shade curves starting from high frequencies and descending with time. These represent dispersion curves of different modes and demonstrate that, within each mode, different frequencies arrive at different times, because of their different propagation velocities. The same frequencies for different modes propagate with difference velocities as well. On the spectrogram, we have superimposed dispersion curves that have been numerically obtained for the particular waveguide, which illustrate how modal dispersion is demonstrated with spectrograms. Although there is a good match between the spectrogram and computed dispersion curves, spectrogram areas corresponding to such curves are often wide, complicating the accurate selection of frequency-time pairs from the curve. Also, some modal curves are not visible in the spectrogram; for example, the seventh trajectory counting from the bottom left corner is not distinguishable. Another example is the fifth trajectory which appears in the spectrogram only after 0.35 s. Not being able to accurately extract dispersion information can lead to mode misidentification, which, in turn, may result to errors in inversion methods.

The goal of this work is to extract frequency information for each mode at each time sample of the spectrogram; in essence, the focus is on the accurate estimation of arrival times of distinct frequencies and distinct modes. Vectors $\mathbf{y_k}$, the results of the application of STFT to a windowed section of length L of the received time series at time k, are the data that are employed in the observation equation. State vector, $\mathbf{x_k}$, can be written as $\mathbf{x_k}=[\theta_1^k, \theta_2^k, \ldots, \theta_{r(k)}^k, r(k)]$, where $\theta_j^k, j=1, \ldots, r(k)$ are frequencies arriving at time k and $r(k)$ is the number of modes arriving at $k=1$. Frequencies in the state vector evolve with time k following the state equation: $\theta_j^k=\theta_j^{k-1}-\xi_k^j$, where ξ_k^j is a Binomial random variable with parameters $N_k^j$and p_k^j. Parameter $N_k^j=\theta_j^{k-1}-\theta_{j-1}^{k-1}$ reflects distance between frequencies j and $j-1$ and keeps the modal trajectories from intersecting. The choice of a Binomial model for the stochastic component of the state equation is motivated by the nature of our tracking problem. It is *a priori* known that, with increasing

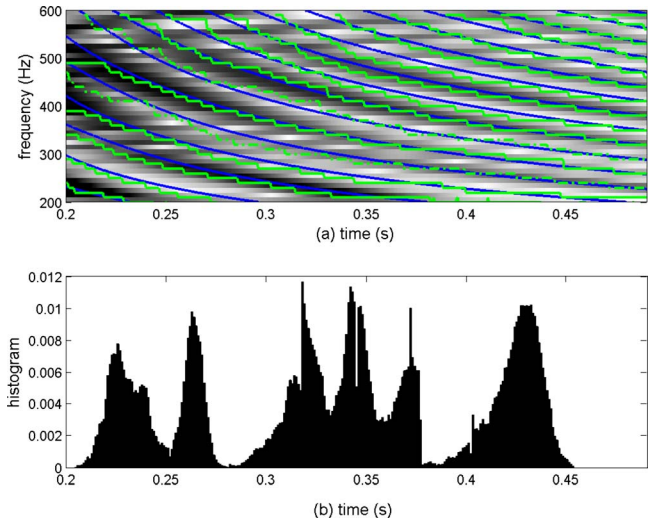


Fig. 2. (Color online). (a) Tracking results superimposed on spectrogram: Green step lines show estimated tracks, and smooth blue curves are the numerically calculated dispersion curves (dot-dashed green lines show trajectories 5 and 7); (b) combined histogram of arrival times of frequency 410 Hz in six consecutive trajectories.

time, each modal trajectory either slowly moves downward or remains approximately constant. At the same time, frequencies can only take discrete values because of Fourier calculations at discrete frequencies.

Binomial parameters p_k^j are estimated off line from the spectrogram before the implementation of the particle filtering process and are provided to the filtering process as prior information. The behavior of the parameters reflects the fact that the spectral trajectories move downward with time. Relatively large parameter values at the beginning of a trajectory indicate that the trajectory is expected to be steeper then than at later times. Close trajectories have similar binomial parameters because of similarity in their descending pattern. We also take into consideration that, for our signal, all trajectories end at 200 Hz; if the lowest frequency of a trajectory attempts to reach values below 200 Hz, it is eliminated and $r(k)$ decreases by 1. If at the moment k a new trajectory is detected at 600 Hz, $r(k)$ increases by 1.

We consider an observation equation relating data $\mathbf{y_k}$ to frequencies present in the signal that is similar to the time delay model of Ref. 21: $\mathbf{y_k}=\mathbf{y_{m,k}}+\boldsymbol{\gamma}_\mathbf{k}=\Sigma_{i=1}^{r(k)}a_i\nu(f-\theta_i^k)+\boldsymbol{\gamma}_\mathbf{k}$, $f=1,2,\ldots,L/2$, where perturbation $\boldsymbol{\gamma}_\mathbf{k}$ is assumed to be zero mean, normally distributed. Function $\nu(f-\theta_i^k)$ is centered at θ_i^k and has a maximum height of 1 and a spread parameter ϕ: $\nu(f-\theta_i^k)=\exp(-(f-\theta_i^k)/\phi)^2, f=1,2,\ldots L/2$. Parameter ϕ can be either estimated *a priori* before the particle filtering process or incorporated within the process as an unknown. Frequencies θ_i^k are generated via the state equation. Corresponding amplitudes a_i are obtained by identifying spectrogram amplitudes at frequencies θ_i^k. Model $\mathbf{y_{m,k}}=\Sigma_{i=1}^{r(k)}a_i\nu(f-\theta_i^k)$ provides a close match to the spectrogram. Figure 1(b) demonstrates the magnitude of a column of the signal spectrogram and the corresponding $\mathbf{y_m}$.

Under the above assumptions, the likelihood for the unknown frequencies is defined as: $p(\mathbf{y_k}|\mathbf{x_k})\propto\exp(-(\mathbf{y_k}-\mathbf{y_{m,k}})^T(\mathbf{y_k}-\mathbf{y_{m,k}})/2\sigma_\gamma^2)$. Using the sequential method of Sec. 2 and the state and observation equations of the present section, we estimated modal trajectories in the signal of Fig. 1(a) as: $\hat{\theta}_j^k=MAP(\theta_{j,l}^k)$, $l=1,2,\ldots,M$, where *MAP* is the *maximum a posteriori* estimate of θ_j^k (essentially, its most frequent value at time k). Trajectory estimates are shown in Fig. 2(a) (green step tracks). These tracks are superimposed on the dispersion curves also shown in Fig. 1(a) (blue lines) and the spectrogram. The estimated tracks with the proposed particle filtering method follow closely the true dispersion curves, unlike areas of the spectrogram under

these tracks which are often light colored, indicating that no trajectory is detected. As expected, in time-frequency regions where there is insufficient evidence about the behavior of the spectrogram, estimated trajectories show more variability (tracks 5 or 7, for example, counting from the left bottom corner).

As discussed in Sec. 1, arrival times for specific frequencies are often selected from time-frequency representations for geoacoustic inversion and source localization. It is of interest to also compute a measure of uncertainty on the extracted arrival times, which will then provide information on uncertainty in the estimation of the unknown geoacoustic and location parameters. To do so, we focus on each trajectory and one frequency within the selected trajectory at time k of the time-frequency representation; we then scan several time instances neighboring k to identify in how many particles the selected frequency appears for each instant within the time interval that is being examined. This process generates a histogram demonstrating how many times the arrival time of a frequency occurs within a given time interval and provides information on the uncertainty in arrival time estimates for a specific frequency. Figure 2(b) illustrates the process for six consecutive trajectories (between 5 and 10) and a frequency of 410 Hz, presenting a histogram of all arrival time samples for that frequency. Trajectories in the combined histogram have distinct peaks and different spreads, which point to different variances in the estimation of arrival times for the same frequency but different modes. Arrival time distributions for some trajectories are much wider than those of others, pointing to the need to model carefully uncertainty for each modal arrival and avoid equal variance assumptions.

4. Conclusions

A new approach is proposed for the extraction of dispersion curve estimates from multi-modal signals in dispersive underwater environments. The approach is based on particle filtering and provides estimates of posterior probability distributions on frequencies arriving at several time steps. Results demonstrate that the method provides sharper dispersion curve estimates than simple time-frequency analysis. In addition to providing connected modal trajectories that facilitate the computation of arrival times for distinct frequencies and modes, the method provides uncertainty information on arrival times, quantifying errors that can then be propagated into a geoacoustic inversion process employing modal arrival time information. Results from dispersion estimation with particle filters are expected to further improve by employing dispersion models such as those proposed in Refs. 6, 22, and 23.

Acknowledgments

This work was supported by the Office of Naval Research through Grant No. N000140510262.

References and links

[1]J. F. Lynch, S. D. Rajan, and G. V. Frisk, "A comparison of broadband and narrowband modal inversions for bottom geoacoustic properties at a site near Corpus Christi, Texas," J. Acoust. Soc. Am. **89**, 648–665 (1991).
[2]C.-S. Chen, J. Miller, F. Boudreaux-Bartels, G. Potty, and C. Lazauski, "Time-frequency representations for wideband acoustic signals in shallow water," in *Proceedings of OCEANS 2003*, San Diego, pp. 2903–2907.
[3]M. Taroudakis and G. Tzagkarakis, "On the use of the reassigned wavelet transform for mode identification," J. Comput. Acoust. **12**(2), 175–196 (2004).
[4]G. Potty, J. Miller, P. Dahl, and C. Lazauski, "Geoacoustic inversion results from the ASIAEX East China Sea Experiment," J. Acoust. Soc. Am. **29**(4), 1000–1010 (2004).
[5]J. Candy and E. Sullivan, "Model-based identification: An adaptive approach to ocean-acoustic processing," IEEE J. Ocean. Eng. **23**, 273–289 (1996).
[6]J. Candy and D. Chambers, "Model-based dispersive wave processing: A recursive Bayesian solution," J. Acoust. Soc. Am. **105**, 3364–3374 (1999).
[7]C. Dubois, M. Davy, and J. Idier, "Tracking of time-frequency components using particle filtering," in *Proc. IEEE ICASSP*, Philadelphia, 2005, Vol. **2**, pp. 887–891.
[8]C. Dubois and M. Davy, "Joint detection and tracking of time-varying harmonic components: A flexible Bayesian approach," IEEE Trans. Audio, Speech, Lang. Process. **4**, 1283–1295 (2007).
[9]N. Ikoma, "Estimation of time varying peak of power spectrum based on non-Gaussian nonlinear state space modeling," Signal Process. **49**, 85–95 (1996).
[10]C. Andrieu, M. Davy, and A. Doucet, "Efficient particle filtering for jump Markov systems. Application to time-

varying autoregressions," IEEE Trans. Signal Process. **51**, 1762–1770 (2003).
[11]S. Godsill, A. Doucet, and M. West, "Maximum a posteriori sequence estimation using Monte Carlo particle filter," Ann. Inst. Stat. Math. **53**, 82–96 (2001).
[12]R. Prado, M. West, and A. Krystal, "Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag-lead structure," J. R. Stat. Soc., Ser. C, Appl. Stat. **50**(1), 95–109 (2001).
[13]R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME, Ser. B **82**, 35–45, (1960).
[14]N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," IEE Proc. F, Radar Signal Process. **140**i2, 107–113 (1993).
[15]W. Gilks and C. Berzuini, "Following a moving target - Monte Carlo inference for dynamic Bayesian models," J. R. Stat. Soc. Ser. B (Stat. Methodol.) **63**, 127–146 (2001).
[16]A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer, New York, 2001).
[17]B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications* (Artech House, Boston, 2004).
[18]J. Vermaak, S. Godsill, and P. Perez, "Monte Carlo filtering for multi-target tracking and data association," IEEE Trans. Aerosp. Electron. Syst. **41**, 309–332 (2005).
[19]J. V. Candy and S. J. Godsill, "Bayesian space-time processing for acoustic array source estimation using a towed array," J. Acoust. Soc. Am. **120**, 3179 (2006).
[20]A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," Stat. Comput. **10**, 197–208 (2000).
[21]Z.-H. Michalopoulou and M. Picarelli, "Gibbs sampling for time-delay and amplitude estimation in underwater acoustics," J. Acoust. Soc. Am. **117**, 799–808 (2005).
[22]J. Candy and D. Chambers, "Internal wave signal processing: A model-based approach," IEEE J. Ocean. Eng. **21**, 37–52 (1996).
[23]G. Okopal, P. J. Loughlin, and L. Cohen, "Dispersion-invariant features for classification," J. Acoust. Soc. Am. **123**, 832–841 (2008).

Adaptive echolocation sounds of insectivorous bats, *Pipistrellus abramus*, during foraging flights in the field

Shizuko Hiryu,[a] Tomotaka Hagino, and Emyo Fujioka
Faculty of Engineering, Doshisha University, Kyotanabe 610-0321, Japan
shiryu@mail.doshisha.ac.jp, jgcqx789@yahoo.co.jp, dth0902@mail4.doshisha.ac.jp

Hiroshi Riquimaroux[a] and Yoshiaki Watanabe[a]
Faculty of Engineering and Bio-navigation Research Center, Doshisha University, Kyotanabe 610-0321, Japan
hrikimar@mail.doshisha.ac.jp, kwatanab@mail.doshisha.ac.jp

Abstract: Echolocation pulses emitted by wild *Pipistrellus abramus* were investigated while foraging for insects in the field. Similar to other European pipistrelles, the frequency structure during foraging varied. During the search phase, the bats emitted long shallow frequency-modulated pulses 9–11 ms in duration, whereas the maximum pulse duration of the bats approaching a large target wall in the laboratory was 3 ms. No significant difference was observed between decreases in the interpulse interval during these two approach flights. It is concluded that the bats use a long quasi-constant frequency pulse to find a weak echo from a small prey target.

PACS numbers: 43.80.Ka [CM]
Date Received: February 22, 2008 **Date Accepted:** May 4, 2008

1. Introduction

Echolocating bats emit brief sounds at frequencies from 15 kHz to over 200 kHz (Neuweiler, 1984) through their mouths or nostrils. Echolocation pulses generally consist of the constant frequency (CF) component or the frequency modulated (FM) component (or CF-FM combination type), and previous studies have examined how pulse type is related to the respective foraging habits of echolocating bats (e.g., Simmons *et al.*, 1979; Neuweiler, 1984; Schnitzler *et al.*, 2003; Thomas *et al.*, 2003). Given that CF pulses are well suited for detecting the movement of a target flying in a cluttered environment (Schnitzler and Henson, 1980), bat species such as members of the Rhinolophidae and Hipposideridae families that forage for insects close to or within foliage employ CF-FM pulses. Conversely, FM pulses are well suited to the precise measurement of echo travel time, which indicates the target distance from the bat (Simmons *et al.*, 1979). The structure of the echolocation pulse reflects adaptations to specific ecological constraints associated with the foraging behavior of different echolocating bat species (Simmons *et al.*, 1979; Neuweiler, 1984; Schnitzler *et al.*, 2003; Thomas *et al.*, 2003).

Pipistrellus bats, which are members of the family Vespertilionidae, use flexible FM pulses in terms of frequency structure during echolocation (Schnitzler *et al.*, 1987; Kalko and Schnitzler, 1993; Boonman and Schnitzler, 2005). In Japan, foraging behavior of *P. abramus* (Japanese house bats) is typically observed in large open spaces such as above rice fields and riparian areas during the summer. Although closely related pipistrelle bats such as *P. pipistrellus* and *P. kuhli* have been studied extensively (e.g. Schnitzler *et al.*, 1987; Kalko, 1995), the echolocation behavior of wild *P. abramus* has not been well investigated. Our recent study using a telemetry sound-recording technique allowed us to conduct detailed investigations of the echolocation behavior of *P. abramus* in the laboratory (Hiryu *et al.*, 2007). The purpose of the

[a]Present address: Faculty of Life and Medical Sciences, Doshisha University.

present study was to record echolocation pulses emitted by wild *P. abramus* to examine the basic acoustic characteristics of the pulses during foraging flights in the field. By comparing these observations to acoustical investigations of approach flights under the controlled environment of the laboratory, we investigated how *P. abramus* adapt their sonar signal design to obtain necessary information regarding a small prey target.

2. Materials and methods

Sound recordings were conducted over five days between June and July 2006 in Kyotanabe, southern Kyoto Prefecture, Japan. The study site was an open area over a large rice field near the campus of Doshisha University, where only *P. abramus* are regularly seen capturing airborne insects during summer evenings. Sound data were collected while foraging *P. abramus* flew from 30 min before sunset to 1 h after sunset. The experiments complied with the *Principles of Animal Care*, publication No. 86-23, revised in 1985, of the National Institutes of Health, and with current Japanese laws.

Echolocation pulses were recorded with a condenser microphone (B&K, 4939, Denmark) positioned 3 m above the ground. The sounds were amplified (B&K, 4939, Denmark), high-pass filtered (20 kHz; NF Corporation, model 3625, Yokohama, Japan), and stored on the magnetic tape of a DAT recorder (SONY, SIR-1000W, Tokyo, Japan). The data were digitally captured on a personal computer with 16 bits at a sampling rate of 384 kHz. Acoustic parameters of the echolocation pulse were analyzed from sonograms using the custom made program MATLAB. Only the fundamental frequency component was analyzed, and each pulse was extracted from the displayed sonogram. Pulse duration was determined from the sonogram at −25 dB relative to the peak intensity of the pulse. We measured inter-pulse intervals (IPI) and the minimum frequency of the FM sweep (terminal frequency). We used a Bat detector (Petterson, D200, Uppsala, Sweden) to locate emerging bats, and then visually monitored the foraging flights to identify periods during which only a single bat was flying near the microphone.

Flight experiments were also conducted with *P. abramus* in a steel-walled laboratory (8 mL × 3 mW × 2 mH). The animals were captured from a large colony roosting in bridge girders near the campus of Doshisha University. The bats were released at one end of the flight chamber and made to fly to the opposite end where a landing mesh (1 mW × 0.7 H) was attached to the wall. Flight behavior was recorded as a flying bat approached this wall for landing, and the recording procedure was the same as that reported previously (Hiryu *et al.*, 2007). Echolocation pulses were recorded using a custom-made wireless telemetry microphone system (Telemike) attached to the back of a bat with a piece of double-sided glue tape, with the microphone positioned approximately 1 cm above the mouth. The Telemike-transmitted signals were demodulated using a custom-made FM receiver and high-pass filtered at 20 kHz (NF Corporation, model 3625, Yokohama, Japan), then digitized using a DAT recorder with 16 bits at a sampling rate of 384 kHz. The flight behavior of the bats was recorded using two digital high-speed video cameras (NIPPON ROPER Co., Ltd., CR Imager model 2000s, Chiba, Japan) with 125 frames per second, and three-dimensional coordinates of the flying bats were reconstructed from these video images using motion analysis software (DITECT, Dipp-Motion 2D ver. 2.1). Five bats were used in the laboratory experiments.

3. Results

At the recording site, the bats usually foraged alone or with a few conspecifics, flying approximately 3–5 m above the ground over the large rice field. Figure 1(A) shows a typical sequence of echolocation sounds for hunting produced by *P. abramus*. Similar to other pipistrelles, the echolocation could be described with three particular patterns including search, approach, and terminal phases (Schnitzler *et al.*, 1987). The terminal phase was followed by a 100 ms silent period, suggesting a successful prey capture. During the search phase, IPI remained at approximately 100 ms and the pulse duration was kept at 9–11 ms [Fig. 1(B)]. Our recordings indicate that flying *P. abramus* is capable of capturing insects every 2–3 s. During the search phase, the terminal frequency of the pulse was almost constant at 40–43 kHz [41.23 ± 1.76 kHz, $n = 221$,

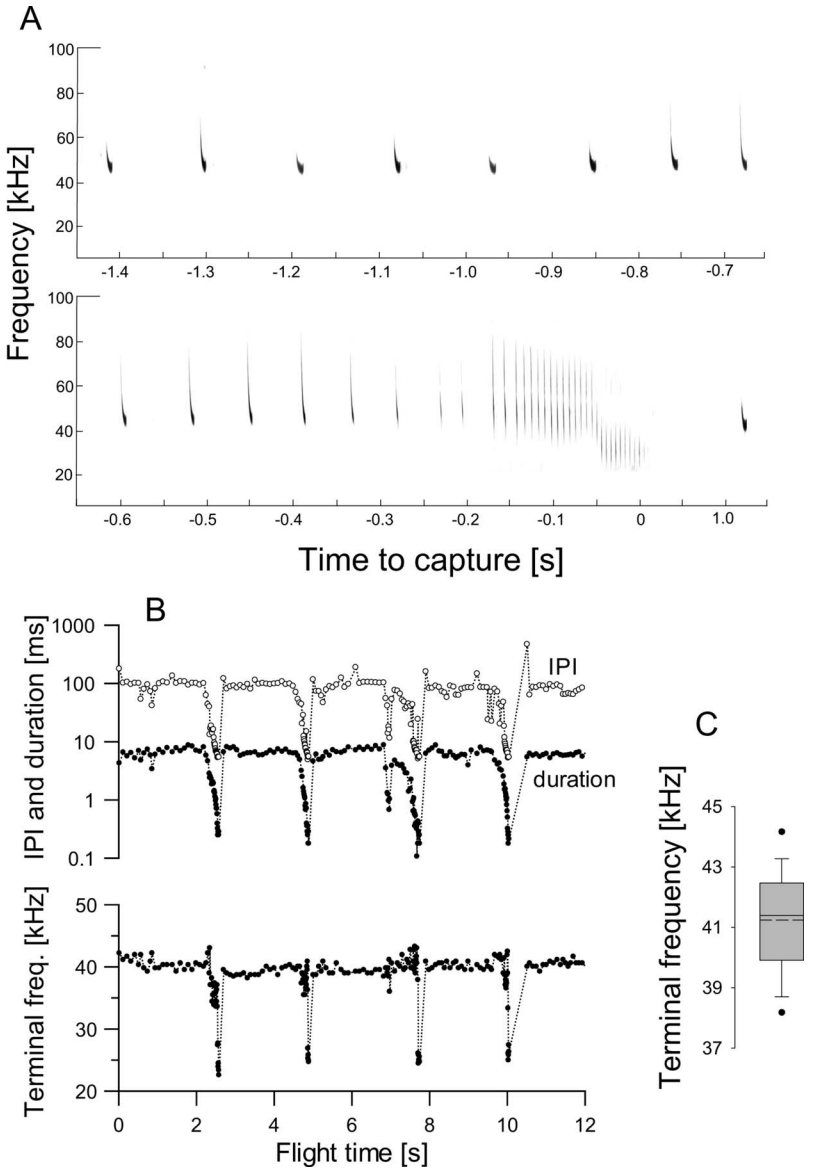


Fig. 1. Typical pulse emission sequence of *P. abramus* during foraging behavior in the field. (A) Sonogram of pulse train during pursuit and capture of insect prey. (B) Interpulse interval (top; open circle) and pulse duration (top; solid circle) and the terminal frequency (bottom) of the pulse emitted by a bat exhibiting continuous foraging. (C) Variation of the terminal frequency of the pulse during the search phase (five capturing flights; total: 221 pulses). The bottom and top boundaries of the box indicate 25% and 75% of the distribution of the data, respectively. The whiskers indicate the 10th and 90th percentiles, and the lower and upper dots indicate 5th and 95th percentiles, respectively. The dashed and solid lines within the box represent the mean (41.23 ± 1.76 kHz; mean $\pm$ SD) and median (41.39 kHz) values, respectively.

Fig. 1(C); Doppler effects caused by relative flight speeds between the bat and the microphone were not considered]. The approach phase began 0.7–0.8 s before capturing the prey, after which the bats shortened their IPI from 100 to 10 ms while the pulse duration decreased from approximately 10 to 1 ms. This was followed by a further decrease in IPI from 10 to 5 ms (buzz I), and the terminal frequency of the pulse was dropped from 40–43 kHz to approximately

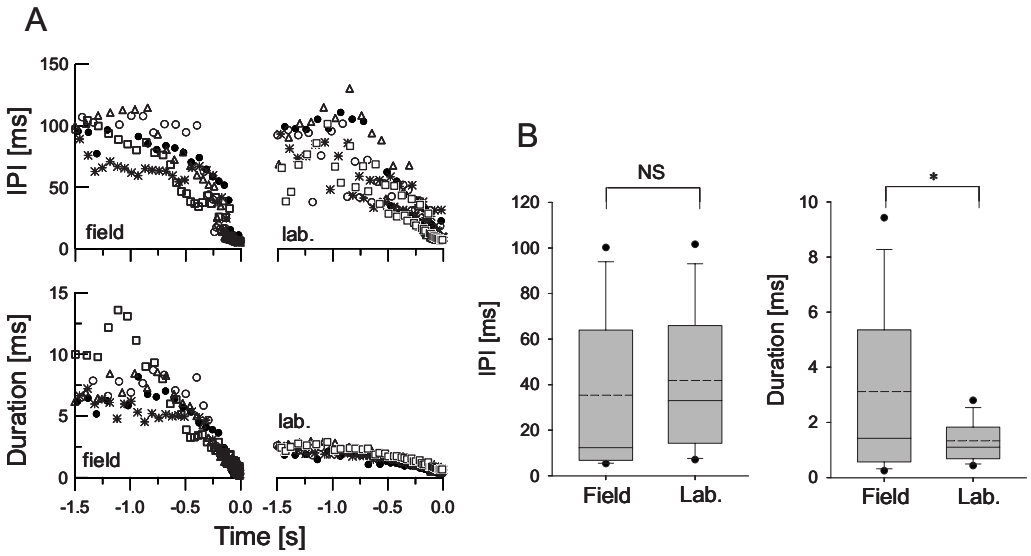


Fig. 2. Comparison of call parameters during approaching behaviors between laboratory and field recordings. (A) Changes in interpulse interval (IPI) and pulse duration as a function of time to landing/capturing. Data were taken from five flight sessions each for field and laboratory recordings. For comparative investigations on laboratory and filed recordings, we only showed the changes in IPI and duration after the approach phase started (1.5 seconds prior to a landing or capture). (B) Box plots of variations of IPI and pulse duration. The IPI during foraging was similar to the IPI of bats exhibiting landing flight in the laboratory, whereas the pulse duration change was significantly different between the two approach behaviors (Kruskal-Wallis; $P < 0.05$).

25 kHz (buzz II). This sudden drop in terminal frequency was never observed in the laboratory recordings for the bats during landing flights.

Decreases in IPI and pulse amplitude usually began at a distance between 1 and 2 m from the wall in the laboratory (Hiryu *et al.*, 2007). We found that the IPI of the foraging bats showed similar changes to those observed in bats approaching the wall in the laboratory, whereas the pulse duration was significantly different between the field and laboratory recordings (Fig. 2). Similar to other pipistrelles, foraging *P. abramus* was found to use highly flexible pulse repertoires in which the bats emphasized sound energy in the low frequencies (terminal frequency of the pulse), and a quasi-CF portion followed the initial FM sweep during the search phase [Fig. 3(A)]. Even though the FM sweep pulse with high-frequency range [i.e., the fundamental frequency of pulse is modulated from 100 to 40 kHz in *P. abramus* (Hiryu *et al.*, 2007)] would be attenuated on the recording by a distant microphone, no long quasi-CF pulses were recorded by the Telemike during landing in the laboratory. The maximum pulse duration was approximately 3 ms in the laboratory; however, as the target distance increased, the duration of the pulse increased with emphasis on the terminal frequency portion at the end of the pulse [Fig. 3(B)].

4. Discussion

Pipistrelles use flexible pulse repertoires in terms of frequency structure during foraging echolocation in the field (Schnitzler *et al.*, 1987; Kalko and Schnitzler, 1993; Boonman and Schnitzler, 2005). We confirmed that wild foraging *P. abramus* emitted long shallow FM pulses (quasi-CF) during the search phase, and the pulse was consistently emphasized at the terminal frequency of 40–43 kHz, which is consistent with other FM bat species [e.g., European *Pipistrellus* bats (Miller and Degn, 1981; Kalko and Schnitzler, 1993; Boonman and Schnitzler, 2005; Kalko, 1995) and *Eptesicus fuscus* (Griffin, 1958; Surlykke and Moss, 2000)]. Given that FM bat species adjust their pulse duration to avoid overlap of the returning echo with the emitted pulse, the distance to the object which the bat is inspecting by echolocation can be estimated

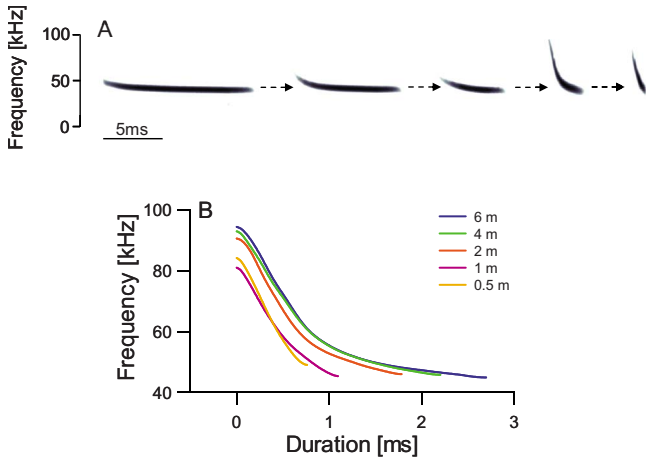


Fig. 3. (Color online) Changes in the frequency structure of echolocation pulses emitted by *P. abramus* in the field and laboratory. (A) Sonograms of representative pulses during echolocation. The bat shortened the pulse duration with decreasing target range (from left to right). Only the first and second pulses from the right were seen from the laboratory recording. (B) Typical changes in FM sweeps of echolocation pulses of the bat during landing flight at various distances from the target wall in the laboratory. The data were from one typical flight.

by the pulse duration (Kalko and Schnitzler, 1993). Our recordings of wild *P. abramus* indicated that the duration of quasi-CF pulses was 9–11 ms [Fig. 1(B)], corresponding to the distance of approximately 1.5–1.9 m. Indeed, we confirmed that the bats intending to land on the wall mesh started to decrease pulse amplitude, duration, and IPI at approximately 1.5–2 m from the target in the laboratory recordings (Hiryu *et al.*, 2007). These findings suggest that the approach distance of the bats during landing in the laboratory may correspond to that of foraging bats in the field, which would explain the lack of significant difference detected between decreases in IPI for the two approach flights (see Fig. 2).

Bates *et al.* (2007) reported that as the background noise level increased during psychological target detection tasks, big brown bats (*E. fuscus*) lengthened the duration emphasizing the end portion of the pulse to concentrate transmitted energy of echolocation pulses in the low frequency range (see Fig. 1 in Bates *et al.*, 2007). Our recordings also show that during landing flights in the laboratory, *P. abramus* emphasized the terminal frequency portion of the pulse as the target distance increased [see Fig. 3(B)]. The body mass of *P. abramus* ranges from 5 to 8 g, and the wingspan measures approximately 10 cm. Given their small body size (e.g., *E. fuscus*, 14–15 g, Bates *et al.*, 2007), the insects on which wild *P. abramus* feed are less than a few millimeters in body length (Hirai and Kimura, 2004). In addition, the sound pressure level of the emitted pulse by pipistrelles is apparently weaker compared to other well-studied FM bat species (personal observations). Thus, we propose that pipistrelles may have established highly flexible pulse repertoires to overcome this disadvantage, thereby enabling them to change the frequency structure of a pulse from a long quasi-CF type used to find a weak echo resulting from an small insect prey (Schnitzler *et al.*, 1987) to a wideband FM type during the approach phase of foraging for precise target ranging (Simmons *et al.*, 1979). Such dynamic and flexible adaptation of sonar signal design is one of the unique strategies by these FM bat species. Comparative investigations on laboratory and field recordings will yield useful information to understand flexible echolocation by biosonar.

Acknowledgments

This work was partly supported by a grant to the Research Center for Advanced Science and Technology (RCAST) at Doshisha University from the Ministry of Education, Culture, Sports,

Science, and Technology (MEXT) of Japan: Special Research Grants for the Development of Characteristic Education from the Promotion and Mutual Aid Corporation for Private Schools of Japan, and the Innovative Cluster Creation Project.

References and links

Bates, M. E., Stamper, S., and Simmons, J. A. (**2007**). "Jamming avoidance response of big brown bats in target detection," J. Exp. Biol. **211**, 106–113.

Boonman, A., and Schnitzler, H. U. (**2005**). "Frequency modulation patterns in the echolocation signals of two vespertilionid bats," J. Comp. Physiol., A **191**, 13–21.

Griffin, D. R. (**1958**). *Listening in the Dark* (Yale University Press, New Haven).

Hirai, T., and Kimura, S. (**2004**). "Diet composition of the common bat *Pipistrellus abramus* (Chiroptera; Vespertilionidae), revealed by fecal analysis," Jpn. J. Ecol. **54**, 159–163 (in Japanese).

Hiryu, S., Hagino, T., Riquimaroux, H., and Watanabe, Y. (**2007**). "Echo-intensity compensation in echolocating bats (*Pipistrellus abramus*) during flight measured by a telemetry microphone," J. Acoust. Soc. Am. **121**, 1749–1757.

Kalko, E. (**1995**). "Insect pursuit, prey capture and echolocation in pipistrelle bats (Microchiroptera)," Anim. Behav. **50**, 861–880.

Kalko, E., and Schnitzler, H. U. (**1993**). "Plasticity in echolocation signals of European pipistrelle bats in search flight: Implications for habitat use and prey detection," Behav. Ecol. Sociobiol. **33**, 415–428.

Miller, L. A., and Degn, H. J. (**1981**). "The acoustic behavior of four species of vespertilionid bats studied in the field," J. Comp. Physiol. [A] **142**, 67–74.

Neuweiler, G. (**1984**). "Foraging, echolocation and audition in bats," Naturwiss. **71**, 446–455.

Schnitzler, H. U., and Henson, O. W., Jr. (**1980**). "Performance of airborne animal sonar system, I. Microchiroptera.," in *Animal Sonar Systems*, edited by R.-G. Busnel and F. F. James (Plenum, New York), pp. 109–181.

Schnitzler, H. U., Kalko, E., Miller, L., and Surlykke, A. (**1987**). "The echolocation and hunting behavior of the bat, *Pipistrellus kuhli*," J. Comp. Physiol., A **161**, 267–274.

Schnitzler, H. U., Moss, C. F., and Denzinger, A. (**2003**). "From spatial orientation to food acquisition in echolocating bats," Trends Ecol. Evol. **18**, 386–394.

Simmons, J. A., Fenton, M. B., and O'Farrell, M. J. (**1979**). "Echolocation and pursuit of prey by bats," Science **203**, 16–21.

Surlykke, A., and Moss, C. F. (**2000**). "Echolocation behavior of big brown bats, *Eptesicus fuscus*, in the field and the laboratory," J. Acoust. Soc. Am. **108**, 2419–2429.

Thomas, J. A., Moss, C. F., and Vater, M. (**2003**). *Echolocation in Bats and Dolphins* (University of Chicago Press, Chicago).

Dispersion relation for air via Kramers-Kronig analysis

Fernando J. Álvarez
Department of Electrical Engineering, Electronics and Automatics, University of Extremadura, 06071 Badajoz, Spain
fafranco@unex.es

Roman Kuc
Department of Electrical Engineering, Yale University, New Haven, Connecticut, 06520
roman.kuc@yale.edu

Abstract: A general expression for the dispersion of acoustic waves in air is obtained by combining the attenuation coefficient given by the ISO:9613-1 standard and the twice-subtracted Kramers-Kronig relation. Good agreement is found with published data of sound velocity at different frequencies and relative humidities. The resulting expression is used to investigate changes in local dispersion with temperature and humidity.

PACS numbers: 43.28.Bj, 43.20.Hq [VO]
Date Received: April 15, 2008 **Date Accepted:** May 17, 2008

1. Introduction

The Kramers-Kronig (K–K) relations link the real and imaginary parts of the frequency response function of any causal and linear system, regardless of its nature. Since their introduction in the 1920s in relation to the absorption of light scattered by atoms, they have been used in many different fields of Physics to build causally consistent models. Ginzberg is credited with being the first one to propose a particular form of the K-K relations for acoustic applications, giving rise to numerous works that have used evolved versions of these expressions to relate attenuation and phase velocity in the measurements of soft tissues,[1] polymers,[2] suspensions,[3,4] and cancellous bone,[5] to give some examples.

Traditionally, the application of these integral relations has involved two problems. First, their nonlocal character requires the knowledge of one of the magnitudes (attenuation or phase velocity) over the entire frequency range in order to obtain the value of the other one at a single frequency. Second, the convergence of the integrals must be ensured. In the last years, much effort has been invested in obtaining new and more useful forms of the K-K relations applicable to a wider class of practical problems.[6] One of these forms, namely the K-K relations with subtractions, has been used in this paper to obtain the dispersion from the attenuation coefficient provided by the ISO:9613-1 standard describing the absorption of sound in the atmosphere.[7]

2. Theory

2.1 Twice-subtracted K-K relation

The transfer function of a passive and linear acoustic system can be written as $H(\omega)=e^{iK(\omega)\cdot l}$, where ω is angular frequency, l represents the length of the system in the direction of interest and $K(\omega)=\omega/c(\omega)+i\alpha(\omega)$ is the complex wave number, expressed in terms of the phase velocity $c(\omega)$ and the attenuation coefficient $\alpha(\omega)$. If $K(\omega)$ diverges as the nth power of the frequency or slower, then the absorption coefficient and the phase velocity can be related by applying Tichmarsh's theorem[8] to the function $R_n(\omega)$, defined as the remainder term in the $(n-1)$th order McLaurin expansion of the complex wave number, i.e.

$$K(\omega) = \sum_{m=0}^{n-1} \frac{1}{m!} \omega^m \frac{d^m}{d\omega^m} K(\omega)\big|_{\omega=0} + \frac{1}{n!} \omega^n R_n(\omega). \tag{1}$$

The complex variable function $R_n(z=\omega+iy)$ is square integrable over the real ω axis, and also along every line parallel to the real axis where it is analytic. If the system is causal, its transfer function $H(z)$ is analytic in the upper half of the complex plane.[9] The complex wave number $K(z)=\ln H(z)/i\cdot l$ is also analytic in this upper half, since $H(z)$ does not have zeros in this region.[1] It is clear from Eq. (1) that $R_n(z)$ is analytic everywhere $K(z)$ is, thus satisfying the second condition of Tichmarsh's theorem. This theorem states that under these conditions, the real and imaginary parts of $R_n(\omega)$ must be Hilbert transforms of each another

$$\mathrm{Re}\{R_n(\omega)\} = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\mathrm{Im}\{R_n(\omega')\}}{\omega' - \omega} d\omega', \tag{2}$$

$$\mathrm{Im}\{R_n(\omega)\} = -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\mathrm{Re}\{R_n(\omega')\}}{\omega' - \omega} d\omega', \tag{3}$$

where $\mathcal{P}$ denotes the Cauchy principal value of the integral. For the particular case in which $K(\omega) \sim \omega^2$—as it is in air—the R_2 function is given by

$$R_2(\omega) = \frac{2}{\omega^2}\left[K(\omega) - K(0) - \omega\left(\frac{d}{d\omega} K(\omega)\big|_{\omega=0}\right)\right] = \frac{2}{\omega^2}\left[\frac{\omega}{c(\omega)} - \frac{\omega}{c(0)} + i\alpha(\omega)\right]. \tag{4}$$

Using Eqs. (2) and (4) we have

$$\frac{1}{\omega}\left[\frac{1}{c(\omega)} - \frac{1}{c(0)}\right] = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha(\omega')/\omega'^2}{\omega' - \omega} d\omega'. \tag{5}$$

Since the expression $\mathcal{P}\int_{-\infty}^{\infty} d\omega'/\omega' - \omega$ equals 0, any constant can be subtracted from the numerator of the integrand above. In particular, the value of the numerator at the frequency of interest ω can be subtracted to obtain:

$$\frac{1}{\omega}\left[\frac{1}{c(\omega)} - \frac{1}{c(0)}\right] = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha(\omega')/\omega'^2 - \alpha(\omega)/\omega^2}{\omega' - \omega} d\omega'. \tag{6}$$

Finally, noting that $\mathcal{P}\int_{-\infty}^{\infty} G(\omega')d\omega' = \lim_{\epsilon\to 0}[\int_{\epsilon}^{\infty} G(\omega')d\omega' + \int_{\epsilon}^{\infty} G(-\omega')d\omega']$, this expression can be simplified to obtain the twice-subtracted K-K relation for the dispersion:

$$\frac{1}{c(\omega)} - \frac{1}{c(0)} = \frac{2\omega^2}{\pi} \int_0^{\infty} \left[\frac{\alpha(\omega')}{\omega'^2} - \frac{\alpha(\omega)}{\omega^2}\right] \frac{d\omega'}{\omega'^2 - \omega^2}. \tag{7}$$

2.2 Dispersion relation in air

Air is one of the polyatomic gases more deeply studied, and currently it is known that there are basically two different mechanisms that cause absorption of acoustic waves: the viscothermal losses (or classical absorption) and the oxygen and nitrogen molecular relaxation processes. The theoretical analysis of both processes led to a set of equations that have been later experimentally adjusted to increase agreement with real data. Today, these equations are grouped into the ISO-9613 standard.[7] This norm establishes that the absorption coefficient can be calculated as

$$\alpha(f) = f^2 \left\{ 1.84 \cdot 10^{-11} \left(\frac{P}{P_{\text{ref}}} \right)^{-1} \cdot \left(\frac{T}{T_{\text{ref}}} \right)^{\frac{1}{2}} + \left(\frac{T}{T_{\text{ref}}} \right)^{\frac{-5}{2}} \cdot \left[0.01275 \frac{e^{\frac{-2239.1}{T}}}{f_{rO} + \frac{f^2}{f_{rO}}} + 0.1068 \frac{e^{\frac{-3352}{T}}}{f_{rN} + \frac{f^2}{f_{rN}}} \right] \right\} (\text{Np/m}), \tag{8}$$

where f is the sound frequency in Hz, P is the atmospheric pressure in Pa (P_{ref}=101 325 Pa), T is the absolute temperature in K (T_{ref}=293.15 K), and f_{rO}, f_{rN} represent the relaxation frequencies of oxygen and nitrogen, respectively. As can be seen from the expression above, α dependence on frequency is approximately quadratic except in the proximities of f_{rO} and f_{rN}. These frequencies are strongly dependent functions on temperature and humidity, and they determine the changes in the dispersion with these magnitudes.

Introducing Eq. (8) into the twice-subtracted K-K relation for the dispersion Eq. (7) and performing the integration, yields

$$\frac{1}{c(\omega)} - \frac{1}{c(0)} = \frac{-\omega^2}{2\pi} \left(\frac{T}{T_{\text{ref}}} \right)^{\frac{-5}{2}} \cdot \left[0.01275 \frac{e^{\frac{-2239.1}{T}}}{\omega^2 + \omega_{rO}^2} + 0.1068 \frac{e^{\frac{-3352}{T}}}{\omega^2 + \omega_{rN}^2} \right], \tag{9}$$

where ω_{rO}, ω_{rN} are the angular counterparts of f_{rO} and f_{rN}, respectively. From this expression, the dispersion with respect to any arbitrary frequency ω_0 can be easily obtained as

$$\frac{1}{c(\omega)} - \frac{1}{c(\omega_0)} = \frac{\omega_0^2 - \omega^2}{2\pi} \left(\frac{T}{T_{\text{ref}}} \right)^{\frac{-5}{2}} \cdot \left\{ 0.01275 \frac{\omega_{rO}^2 \cdot e^{\frac{-2239.1}{T}}}{(\omega^2 + \omega_{rO}^2)(\omega_0^2 + \omega_{rO}^2)} + 0.1068 \frac{\omega_{rN}^2 \cdot e^{\frac{-3352}{T}}}{(\omega^2 + \omega_{rN}^2)(\omega_0^2 + \omega_{rN}^2)} \right\}. \tag{10}$$

This equation is completely general, and can be used to obtain the dispersion of any acoustic signal in air provided the expression for the attenuation coefficient given by the ISO standard remains valid.

3. Results

The theoretical values of dispersion predicted by Eq. (10) have been compared with the data of sound velocity obtained by Evans and Bass.[10] This extensive compilation provides calculations in still air at 20 °C, relative humidities from 0% to 100%, and the range of frequencies 12 Hz–1 MHz. Figure 1 shows this comparison for three different values of relative humidity: 0%, 50% and 100%. As can be seen, Eq. (10) predicts the tendency of the data in the three cases, and this is also true for the rest of values reported in Ref. 10 under different conditions of humidity. Moreover, the predictions are fairly accurate in all cases except when these humidities are very low. In their work, Evans and Bass admitted that, for extremely dry air ($H<10\%$), there was some difficulty in assessing the accuracy of the calculations below 100 Hz, a fact that could explain the discrepancies observed in this range of humidities. For example, a small increment of 30 parts per million in the sound velocity calculated for H=0% and f=12 Hz, would make the differences seen in Fig. 1(a) for frequencies above 31.5 kHz negligible.

Equation (10) can be also used to investigate dispersion variations with temperature and pressure (the latter magnitude does not explicitly appear in the equation, but it determines the values of the relaxation frequencies). Unfortunately, to the authors' knowledge, there are no

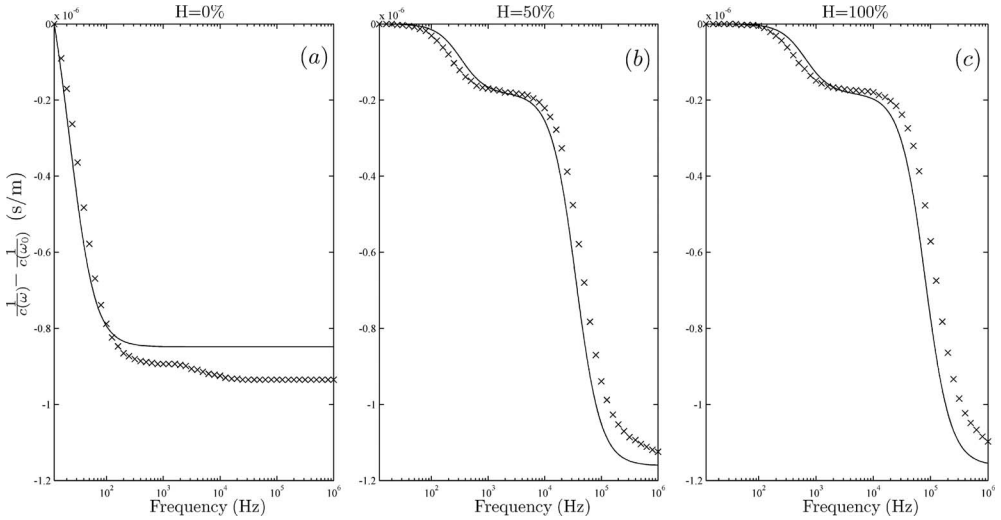


Fig. 1. Comparison between the dispersion predicted by Eq. (10) (solid line) and the data extracted from Evans and Bass (Ref. 10) (crosses) for different humidities, $T=20\ ^\circ\mathrm{C}$ and $P=101\ 325$ Pa. The reference frequency $f_0=12$ Hz is the lowest value reported in Ref. 10.

accurate published calculations of sound velocity in air as a function of frequency for different temperatures or pressures against which to compare the predictions of this expression.

Particularly interesting is the study of the frequency derivative of Eq. (9) $d/d\omega\, 1/c(\omega)$, showing how the local dispersion changes with temperature and humidity. Figure 2 shows the magnitudes for three different humidities (0, 50 and 100%) and four different temperatures (−20, 0, 20 and 40 °C). Note that different scales were used in Figs. 2(b) and 2(c) in comparison with Fig. 2(a). Figure 2(a) shows the results in dry air, when the relaxation frequencies of nitrogen (f_{rN}) and oxygen (f_{rO}) remain essentially constant with temperature. The

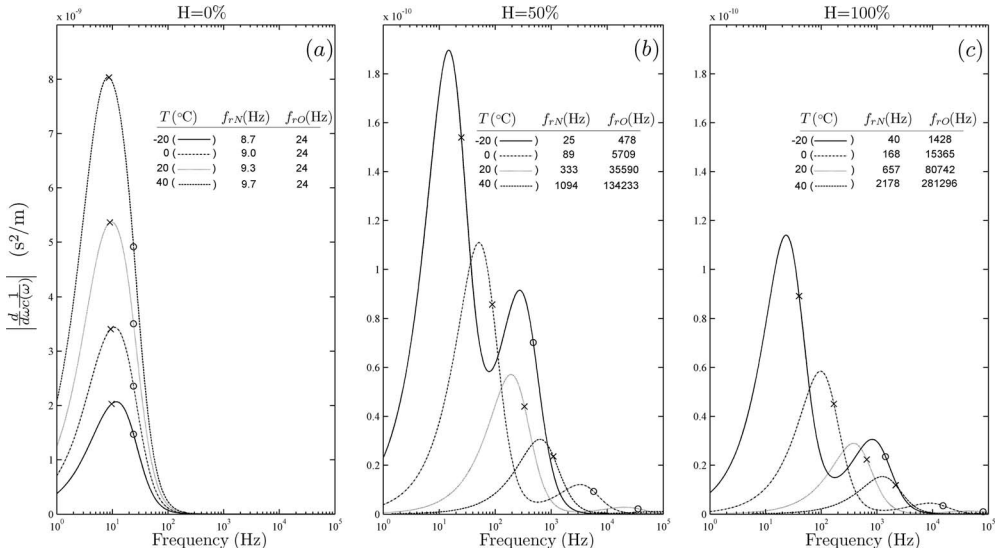


Fig. 2. Local dispersion dependence on temperature and humidity. The symbols X and O represent the values of f_{rN} and f_{rO} in each case, respectively. All the curves were calculated assuming $P=101\ 325$ Pa.

values of these frequencies have been plotted in all cases over the corresponding curve, with the symbols "X" and "O" representing f_{rN} and f_{rO}, respectively. As can be seen, in dry air the maximum local dispersion always occurs around f_{rN} and increases in magnitude with rising temperature. In this case the maximum values of local dispersion are obtained.

When the relative humidity is different from 0, f_{rN} and f_{rO} are increasing functions of temperature and, opposite to the results in dry air, the maximum dispersions occur at the lower temperatures. In this case, two maxima of local dispersion can be clearly identified just below the corresponding relaxation frequencies, being the maximum associated with f_{rO} of lower magnitude. With increasing temperature, the local maxima move toward higher frequencies following the behavior of f_{rN} and f_{rO}, and decrease in magnitude. A similar trend occurs with increasing humidity, as can be deduced from the comparison between Figs. 2(b) and 2(c).

4. Summary

A general expression for the dispersion of acoustic waves in air has been obtained following arguments of linearity and causality. The predictions of this expression have been compared with published data of sound velocity and agree in all cases except in extremely dry air, a fact that could be attributed to the low accuracy of these data for relative humidities below 10% and frequencies lower than 100 Hz. The theoretical trends of these data are predicted accurately.

The obtained expression has been used to derive changes in the local dispersion with temperature and relative humidity. It has been shown that in dry air the dispersive region remains invariant with temperature and centered around the relaxation frequency of nitrogen, although the magnitude of this dispersion increases with rising temperatures. In humid air, the dispersive region moves toward higher frequencies and decreases in magnitude with rising temperatures, showing a maximum value just before the relaxation frequency of nitrogen.

Acknowledgments

This work was performed while F. J. Álvarez was on a research sabbatical in the Intelligent Sensors Laboratory at Yale University. This stay was financed by the Spanish Ministry of Science and Education through the "Jose Castillejo" program.

References and links

[1]The system would be a perfect reflector otherwise.

[1]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**(3), 696–701 (1981).
[2]H. J. Wintle, "Kramers-Kronig analysis of polymer acoustic data," J. Appl. Phys. **85**, 44–48 (1999).
[3]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**(5), 2091–2106 (2000).
[4]J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Erratum: Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **112**(2), 760–761 (2002).
[5]K. R. Waters, B. K. Hoffmeister, and J. A. Javarone, "Application of the Kramers-Kronig relations to measurements of attenuation and dispersion in cancellous bone," in *IEEE Ultrasonic Symposium* (UFFC Society, Montreal, 2004), Vol. **1**, pp. 561–564.
[6]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**(5), 822–833 (2005).
[7]ISO/TC 43 Technical Comittee, Acoustics, Sub-Comittee SCI, Noise, "Attenuation of sound during propagation outdoors. Part 1: Calculation of the absorption of sound by the atmosphere," Technical Report No. ISO 9613-1:1993(E), International Organization for Standarization, Geneve, Switzerland.
[8]E. C. Tichmarsh, *Introduction to the Theory of Fourier Integrals*, 3rd ed. (Chelsea, New York, 1986).
[9]R. L. Weaver and Y.-H. Pao, "Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media," J. Math. Phys. **22**(9), 1909–1918 (1981).
[10]L. B. Evans and H. E. Bass, "Tables of absorption and velocity of sound in still air at 68F (20 °C)," Technical Report No. AD0738576, National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161.

ACOUSTICAL NEWS—USA

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

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

Preliminary Notice: 156th Meeting of the Acoustical Society of America

The 156th Meeting of the Acoustical Society of America will be held Monday through Friday, 10–14 November 2008 at the Doral Golf Resort and Spa, Miami, Florida, USA. A block of rooms has been reserved at the Doral Golf Resort and Spa.

Information about the meeting also appears on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩.

Charles E. Schmid
Executive Director

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 10–14 November.

Special Sessions

Acoustical Oceanography (AO)

Attenuation coefficient of sediments from low- to mid-frequencies
(Joint with Underwater Acoustics)
There is much debate about the slope of the frequency dependance of the low- to mid-frequency sediment attenuation coefficient
Three-dimensional acoustics and inversions on the Continental Shelf and canyons
(Joint with Underwater Acoustics)
Complex oceanography, geology, and acoustic propagation for the Continental shelf and the canyons that cross it

Animal Bioacoustics (AB)

Acoustics of manatees and alligators
Bioacoustics of manatees: hearing, behavior, sound communication, and considerations for the avoidance of potentially dangerous sound sources
Marine mammal acoustics in honor of Sam Ridgway
Acoustics of marine mammals, reflecting the wide interests and influence of Sam Ridgway

Architectural Acoustics (AA)

Acoustic challenges of implementing curtain wall construction
Discussion of test results and potentially successful methods for improving the acoustic isolation of curtain wall systems
Acoustical issues of Green buildings
(Joint with Noise)
Existing performance data and issues and challenges of Green building acoustical design
Acoustics of retrofitted performance spaces
Case studies/acoustical challenges of spaces retrofitted for performance
Acoustics of single family residences
Explore the different sound characteristics in today's single family residences
Acoustics of small, multipurpose performance spaces
Theory and experience related to acoustics of small, multipurpose performance spaces
Classroom acoustics in honor of Michael Nixon
(Joint with Noise and Speech Communication)
Ongoing developments in classroom acoustics, honoring a recently deceased pioneer
Fostering productive architect-acoustician relationships
Professional development discussion about collaboration between architects and acousticians and how it works best in the design process
Innovative integration of acoustic treatment into modern architecture
Discussion of unique ways to integrate acoustic treatments in the current architectural trend of using glass, steel, and concrete
Special session celebrating the work of Russell Johnson
Russell Johnson's contributions to architectural acoustics at Bolt, Beranek and Newman and Artec

Engineering Acoustics (EA)

Acoustics for battlefield operations and homeland security
Sensors, systems, and methods to enhance mission performance in tactical environments and ensure persistent surveillance
High precision acoustical measurements
(Joint with ASA Committee on Standards and Biomedical Ultrasound/Bioresponse to Vibration)
Techniques and applications for highly accurate or precise acoustical measurements used in metrology, quality assurance, research, and development

Education in Acoustics (ED)

Hands-on experiments for high school students
Experiments for high school students
"Project Listen Up"
(Joint with ASA Student Council)
Descriptions of acoustic demonstrations, laboratory experiments or discovery activities for learners of all ages. Apparatus may be shown but the talks should focus on concepts, explanations, diagrams, and drawings with an emphasis on careful scientific approach

Musical Acoustics (MU)

Dynamical approaches in the study of music perception and performance
Dynamical models of and empirical studies on time-varying processes in the perception and production of musical events and sequences
Statistical approaches for analysis of music and speech audio signals
(Joint with Speech Communication and Signal Processing in Acoustics)
Exploration of recent advances in data-driven approaches which provide promising alternatives to traditional methods for extracting structure and information from audio
Structural vibrations in musical instruments
(Joint with Structural Acoustics and Vibration)
Holographic interferometry and modal analysis as experimental tools and finite element analysis as a theoretical tool to study instrumental vibrations
Telematic music technology
New technological developments and requirements for internet-based music collaborations utilizing broad bandwidth and new telepresence applications

Noise (NS)

Advances in measurement and noise and noise effects on humans and nonhuman animals in the environment
(Joint with Animal Bioacoustics)
Methodologies and metrics for quantifying noise in the environment to quantify effects
Fire codes and acoustics
(Joint with Architectural Acoustics)
Impact of fire codes on acoustical design
Noise control and acoustics of marine vessels
Methods and approaches which address the unique structures and environments of marine interiors and equipment
Sound levels and acoustical characteristics of modular classrooms
(Joint with Architectural Acoustics)
Modular classroom within the concept of classroom acoustics

 0001-4966/2008/124(2)/689/5/$23.00

Soundscape and sound quality—Measurement and lexicon
(Joint with Architectural Acoustics and ASA Committee on Standards)
Interdisciplinarity in soundscape approaches to integrate different disciplines
Special issues for classroom acoustics in tropical environments
(Joint with Architectural Acoustics)
Requirements through the tropical environments-research on specific needs

Signal Processing in Acoustics (SP)

Autonomous system acoustic sensors and processors
(Joint with Underwater Acoustics and Engineering Acoustics)
Acoustic sensors and signal processors for the development of fully autonomous systems in unmanned vehicles for detection, localization, tracking, classification, or obstacle avoidance
Recent developments in coded signals in acoustics
(Joint with Underwater Acoustics, Architectural Acoustics and Biomedical Ultrasound/Bioresponse to Vibration)
Overview of improvements in the theory and application of coded signals to architectural, audio, biomedical, environmental, remote sensing, and underwater acoustics
Signal processing for high clutter environments
(Joint with Underwater Acoustics)
Signal processing methods used in the presence of high clutter such as dense shipping lanes, harbors, complex machinery, and battlefields using both single sensors and distributed networks

Speech Communication (SC)

A quantal transition: Ken Stevens in "retirement"
Honoring Ken Stevens upon his retirement from teaching
James J. Jenkins: Teacher, mentor, researcher
Honoring Dr. Jenkins' contributions to speech acoustics and perception through training and mentoring many ASA members as well as his own research

Structural Acoustics and Vibration (SA)

Aeroacoustic and hydrodynamic interactions with structural acoustics and vibrations
Impacts of fluid loading and fluid-structure interactions on structural acoustics and vibrations
Ambient noise correlations in structural acoustics and vibrations
Impacts of ambient and environmental noise on structural acoustics and vibrations
Building structural acoustics and vibrations
Transmissions of sound and vibrations through building structures
Causalities in structural acoustics and vibrations
Vibrations can be excited by sound waves. Conversely, vibrations can generate sound. The interrelationship between sound and vibrations are studied

Underwater Acoustics (UW)

Acoustics of harbors, ports, and shallow navigable waterways
(Joint with Acoustical Oceanography)
Topics and applications relating to acoustic propagation in, or acoustic characterization of, near shore waterways. Relevant topics include acoustic applications for harbor defense, low-frequency acoustic propagation including tidal effects and temporal water column variability, and ambient noise levels in harbors
Robust array processing
(Joint with Signal Processing in Acoustics)
Techniques to reduce the sensitivity of array processors to environmental and array element mismatch

Other Technical Events

Hot Topics

A "Hot Topics" session sponsored by the Tutorials Committee will cover the fields of Engineering Acoustics, Physical Acoustics, and Underwater Acoustics.

The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

Barry Blesser, author of the book *Spaces Speak, Are You Listening*? *Experiencing Aural Architecture* will discuss the topic of his research.

Exhibit

The meeting will be highlighted by an exhibit which will feature displays with instruments, materials, and services for the acoustical and vibration community. The exhibit, which will be conveniently located near the registration area and meeting rooms, will open at the Doral with a reception on Monday evening, 10 November, and will close Wednesday, 12 November, at noon. Morning and afternoon refreshments will be available in the exhibit area.

The exhibit will include computer-based instrumentation, sound level meters, sound intensity systems, signal processing systems, devices for noise control, sound prediction software, acoustical materials, passive and active noise control systems, and other exhibits on vibrations and acoustics. For further information, please contact: Robert Finnegan, American Inst. of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747; (516) 576-2433; rfinneg@aip.org.

Online Meeting Papers

The ASA provides the "Meeting Papers Online" website where authors of papers to be presented at meetings will be able to post their full papers or presentation materials for others who are interested in obtaining detailed information about meeting presentations. The online site will be open for author submissions in September. Submission procedures and password information will be mailed to authors with the acceptance notices.

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

The url is ⟨http://scitation.aip.org/asameetingpapers/⟩.

Proceedings of Meetings on Acoustics (POMA)

The upcoming meeting of the Acoustical Society of America will have a published proceedings, and submission is optional. The proceedings will be a separate volume of the online journal, "Proceedings of Meetings on Acoustics" (POMA). This is an open access journal, so that its articles are available in pdf format without charge to anyone in the world for downloading. Authors who are scheduled to present papers at the meeting are encouraged to prepare a suitable version in pdf format that will appear in POMA. The format requirements for POMA are somewhat more stringent than for posting on the ASA Online Meetings Papers Site, but the two versions could be the same. The posting at the Online Meetings Papers site, however, is not archival, and posted papers will be taken down six months after the meeting. The POMA online site for submission of papers from the meeting will be opened at the same time when authors are notified that their papers have been accepted for presentation. It is not necessary to wait until after the meeting to submit one's paper to POMA. Further information regarding POMA can be found at the site http://asa.aip.org/poma.html. Published papers from previous meetings can be seen at the site http://scitation.aip.org/POMA.

Meeting Program

An advance meeting program summary will be published in the September issue of JASA and a complete meeting program will be mailed as Part 2 of the October issue. Abstracts will be available on the ASA Home Page ⟨http://asa.aip.org⟩ in October.

Tutorial Lecture

A tutorial presentation on "Aircraft Noise Prediction" will be given by Joe Posey of NASA Langley Research Center on Monday, 10 November, at 7:00 p.m.

This tutorial lecture will consider noise from present and future subsonic jet aircraft, rotorcraft, propeller aircraft, and supersonic transports. Noise prediction capabilities will be identified, along with an overview of noise control technology. Acousticians are preparing to predict and control community and interior noise for arbitrary configurations using models based more on first principles, whereas the state of the art is largely semi-empirical. It is imperative that acousticians be included in the early stages of the design process and for all design team members to have some exposure to noise control principles to avoid wasting resources on designs that should be nonstarters from the noise perspective.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15.00 USD for registration received by 20 October and $25.00 USD at the meeting. The fee for students with current ID cards is $7.00

USD for registration received by 20 October and $12.00 USD at the meeting. To register, use the registration form in the printed call for papers or register online at ⟨http://asa.aip.org⟩.

Short Course

Ultrasonic nondestructive evaluation (NDE) is the central topic of this short course. NDE is the process of using a form of energy to interrogate industrial parts or components for the purpose of assessing their fitness for service or manufacture.

This short course provides an introduction to the methods of ultrasonic NDE and analytical modeling of ultrasonic scattering at planar interfaces.

The instructor, Dale E. Chimenti, is Professor of Aerospace Engineering and Senior Scientist in the Center for Nondestructive Evaluation at Iowa State University, Ames IA. He has published nearly 200 scientific and technical papers, over 75 in the archival journal literature. He has held faculty and research positions both at Iowa State and at the Center for NDE at The Johns Hopkins University in Baltimore, Maryland. Dr. Chimenti serves as Editor-in-Chief of NDT&E International and served as Associate Editor of JASA from 1998 to 2004. He is also co-editor of the Review of Progress in Quantitative Nondestructive Evaluation published by AIP and now in its 27th volume. In 2007 he was selected to receive Iowa State's Eminent Faculty Research Award, and he is a Fellow of ASA.

The course schedule is Sunday, 9 November 2008, 1:00 to 5:00 p.m. and Monday, 10 November 2008, 8:30 a.m. to 12:30 p.m.

The registration fee is $250.00 USD and covers attendance, instructional materials and coffee breaks. **The number of attendees will be limited so please register early to avoid disappointment**. Only those who have registered by 20 October will be guaranteed receipt of instructional materials. There will be a $50.00 USD discount for registration made prior to 20 October. Full refunds will be made for cancellations prior to 20 October. Any cancellation after 20 October will be charged a $25.00 USD processing fee. To register, use the form in the printed call for paprs or register online at ⟨http://asa.aip.org⟩.

Special Meeting Features

Student Transportation Subsidies

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

Young Investigator Travel Grant

The Committee on Women in Acoustics (WIA) is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Miami meeting. Young professionals who have completed their doctorate in the past five years are eligible to apply if they plan to present a paper at the Miami meeting, are not currently students, and have not previously received the award. Each award will be of the order of $300 with three awards anticipated. Awards will be presented by check at the WIA luncheon at the meeting. Both men and women may apply. Applicants should submit a request for support, a copy of the abstract for their presentation at the meeting, and a current resume/vita which includes information on their involvement in the field of acoustics and in the ASA. Submission by e-mail is preferred to Dr. Sarah Hargus Ferguson at safergus@ku.edu. Deadline for receipt of applications is 30 September.

Students Meet Members for Lunch

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

Plenary Session, Awards Ceremony, Fellows Luncheon and Social Events

Buffet socials with cash bar will be held on Tuesday and Thursday evenings.

The ASA Plenary session will be held on Wednesday afternoon, 12 November, at the Doral Hotel where Society awards will be presented and recognition of newly elected Fellows will be announced.

A Fellows Luncheon will be held on Thursday, 13 November, at 12:00 noon at the Doral. This luncheon is open to all attendees and their guests. To register, use the form in the printed call for papers or register online at ⟨http://asa.aip.org⟩.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 12 November. Those who wish to attend this luncheon must register using the form in the printed call for papers or online at ⟨http://asa.aip.org⟩. The fee is $15 (students $5) for preregistration by 20 October and $20 (students $5) at the meeting.

Transportation and Hotel Accommodations

Air Transportation

The Miami International Airport, (Airport Code MIA) is served by the following airlines: Air Canada, AirTran, American Airlines, America West Airlines, Continental Airlines, Delta Airlines, Northwest Airlines, Southwest Airlines, Spirit, United Airlines, and U.S. Airways. For further information see ⟨http://www.miami-airport.com/⟩.

Ground Transportation

The Doral Golf Resort and Spa is located approximately 7 miles from Miami International Airport. Traffic in Miami can be congested at any time of day, but it is particularly slow on weekdays westbound from the airport from about 4:00 p.m. to 7:00 p.m., and eastbound from the hotel from about 7:00 a.m. to 10:00 a.m. Try to avoid these times, or plan on delays.

•**Taxicabs:** A taxi from Miami International Airport to the Doral costs approximately $32.00. There may be additional charges for excess baggage, fuel surcharge, and airport fees. Taxicabs are available on the Arrival/Ground Level of the Terminals, outside of the baggage claim area.

•**Airport Shuttle:** A continuously operating shuttle service is available from Miami International Airport to the Doral for $21.00 per person, one way. There may be additional charges for fuel surcharge and airport fees. Call Super Shuttle at 305-871-2000 for more details or to make a reservation. Super Shuttle vans are available on the Arrival/Ground Level of the Terminals, outside of the baggage claim area.

•**Automobile Rental:** Miami International Airport is served by all major car rental companies. Select rental car companies have registration counters located inside the baggage claim area, on the Arrival/Ground Level of the North and Central Terminal. South Terminal uses the rental car phone center located at the Information and Reservation Boards located on the Arrival/Ground Level and on the 3rd floor. Actual Vehicle pick-up for all companies is not on airport property and is located at various off-site areas which must be accessed by courtesy shuttle vans for each separate rental car company. The airport website suggests stepping outside of the baggage claim area on the median curb on the Arrival/Ground Level and flagging-down your selected company's vehicle as you see them pass by. Miami International Airport does not have any rental car return locations at the terminal facilities. When returning the rental car, please follow the posted signs to the specific company. Be sure to allow at least a half hour for shuttle to the airport. If your company's sign is not posted, contact the rental agency for directions. For a list of rental car companies, visit ⟨http://www.miami-airport.com/html/car_rental_list.html.⟩ Alternatively, Enterprise Car Rental is

available in the main lobby of the Doral Hotel. Call extension #6667 for more information or to make a reservation. Hours of operation are: Monday through Friday, 7:30 a.m.—6:00 p.m.; Saturday, 8:00 a.m.—4:00 p.m.; and Sunday, 9:00 a.m.—2:00 p.m.

Parking at the Doral

The Doral Resort offers complimentary on-site parking. Valet parking is also available for $22.00 per day.

Hotel Accommodations and Reservations

The meeting will be held at the Doral Golf Resort and Spa ⟨http://www.doralresort.com⟩ in Miami, Florida. The Resort features ten separate lodges with approximately 700 rooms and suites built on 650 acres in northwest Miami. The property includes five championship 18-hole courses, including the prestigious Blue Monster Course, host of the PGA Tour for over 40 years. For attendees interested in golf, please visit the hotel website. ASA has blocked a limited number of rooms so early reservations are advised.

The European inspired Spa at Doral combines traditionally elegant ambience with a thoroughly modern selection of over 100 spa treatments, along with a state-of-the-art fitness center. In addition to swimming, the Blue Lagoon features waterfalls and a 150 foot waterslide. Other amenities include a driving range, a golf school, Camp Doral (children ages five to 12), boutique shops, and barber/beauty shops. The Doral has five restaurants, including Terrazza Restaurant and Cafe, Atrium Restaurant, Bungalov's Bar & Grill, Champion's Grill & Bar, and Windows, overlooking the Blue Monster golf course. The resort also offers valet parking, complimentary self-parking, concierge service, currency exchange, valet laundry service, and room service. All rooms feature high-speed Internet access, work desks, safes, and free weekday newspapers.

Please make your reservation directly with the Doral Golf Resort and Spa. When making your reservation, you must mention the Acoustical Society of America to obtain the special ASA meeting rates. Alternatively, reservations can be made directly online at the site listed below, which has been set up specifically for the Acoustical Society of America, and has the conference rates and all applicable information incorporated into it.

Doral Golf Resort and Spa
4400 NW 87th Avenue
Miami, FL 33178-2192
Tel.: 305-592-2000; Toll Free: 1-800-228-9290
FAX: (305) 591-6653
Online: http://marriott.com/miadl?groupCode=asaasaa&app=resvlink
Rates (excluding taxes, current 13%)
Single/Double: $189.00 USD

Room Sharing

ASA will compile a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, not later than 1 October to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Acoustical Society of America, Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

Weather in Miami is generally mild in November with average temperatures between about 68 to 80 degrees F. Colder temperatures (45–60 degrees F) may accompany passing cold fronts. November is a relatively dry month, averaging only a few inches of rain, but it can rain in Miami at any time. Locally heavy showers and thunderstorms often accompany the passage of cold fronts. Recent observations and forecasts may be found on a number of different web pages (e.g., www.weather.com).

Assistive Listening Devices

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

Child Care

The Women in Acoustics Committee of the ASA is helping to organize on-site child care services for the Miami meeting and evaluate interest in options for future meetings. Members interested in child care services should contact Andone Lavery member of the ASA Committee on Women in Acoustics, at alavery@whoi.edu as early as possible to indicate interest. On-site child care for children under the age of 5 will be arranged through the Resort Concierge (305-592-2000). For children ages 5–12, the Resort offers Camp Doral, which offers activities that include swimming, arts and crafts, fishing, Tug-of-War, Karaoke, games by the pool, basketball, volleyball, soccer, kickball, indoor games, sandcastle contest, golf putting contest, swimming relays, and poolside bingo.

Daily Camp Hours are as follows (additional hours may be accommodated):

Regular Camp 9:00 a.m.–5:00 p.m.

Half-Day Camp 9:00 a.m.–12:00 p.m./1:00 p.m.–5:00 p.m.

For further information on Camp Doral, call (305) 592-2000 ext. 6329. Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ or email alavery@whoi.edu for updates about child care.

Accompanying Persons Program

Spouses and other visitors are welcome at the Miami meeting. The registration fee for accompanying persons is $50.00 for preregistration by 20 October and $75.00 at the meeting. A hospitality room for accompanying persons will be open at the Doral Golf Resort and Spa from 8:00 a.m. to 11:00 a.m., Monday through Friday. Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ for updates about the accompanying persons program.

Registration Information

The registration desk at the meeting will open on Monday, 10 November, at the Doral Golf Resort and Spa. To register use the form on page 17 or register online at ⟨http://asa.aip.org⟩. **If your registration is not received at the ASA headquarters by 20 October you must register on-site.**

Registration fees are as follows:

Category	Preregistration by 20 October	Onsite20 Registration
Acoustical Society Members	$350	$425
Acoustical Society Members One-Day Attendance*	$175	$215
Nonmembers	$400	$475
Nonmembers One-Day Attendance*	$200	$240
Nonmember Invited Speakers One-Day Attendance*	Fee waived	Fee waived
Nonmember Invited Speakers (Includes one-year ASA membership upon completion of an application)	$110	$110
ASA Early Career Associate or Full Members (For ASA members who transferred from ASA student member status in 2006, 2007, or 2008)	$175	$215
ASA Student Members (with current ID cards)	Fee waived	$25
Nonmember Students (with current ID cards)	$45	$55
Emeritus members of ASA (Emeritus status pre-approved by ASA)	$50	$75
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$50	$75

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2009) of membership. Invited speakers who

are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** may register for one-day only without charge. A nonmember invited speaker who pays the full-week registration fee will be given one free year of membership upon completion of an ASA application form.

Note: A $25 fee will be charged to those who wish to cancel their registrations after 20 October.

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.

2008

10–14 Nov — 156th Meeting of the Acoustical Society of America, Miami, FL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

2009

18–22 May — 157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

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.

August 2008

25–28	**1st International Conference on Water Side Security**, Lyngby, Denmark (www.wss2008.org).
25–29	**10th International Conference on Music Perception and Cognition (ICMPC 10)**, Sapporo, Japan (icmpc10.typepad.jp).

September 2008

8–12	**International Symposium on Underwater Reverberation and Clutter**, Lerici, Italy (isurc2008.org).
9–11	**6th International Ssymposium on Ultrasonic Doppler Methods for Fluid Mechanics and Fluid Engineering**, Prague, Czech Republic (isud6.fsv.cvut.cz).
10–12	**Autumn Meeting of the Acoustical Society of Japan**, Fukuoka, Japan (ww.asj.gr.jp/index-en.html).
15–17	**International Conference on Noise and Vibration Engineering (ISMA2008)**, Leuven, Belgium (www.isma-isaac.be).
22–26	**INTERSPEECH 2008-10th ICSLP**, Brisbane, Australia (www.interspeech2008.org).

October 2008

3–5	**7th International conference onAuditorium Acoustics**, Oslo, Norway (Web:ioa.org.uk).
6–8	**Acoustics Week in Canada**, Vancouver, B.C., Canada (www.caa-aca.ca/vancouver2008)
14–15	*__Underwater Noise Measurement, Impact and Mitigation__, Southampton, UK (www.ioa.org.uk/viewupcoming.asp).
21–22	**Insititute of Acoustics (UK) Autumn Conference 2008**, Oxford, UK (ioa.org.uk).
21–23	**International Conference on Low Frequency Noise and Vibration**, Tokyo, Japan (www.lowfrequency2008.org).
21–24	**acústica 2008**, Coimbra, Portugal (www.spacustica.pt).
26–29	**inter-noise 2008**, Shanghai, China (www.internoise2008.org).
27–31	**XX Session of the Russian Acoustical Society**, Moscow, Russia (rao.akin.ru).
30–31	*__Autumn Conference of the Swiss Acoustical Society__, Lansanne, Switzerland (www.sga-ssa.ch)

November 2008

2–4	**IEEE Internatinal Ultrasonics Symposium**, Bejjing, China (www.ieee.org/conf/ius_2008).
5–7	**Iberoamerican Acoustics Congress (FIA 2008)**, Buenos Aires, Argentina (www.adaa.org.ar).
14–18	**20th Session of the Russian Acoustical Society**, Moscow, Russia (www.akin.ru).
20–21	**Reproduced Sound 24**, Brighton, UK (ioa.org.uk).
24–26	**Australian Acoustical Society National Conference,** Geelong, Vic., Australia (www.acoustics.asn.au).
30–5	*18__th Biennial Congress of the Australian Insititute of Physics__, Adelaide, Australia (www.aipc2008.com).

December 2008

9–12	**19th International Acoustics Emission Symposium**, Kyoto, Japan (www.19iaes-kyoto.com/index.html)
17–19	**Symposium on the Acoustics of Poro-Elastic Materials (sapem 2008)**, Bradford, UK (sapem2008.matelys.com).

January 2009

11–17	**International Congress on Ultrasonics**, Santiago, Chile (www.icaltrasonics.org).

March 2009

23–26	**International Conference on Acoustics (NAG/DAGA 2009)**, Rotterdam, The Netherlands (www.nag-daga.nl).

April 2009

5–9	**Noise and Vibration: Emerging Methods (NOVEM 2009)**, Oxford, UK (www.isvr.spton.ac.uk/NOVEM2009).
13–17	**2nd International Conference on Shallow Water Acoustics**, Shanghai, China ([soon]: www.apl.washington.edu).
9–24	**International Conference on Acoustics, Speech, and Signal Processing**, Taipei. R.O.C. (icassp09.com).

July 2009

5–9	**16th International Congress on Sound and Vibration**, Krakow, Poland (www.icsv16.org).

August 2009

23–28	**inter-noise 2009**, Ottawa, Ont., Canada (www.internoise2009.com)

September 2009

6–10	**Inter-Speech 2009**, Brighton, UK (www.interspeech2009.org).
19–23	**IEEE 2009 Ultrasonics Symposium**, Rome, Italy (e-mail: pappalar@uniroma3.it).
23–25	*__TECNIACUSTICA2010__, Cădiz, Spain (www.sea-acustica.es).

October 2009

26–28	**Euronoise 2009**, Edinburgh, UK (www.euronoise2009.org.uk).

June 2010

13–16	*__INTERNOISE2010__, Lisbon, Portugal, (www.internoise2010.org).

August 2010

23–27	**20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (www.ica2010sydney.org).

September 2010

26–30	**Interspeech 2010**, Makuhari, Japan (www.interspeech2010.org).

Nikolai Andreyevich Dubrovsky 1933–2008

Nikolai Andreyevich Dubrovsky, the President of the Russian Acoustical Society died of serious illness on April 16, 2008. He was born in suburban Moscow on April 25, 1933. After graduating from the Moscow Physics and Technology Institute in 1957 he started working at the Acoustical Institute in Moscow where he stayed until his death. Nikolai An-

 0001-4966/2008/124(2)/695/2/$23.00

dreyevich was initially a PhD student at the Institute (1957–1960) where he studied sound reception under the supervision of Academician N. Andreyev. Later, Nikolai Andreyevich was a researcher (1960–1963), head of the laboratory (1963–1976) and division head of the Acoustics Institute (1976–1989) where he investigated dolphin sound systems, signal processing and sound pattern recognition. He earned his PhD in 1963 and became Doctor of Science in physics and mathematics in 1980.

Nikolai Andreyevich started his administrative career in 1989 as the first deputy director of the N. Andreyev Acoustics Institute. He was a director there from 1990 to 2007. Nikolai Andreyevich deservedly accrued many honors. He was a winner of the State Prize of the USSR in 1983, he received the distinction of "Honored Scientist of the Russian Federation" and many medals.

Nikolai Andreyevich founded the Russian Acoustical Society (RAS) in 1991 and was elected as the first RAS president, a position he held until his death. He has done much to join RAS to the European Acoustics Association (EAA) and to establish close relations with the Acoustical Society of America. Nikolai Andreyevich made great efforts to involve FSU acousticians into EAA activities. He was a true leader of Russian acousticians and the memory of him will stay with us.

Elena V. Yudina
Executive Director of the RAS
Moscow, Russia

FORUM

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

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

What exactly is meant by the term "auralization?"

Jason E. Summers
Acoustics Division, Code 7142, U.S. Naval Research Laboratory, Washington, DC 20375-5350

(Received 24 March 2008; accepted 28 May 2008)
[DOI: 10.1121/1.2945708]

Changes in technology precipitate changes in language. This is well illustrated by the term "auralization." In the musical community, the term has a long history of use. The first this author has found in print is by Matthay[1] who, in his text on musical interpretation, noted the "ability keenly to visualize, or auralize things apart from their actual physical happening outside of us," terming auralization the "power of pre-hearing." When Martin[2] introduced the term to the acoustics community, nearly 40 years later, he defined it similarly as the process of forming "a mental impression of a sound not yet heard." An analogy was made with the usage of the term "visualization" to mean "form a mental image of." In this usage, "visualize" acts as a more technically nuanced form of the verb "envisage."

Another 40 years later, technology had enabled computational rendering of virtual auditory scenes and Kleiner[3] recoined the term auralization for the room-acoustics community to mean, "the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space." Here, an analogy was made with the newer usage of the term visualization to mean "make visible." The definition intentionally encompasses the work of Spandöck[4] and his successors, who auralized sound fields in auditoria by radiating sound into small-scale models and then appropriately rescaling the recorded signals (slowing down the tape replay speed). In his earlier article, Martin had partially foreshadowed this new definition by Kleiner in which the sound field rather than the musical score was the subject of the auralization, noting that one could auralize to "compensate for listening means and environment."

Today, usage of auralize and auralization by the musical community remains essentially true to its earliest uses (see, e.g., Refs. 5 and 6). The essential concept is that the mind of the listener allows for something to be heard in the absence of an external stimulus. The mind (or the "imagination," to use Matthay's terminology) is the agent that performs the rendering and, typically, the thing being rendered is a musical score.

In contrast, usage of the term outside of the musical community has grown broader. The essential concept in Kleiner's definition is that modeling is the agent that performs the rendering and the thing being rendered is a sound field. For the room-acoustics community, this definition remains foundational, though Vorländer[7] expands it to include any process that renders audible numerical data arrived at through simulation, measurement, or synthesis while noting that, in modern parlance, both the process and the result of the process are termed (an) auralization. This broadening of the term is insidious, such that, if each of the meanings of the term is employed, it is grammatically correct for one to "auralize the auralization that will be auralized."

Of course, eventual broadening of the meaning of the term auralization is implicit in its analogical definition. Visualization, in the sense of "make visible," always had a broader meaning than Kleiner ascribed to it in his analogy with auralization. Merriam-Webster defines visualization in this sense as the "act or process of interpreting in visual terms."[8] Consequently, there is early usage of the term auralization in this sense (see, e.g., Ref. 9). More recently, the aural analogy of this meaning of the term visualization has been termed "sonification," which is defined as the "transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation."[10] While there is much to be admired in this more nuanced delineation, the analogy with visualization is lost. The visual parallel of sonification is not termed "lumification." Likewise, "audification," which describes a specific subset of sonification, has no visual parallel in "vidification." In fact, the term visualize allows for the same grammatical jumble described earlier.

Contrary to first appearances, the science of sound has been comparatively precise in defining terms. The most logical course now is to accept the broadening of the term auralization to match visualization and maintain symmetry between terms describing the senses of hearing and sight. Yet, as usage of the term auralization grows, there is also now a pressing need for formal definition of the term and standardization of its technical usage.

[1] T. Matthay, *Musical Interpretation, its Laws and Principles, and their Application in Teaching and Performing* (The Boston Music Company, Boston, MA, 1913), p. 10.
[2] D. W. Martin, "Do you auralize?," J. Acoust. Soc. Am. **24**, 416 (1952).
[3] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization—An overview," J. Audio Eng. Soc. **41**, 861–874 (1993).
[4] F. Spandöck, "Akustische Modellversuche," Ann. Phys. **20**, 345–360 (1934).
[5] G. S. Karpinski, *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-Level Musicians* (Oxford University Press, Oxford, UK, 2000).
[6] C. Lawson, *The Cambridge Companion to the Orchestra* (Cambridge University Press, Cambridge, UK, 2003).
[7] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms, and Acoustic Virtual Reality* (Springer, Berlin, 2008).
[8] *Merriam-Webster's Collegiate Dictionary*, 11th ed. (Merriam-Webster, Springfield, MA, 2003, principal copyright 1993).
[9] L. Albright, J. A. Jackson, and J. Francioni, "Auralization of parallel programs," ACM SIGCHI Bulletin **23**(4), 86–87 (1991).
[10] G. Kramer *et al.*, "The sonification report: Status of the field and research agenda," report prepared for the National Science Foundation by members of the International Community for Auditory Display (ICAD), Santa Fe, NM, 1999. The report is available online at the site ⟨http://icad.org/websiteV2.0/References/nsf.html⟩ (last viewed 24 March 2008).

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

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:

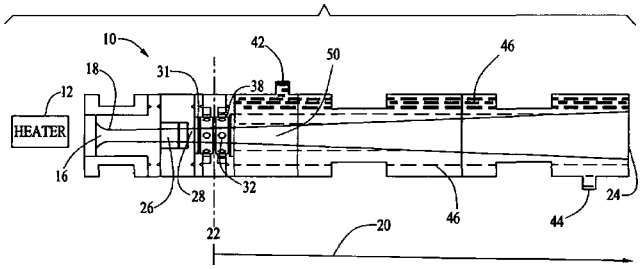
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
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*

7,296,396

43.25.Ts METHOD FOR USING VARIABLE SUPERSONIC MACH NUMBER AIR HEATER UTILIZING SUPERSONIC COMBUSTION

Kenneth J. Wilson *et al.*, assignors to United States of America as represented by the Secretary of the Navy
20 November 2007 (Class 60/211); filed 14 December 2004

In hypersonic aircraft engine laboratory testing, a supersonic flow of low density inlet air is required. The author claims an air heater and accelerator **10** that uses edge **28** that trips eddies at the cavity **26** resonance,



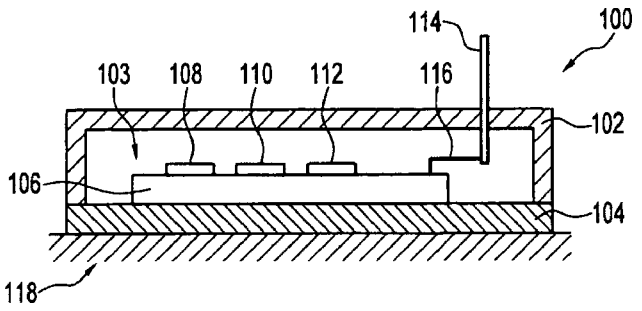
thereby causing mixing of the fuel and oxidizer gases for combustion in channel **50**.—AJC

7,302,864

43.35.Zc TORQUE SENSOR

James Zt Liu and Steven J. Magee, assignors to Honeywell International Incorporated
4 December 2007 (Class 73/862); filed 23 September 2005

The author claims a surface acoustic wave (SAW) shaft torque sensor package **100** that attaches to the surface **118** of a power transmission shaft. Diaphragm **106** is formed as part of, or attached to SAW substrate **104**. Sensors **108–112** can be SAW or bulk acoustic wave (BAW) transducers that



vibrate in any of 15 commonly known modes (flexible plate, acoustic plate, shear horizontal etc.). Claims include the manufacturing methods to create sensor package **100** carrying any transducers, three or so of 15 named varieties.—AJC

7,327,637

43.35.Zc ACOUSTIC PULSE ACTUATOR

Joshua M. Chambers *et al.*, assignors to Massachusetts Institute of Technology
5 February 2008 (Class 367/140); filed 22 February 2006

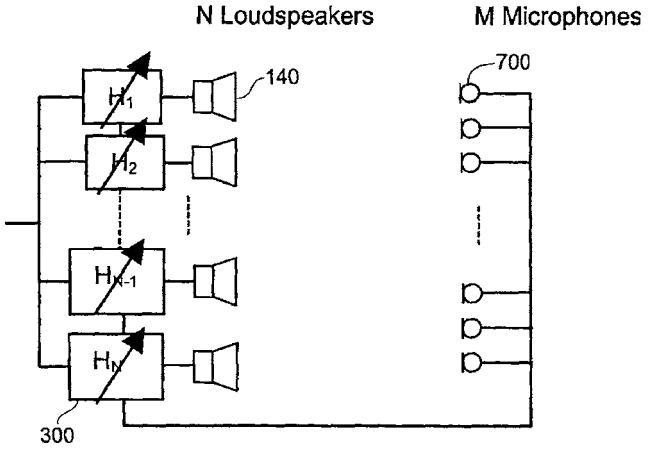
This patent pertains to a technique for actuating materials that demonstrate an actuation response to an applied stress. It is claimed that this invention overcomes the limitations of conventional actuation in that it produces large actuation strokes with fast actuation response time and high output strain at convenient operating temperatures. Key to embodiment of this patent is the finding that actuation materials, including those conventionally actuated by electric or magnetic fields (e.g., piezoelectric or magnetostrictive materials), can be actuated by an acoustic stress wave. Actuators according to this patent, such as those for micropositioning, involve action of acoustic stress waves from a stress wave generator on an actuation material.—EEU

7,336,793

43.38.Hz LOUDSPEAKER SYSTEM FOR VIRTUAL SOUND SYNTHESIS

Ulrich Horbach and Etienne Corteel, assignors to Harman International Industries, Incorporated
26 February 2008 (Class 381/59); filed 8 May 2003

It appears that wave field synthesis may eventually become a practical alternative to multi-channel stereo sound reproduction. Using two-dimensional wave field synthesis, a virtual sound source can be located almost anywhere on a horizontal plane, providing depth as well as lateral placement. In practice, wave field synthesis requires a great many channels



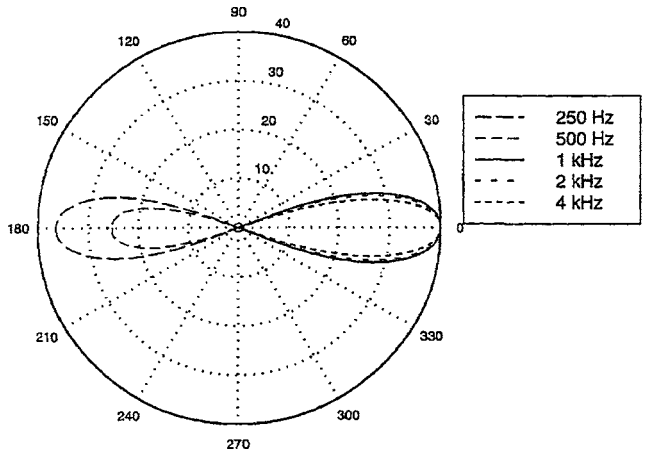
and loudspeakers even when restricted to two dimensions. Several existing patents suggest using multi-driver panel loudspeakers to provide multiple, closely spaced sound sources. The patent at hand describes a method for calibrating and filtering an array of multi-driver panels in a particular listening environment to recreate the original sound field as closely as possible. The benefits of this method are said to extend above the aliasing frequency. Although the procedure is fairly complicated, the patent is clearly written and easy to follow.—GLA

7,343,018

43.38.Hz SYSTEM OF SOUND TRANSDUCERS WITH CONTROLLABLE DIRECTIONAL PROPERTIES

Johan van der Werff, assignors to PCI Corporation
11 March 2008 (Class 381/80); filed in Netherlands 13 September 2000

To suppress side lobes, the individual elements of a loudspeaker line array can be attenuated with respect to location (shading), lowpass filtered (tapering), or a combination of both. The inventor has opted for a combination of both and has worked out a set of formulas to determine the filter coefficients for a desired beamwidth. Some features of the proposed method are a bit puzzling, however. For example, the diagram shows the coverage pattern of a 43-speaker array made up of 13.5 cm (nominal 5 in.) diameter loudspeakers. It appears that individual sources are assumed to be nondirectional, yet total side lobe suppression is achieved all the way up to 4 kHz, well above the aliasing frequency. Some degree of skepticism seems justified. Also, the polar plot format is peculiar: 1 kHz, 2 kHz, and 4 kHz are



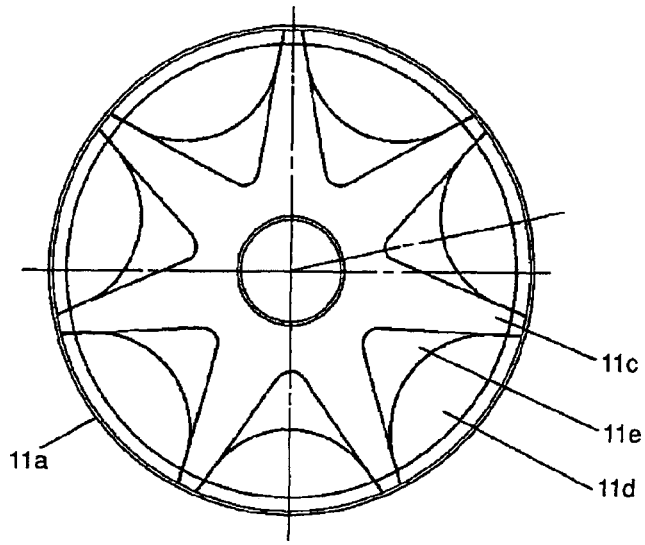
plotted in the front hemisphere; 250 Hz and 500 Hz are plotted in the rear hemisphere, with 250 Hz attenuated about 4 dB and 500 Hz attenuated 15 dB. All in all, the patent poses more questions than it answers.—GLA

7,315,628

43.38.Ja DIAPHRAGM FOR LOUD SPEAKER AND LOUD SPEAKER EMPLOYING IT

Ryo Kuribayashi and Shinsaku Sawa, assignors to Matsushita Electric Industrial Company, Limited
1 January 2008 (Class 381/424); filed in Japan 15 October 2003

Cone **11** is part of an electrodynamics loudspeaker assembly. Looking at the back of the cone one sees that it has several parts—**11a**, which is a conventional surround bonded to the perimeter of the cone body **11d**, to which web shaped part **11e** is affixed, and on top of that an "equiangular



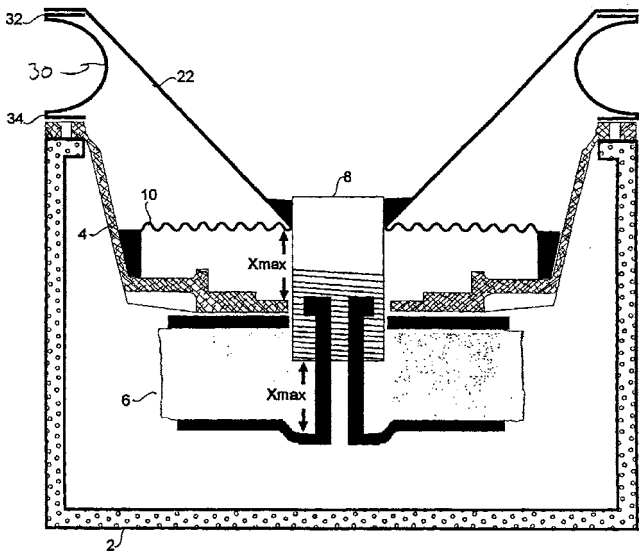
seven thick part" [sic] **11c** is affixed. The description of the patent allows for the web shaped part to decrease in thickness from the periphery towards the center. Two graphs in the patent appear to indicate some smoothing of the frequency response compared to what is called a prior art device.—NAS

7,318,496

43.38.Ja ACOUSTIC RADIATOR WITH A BAFFLE OF A DIAMETER AT LEAST AS LARGE AS THE OPENING OF THE SPEAKER ENCLOSURE TO WHICH IT IS MOUNTED

Joseph Y. Sahyoun, Redwood City, California
15 January 2008 (Class 181/172); filed 16 April 2002

To increase the output of audio subwoofer transducers, designers can increase the radiating area and/or increase the excursion of the cone **22**. For the former, a large part of the area is necessarily needed for the surround so the patent increases the radiating area for a given diameter by mounting the



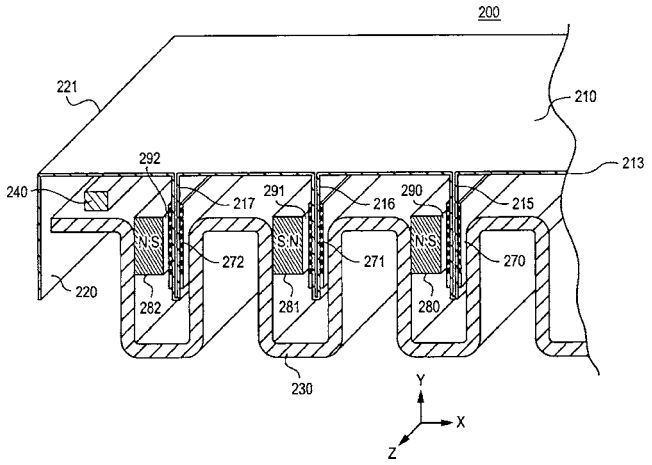
surround **30** below the circumference of the cone instead of radially outward from the edge of the cone. Various implementations where the half roll of the surround is facing inward instead of outward, as seen in the figure, as well as pleated surrounds, are described.—NAS

7,333,620

43.38.Ja ACOUSTIC TRANSDUCER WITH MECHANICAL BALANCING

An Duc Nguyen and Charles M. Sprinkle, assignors to Harman International Industries, Incorporated
19 February 2008 (Class 381/152); filed 9 April 2004

The diagram shows a clever design for the voice coil and magnetic circuit of a planar loudspeaker. It is described in great detail in the patent text, along with several variants, yet the patent claims contain not one word about the motor structure. They are concerned solely with the folded and pleated construction of the diaphragm. How can this be? It turns out that this is just one component of a patent triplet issuing from three applications filed on the same date, each of which incorporates the others by reference. Those



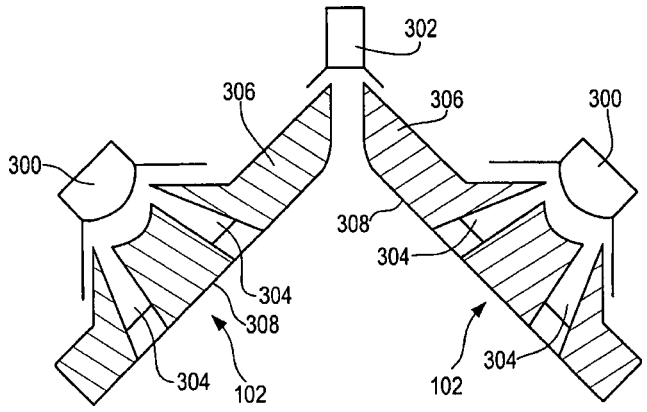
interested in the field are advised to wait until the remaining two patents are published, obtain copies of all three, and staple them together.—GLA

7,333,626

43.38.Ja ARBITRARY COVERAGE ANGLE SOUND INTEGRATOR

Scott M. Opie *et al.*, assignors to Harman International Industries, Incorporated
19 February 2008 (Class 381/345); filed 28 February 2006

My dictionary lists more than six synonyms for the word "arbitrary," including "uncontrolled," "capricious," and "unreasonable." However, the title of this patent probably refers to a mathematical definition, i.e., "not assigned a specific value." The very short patent claims use the words "adjustable angle" instead. Looking at the simplified horizontal section, we see



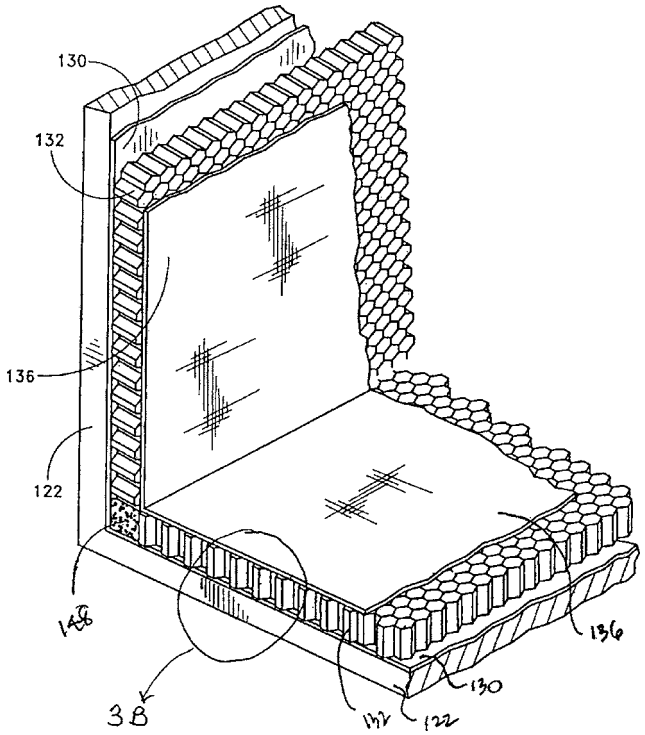
a high frequency driver **302** coupled to a waveguide formed by surfaces **306**, each of which has a curved section and a planar section. Sound from midrange drivers **300** enters the waveguide through restricted openings **304**. This basic geometry is used in each element of a vertical line array to control the horizontal mid and high frequency coverage angle.—GLA

7,337,874

43.38.Ja LIGHTWEIGHT SPEAKER ENCLOSURE

Ross Ritto, assignor to Southern California Sound Image
4 March 2008 (Class 181/199); filed 13 June 2005

For touring or mobile sound applications a loudspeaker enclosure should be lightweight, yet rigid and nonresonant. The panels of this speaker enclosure are made of an outer layer **122**, a middle "sound absorbing layer"



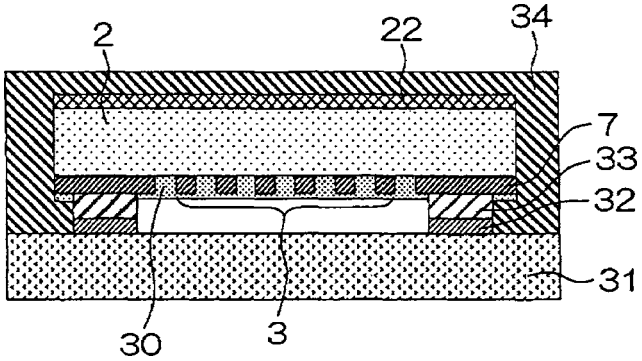
132, and an inner layer **136**. The properties of these three layers are spelled out in three short patent claims.—GLA

7,307,369

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE, SURFACE ACOUSTIC WAVE APPARATUS, AND COMMUNICATIONS EQUIPMENT

Yuuko Yokota *et al.*, assignors to Kyocera Corporation
11 December 2007 (Class 310/313 R); filed in Japan 26 August 2004

Claims include the manufacturing method for a surface acoustic wave (SAW) radio frequency (RF) filter including semiconductor layer **22** on the backside of piezoelectric substrate **2** to reduce electrostatic voltage damage



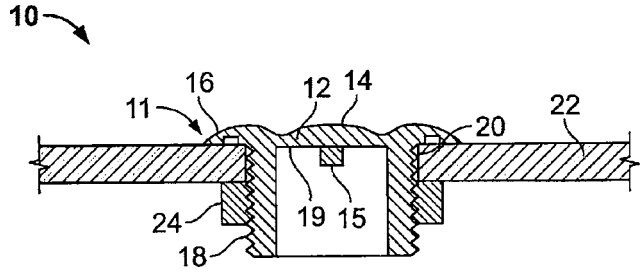
during construction and the inverted mounting position to reduce parasitic capacitance between the interdigital transducer fingers **3**, thereby reducing the out-of-band RF transmission.—AJC

7,307,627

43.38.Rh INDIVIDUAL ACOUSTIC WAVE SWITCH

Terence J. Knowles and Charles F. Bremigan III, assignors to Illinois Tool Works, Incorporated
11 December 2007 (Class 345/177); filed 12 May 2003

A touch switch is claimed where sound emitted from shear transducer **15** resonates in cylinder **18**. Such sound is damped by pressing an acoustical



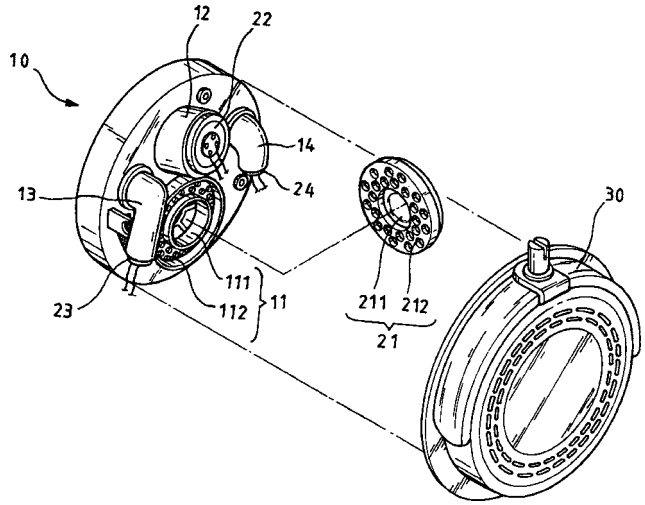
absorber, e.g., a finger tip, to cap **12**. Sound detection is also by transducer **15**.—AJC

7,340,071

43.38.Si HEADPHONES WITH A MULTICHANNEL GUIDING MECHANISM

Jui-Shu Huang, Taoyuan City, Taiwan
4 March 2008 (Class 381/309); filed 2 June 2004

A number of inventors have patented headphones that attempt to create miniature listening rooms attached to the user's head. In this case, however, the goal seems to be acoustical separation yet integration of sound from a front main speaker **21**, subwoofer **22**, and rear surround speaker **23**,



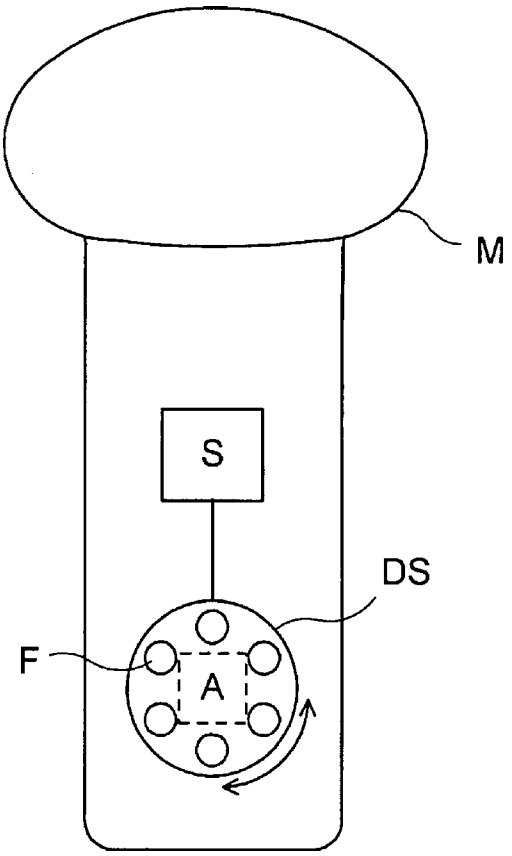
"…thereby allowing the front main channel (F) to have the treble area (H), the bass area (F), the subwoofer (W), and the rear surround channel (R). These are what the conventional headphones don't have."—GLA

7,317,808

43.38.Tj MICROPHONE

Wolfgang Niehoff, assignor to Sennheiser Electronic GmbH & Company, KG
8 January 2008 (Class 381/355); filed in Germany 24 July 2002

In live sound performances where multiple handheld wireless microphone transmitters are used, a piece of colored tape applied to the microphone body has been a simple, but some would say, inelegant, way of



identifying the microphone. A means of changing the identification color that is a bit more elegant than tape is described—by providing a bezel **DS** that can rotate, with colored ports **F**. Thus a visible indication of the microphone channel is provided so long as the talent does not cover the bezel while holding the microphone. In addition, the patent appears to also claim a means of encoding this information so that the receiving equipment can display the color associated with that microphone, but perhaps the reviewer is reading more into the patent than is there.—NAS

7,320,455

43.40.Tm INSTRUMENTED PLATFORM FOR VIBRATION-SENSITIVE EQUIPMENT

Vyacheslav M. Ryaboy and Warren Booth, assignors to Newport Corporation
22 January 2008 (Class 248/638); filed 24 October 2003

A table top, such as that for an optical table, consists of two plates separated by an inner core. The upper plate typically supports a payload that is to be protected from vibrations. In a platform as described by this patent, vibration sensors may be attached to the upper plate and a damper may be placed between the two plates. The damper may be passive or it may be actively controlled on the basis of signals generated by the sensors.—EEU

7,328,770

43.40.Tm STRAP SILENCER

Jaffrey A. Owens, Lansing, Michigan *et al.*
12 February 2008 (Class 181/207); filed 16 June 2005

This silencer is intended to be attached to straps, such as those that are used to fasten luggage and the like atop vehicles, which straps tend to vibrate and to generate annoying noise as air passes over them. A damper according to this patent in essence consists of a bundle of fibers that may be attached to a strap via a hook or similar device, with the fibers extending outward, into the air stream.—EEU

7,329,198

43.40.Tm SILENT CHAIN

Kazufumi Kotani, assignor to Tsubakimoto Chain Company
12 February 2008 (Class 474/212); filed in Japan 27 September 2005

As stated in the patent's abstract, "in a silent chain composed of interleaved rows of link plates held by connecting pins secured to guide plates disposed along the sides of the chain, the guide plates are formed with protrusions, and are deformed so that they exert a spring action on the rows of link plates through the protrusions, pressing the link plates against one another to increase the frictional contact between the link plates, and thereby reduce vibration and vibration noise."—EEU

7,331,847

43.40.Tm VIBRATION DAMPING IN CHEMICAL MECHANICAL POLISHING SYSTEM

Hung Chih Chen *et al.*, assignors to Applied Materials, Incorporated
19 February 2008 (Class 451/285); filed 17 January 2006

Smoother polishing of substrates on which integrated circuits are formed is accomplished by use of polishing heads whose vibration is suppressed. The polishing heads described in this patent incorporate viscoelastic elements in their supports and/or include a layer of viscoelastic material in their construction in order to reduce the heads' vibrations.—EEU

7,332,118

43.40.Tm METHOD OF PREPARING AND METHOD OF APPLYING A VIBRATION DAMPING SYSTEM

John Asmussen, assignor to Rockwool International A/S
19 February 2008 (Class 264/257); filed in Denmark 4 April 2001

This patent pertains to vibration isolation pads to be installed under rail systems or under roads. The isolation pads described here are in the form of plates made of mineral fibers and of foamed or nonfoamed polymeric material. Voids may be provided to reduce the pads' stiffness and drainage provisions may be incorporated in the pad's configurations. The pads are claimed to be durable and to have long-term stable dynamic characteristics.—EEU

7,334,552

43.40.Tm INTERNAL VISCOUS DAMPER MONITORING SYSTEM AND METHOD

Peter Möller and Erik Svenske, assignors to Ford Global Technologies, LLC
26 February 2008 (Class 123/192.1); filed 7 October 2005

Internal combustion engines in vehicles often have rotational dampers mounted on their crankshafts to suppress vibrations. These dampers may contain a viscous fluid, along with a flywheel, as well as elastomeric elements. In recognition of the fact that deterioration of the damper may result in excessive engine vibrations, this patent describes an approach to monitoring the engine vibrations and limiting the engine speed if vibrations are detected that exceed certain limits and that are ascribable to damper deterioration.—EEU

7,337,586

43.40.Tm ANTI-SEISMIC DEVICE WITH VIBRATION-REDUCING UNITS ARRANGED IN PARALLEL

Chi-Chang Lin, Nan Dist., Taichung City and Jer-Fu Wang, Tso-Yuan City, both of Taiwan
4 March 2008 (Class 52/167.1); filed 14 June 2004

Several dynamic absorbers, each tuned to a different frequency in a given range, are arranged in parallel and attached to a structure whose vibrations are to be suppressed. Each absorber consists of a spring-restrained piston that can move axially in a cylinder. The absorber array may be attached so as to act horizontally on a building to reduce its response to seismic excitation; vertically acting arrays may be mounted on bridges to suppress their vibrations.—EEU

7,337,981

43.40.Tm VIBRATION DAMPING SYSTEM

John Christian Asmussen, assignor to Rockwool International A/S
4 March 2008 (Class 238/283); filed in the European Patent Office 14 November 2001

This patent, which is closely related to U. S. Patent No. 7,332,118, also deals with isolation pads to be placed under rail installations. Pads according to this patent have a relatively rigid top layer that can distribute the loads acting on it—for example, those transmitted via ballast. The pads are configured so that they can be placed in layers, one atop the other, so that the upper layer covers the joints in the lower layer.—EEU

7,341,550

43.40.Tm ROLL, IN PARTICULAR MIDDLE ROLL OF A CALENDAR, AND CALENDAR

Rolf van Haag, assignor to Voith Paper Patent GmbH
11 March 2008 (Class 492/42); filed in Germany 17 October 2002

Vibrations of calendar rolls, which essentially are tubes, are suppressed by means of dynamic absorbers located inside the tubes. A typical dynamic absorber here consists of a mass that is centered on the axis of the tube and held in place by resilient elements, such as rubber rings. A single cylindrical mass may extend along the entire length of the tube; alternatively, several separate masses may be spaced along the tube.—EEU

7,333,882

43.40.Vn SUSPENSION CONTROL APPARATUS

Toru Uchino *et al.*, assignors to Hitachi, Limited
19 February 2008 (Class 701/37); filed in Japan 12 February 2004

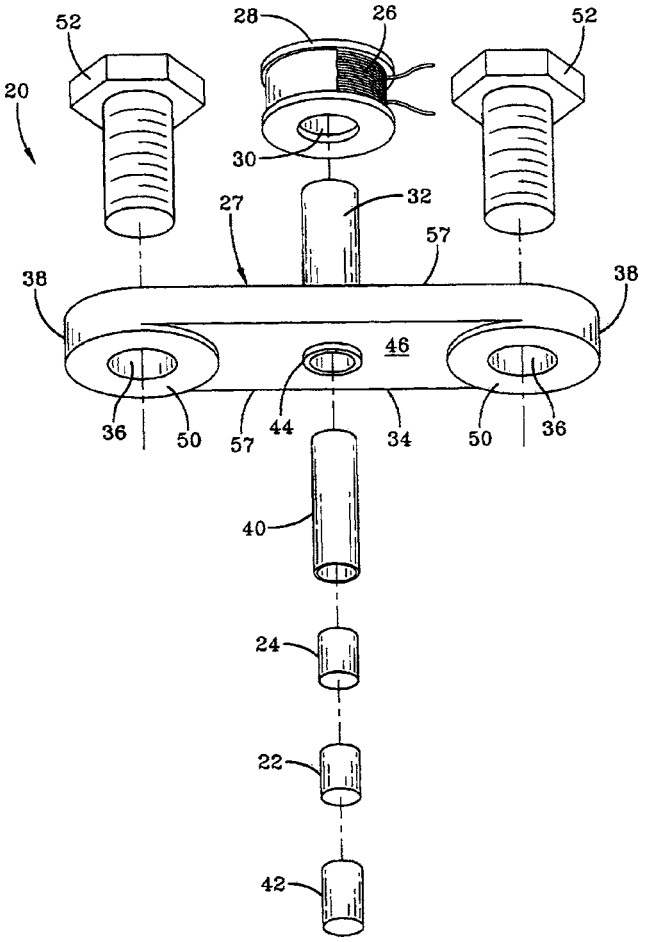
In this active suspension system for a motor vehicle the vertical motions of both the sprung mass (the vehicle body) and the unsprung mass (the wheels) are sensed and the signals are fed to a controller that causes the shock absorber characteristics to be adjusted. In one embodiment there is provided a filter that passes signals in the vicinity of the natural frequency of the unsprung mass, with the filter adjusted in accordance with the weight of the unsprung mass. There is included a function for judging the road surface condition on the basis of the vertical acceleration of the unsprung mass or from that of the sprung mass. In one embodiment the shock absorber characteristics for the rear wheels are controlled on the basis of the signal from the unsprung mass at the front of the vehicle.—EEU

7,298,237

43.40.Yq MAGNETOSTRICTIVE STRESS WAVE SENSOR

Steven R. Stuve, assignor to Key Safety Systems, Incorporated
20 November 2007 (Class 335/215); filed 20 July 2006

The author claims a vehicle crash shock wave sensor **20** comprising a Terfonal-D "giant magnetostrictive core" **22** next to a biasing magnet **24** and a spacer **42**, all bonded end to end inside tube **40** and bolted by **52** and **38** against the vehicle frame, contacting at **44**. Coil **26** produces a 0.2–2 V



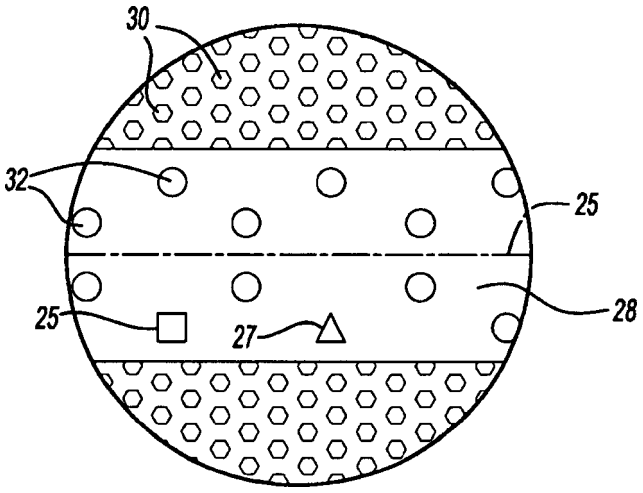
pulse when a vehicle shock wave occurs as the result of a crash. The assembled unit apparently has no moving parts and relies on the magnetostriction EMF excited by the structural frame vibration impulse induced by the crash.—AJC

7,296,656

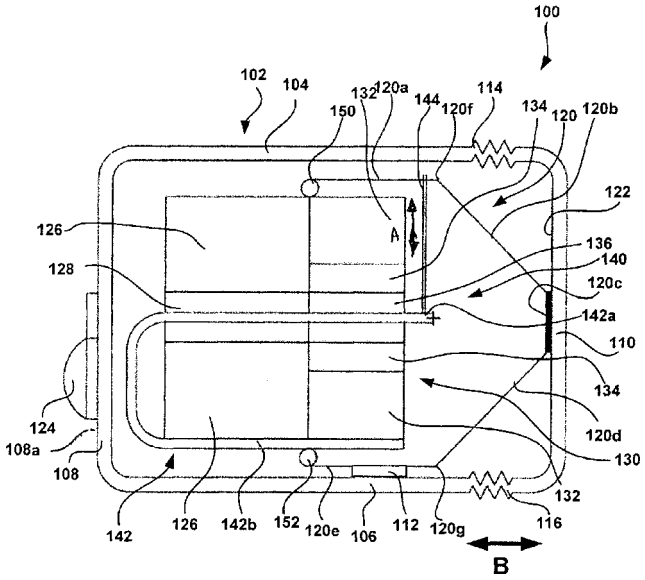
43.50.Gf ACOUSTIC MECHANICAL RETAINER

Scott Sanicki and Mark W. Costa, assignors to United Technologies Corporation
20 November 2007 (Class 181/210); filed 22 April 2005

In the sound absorber lining of turbine engine inlet surfaces, the protective perforated surface **30** is fabricated from several panels that must be bonded in place at their respective perimeters, resulting in some surface



areas **28** that are no longer perforated. The author claims the method of drilling holes **32**, **25**, **27**, usually larger in diameter than and fewer in number than the perforated holes **30**, would restore the sound absorption capability of those bond surface areas.—AJC



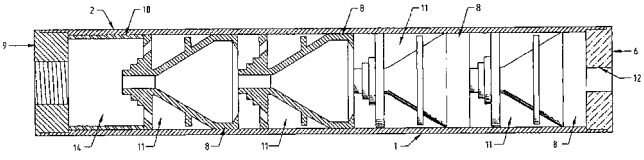
the movable portion of the housing and on the other end to a drive rod in the magnetic motor assembly, translates displacement of the drive rod to the shiftable housing portion.—DAP

7,308,967

43.50.Gf SOUND SUPPRESSOR

Thomas Trail Hoel, assignor to Gemini Technologies, Incorporated
18 December 2007 (Class 181/223); filed 21 November 2005

Silencers can decrease the accuracy of the firearm to which they are attached due to interior baffle design, among other things. The patent presents a silencer that is said to increase the sound reflection from the internal



baffles by using an improved profile as well as subvents and better controlled venting between baffles, which are also said to provide better accuracy over prior art.—NAS

7,336,797

43.50.Gf APPARATUS AND METHOD FOR GENERATING ACOUSTIC ENERGY IN A RECEIVER ASSEMBLY

Stephen C. Thompson *et al.*, assignors to Knowles Electronics, LLC
26 February 2008 (Class 381/418); filed 10 May 2004

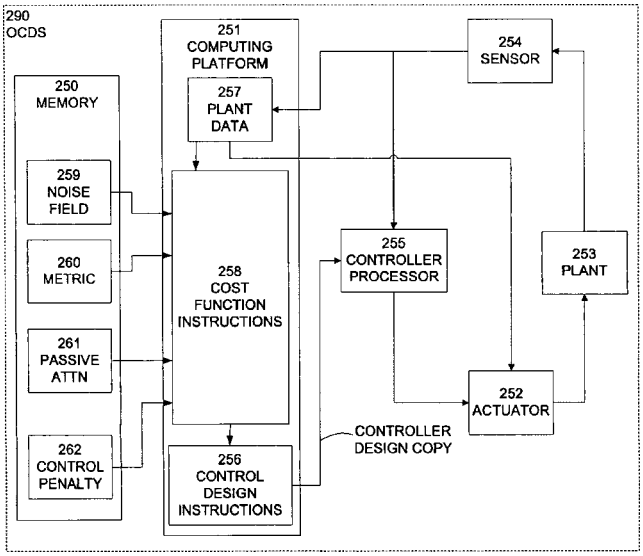
To reduce size while improving sensitivity, stability, and robustness, a flexible membrane in the surface of the receiver housing acts as an accordion-like bellows to allow relative motion between the fixed and movable portions of the housing. A linkage assembly, connected on one end to

7,308,106

43.50.Hg SYSTEM AND METHOD FOR OPTIMIZED ACTIVE CONTROLLER DESIGN IN AN ANR SYSTEM

Michael A. Vaudrey *et al.*, assignors to Adaptive Technologies, Incorporated
11 December 2007 (Class 381/72); filed 17 May 2004

The author claims an active noise reduction (ANR) hearing protection system that incorporates an optimizing controller design system (OCDS) applicable to both ANR earmuffs and ANR earplugs. OCDS **129** adjusts for the plant ("the dynamics associated with the actuator, sensor, and the acoustic dynamics in the occluded space") of individual users and individual ANR



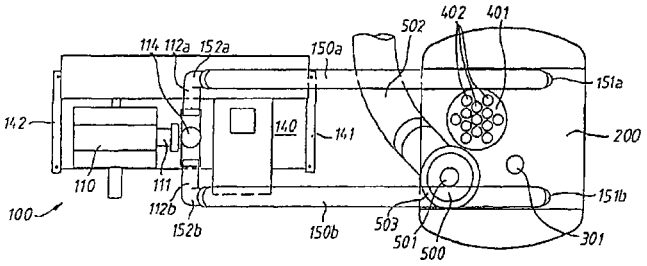
devices. Control objectives **256** can be the limitation of any or all of perceived loudness, hearing damage (minimizing the A-Weighted sound onto the ear drum), cost **258**, or damage to actuator **252**.—AJC

7,311,062

43.50.Jh MILKING PLANT AND METHOD FOR REDUCING SOUND EMISSIONS IN A MILKING PLANT

Erwin Bilgery, assignor to Moser Stalleinrichtungen and Bitec Engineering
25 December 2007 (Class 119/14.07); filed in Switzerland 7 June 2001

Commercial milking operations use milking machines that typically require the maintenance of a vacuum in the milk extraction tubing and apparatus. The invention describes how to reduce the noise and vibration from the vacuum control valve **500**, the use of flexible hoses to reduce



vibration and noise via this path, and a means for mounting the pulsator unit to reduce vibration from same. The patent provides a brief description of how the device works. Got cookies?—NAS

7,340,407

43.55.Rg SOUND CONTROL METHOD

Ralph Michael Fay *et al.*, assignors to Johns Manville
4 March 2008 (Class 705/7); filed 13 November 2000

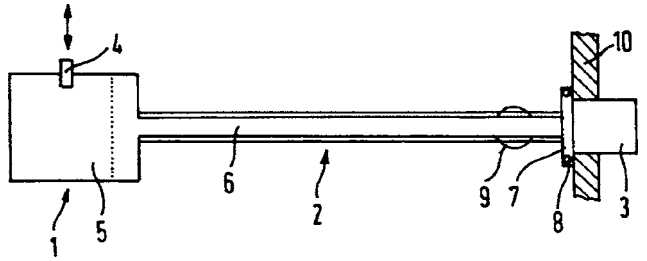
At first glance, this appears to be a presumptuous and shallow patent that describes a basic engineering approach for sound isolation analysis (part of professional practice for about 100 years) and deems it to be an "invention" due to the fact that common sense has been codified and data is accessible via a computer. On closer examination and more careful reading of the patent, it still seems to be no more than it appeared to be at first glance.—CJR

7,299,676

43.58.Vb ACOUSTIC PRESSURE CALIBRATOR

Guenther Mueller and Joseph Steigenberger, assignors to EADS Deutschland GmbH
27 November 2007 (Class 73/1.82); filed in Germany 5 October 1999

A portable means to calibrate pressure sensors by adapting a standard pistonphone is claimed. The port of pistonphone **1** is coupled by resonating tube **2**, **6** to the port of pressure sensor **3**, sealed by gasket **8**. Resonance is



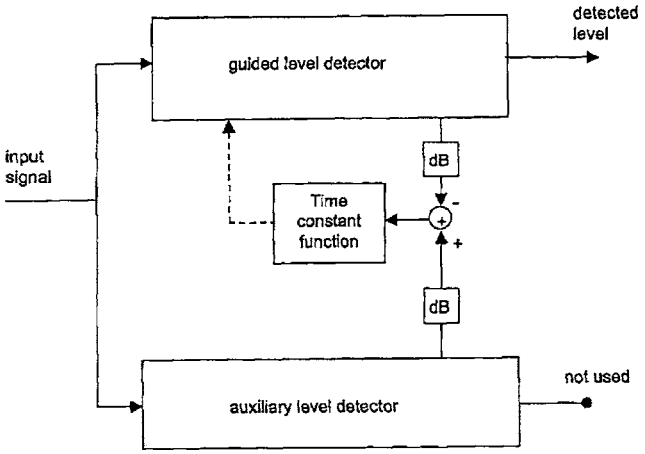
assured by adjustment **9**, resulting in a resonant sound pressure level of 151 dB at 315 Hz. Tests over several days indicate a repeatability of ±0.3 dB.—AJC

7,333,623

43.66.Ts METHOD FOR DYNAMIC DETERMINATION OF TIME CONSTANTS, METHOD FOR LEVEL DETECTION, METHOD FOR COMPRESSING AN ELECTRIC AUDIO SIGNAL AND HEARING AID, WHEREIN THE METHOD FOR COMPRESSION IS USED

Joachim Neumann, assignor to Oticon A/S
19 February 2008 (Class 381/312); filed 26 March 2002

Time constants for a guided level detector in a hearing aid compression circuit are determined by subtracting, on a dB scale, the output of a guided level detector from the output of a fast-acting, fixed time constant auxiliary level detector. The guided level detector is set to longer time constants when the difference is negative. The guided level detector is set to



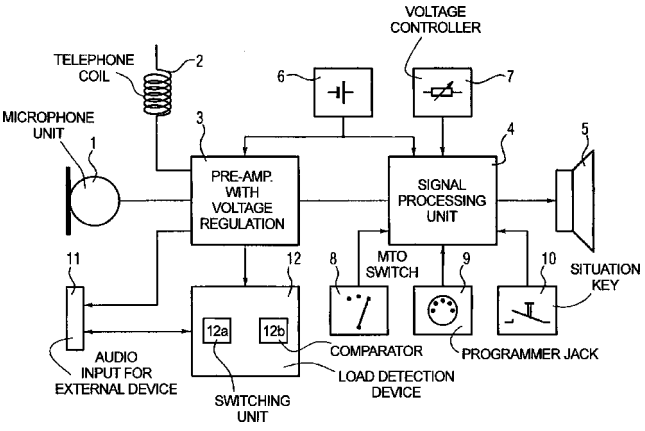
shorter time constants when the difference is positive and when the difference is greater. Either the input signal or the output of the auxiliary level detector is fed to the guided level detector.—DAP

7,333,624

43.66.Ts HEARING AID DEVICE AND OPERATING METHOD FOR AUTOMATICALLY SWITCHING VOLTAGE SUPPLY TO A CONNECTED EXTERNAL DEVICE

Kunibert Husung, assignor to Siemens Audiologische Technik GmbH
19 February 2008 (Class 381/312); filed in Germany 24 September 2003

The hearing aid device senses the presence and amount of power supply load from an external device, such as an FM receiver plug-in accessory.



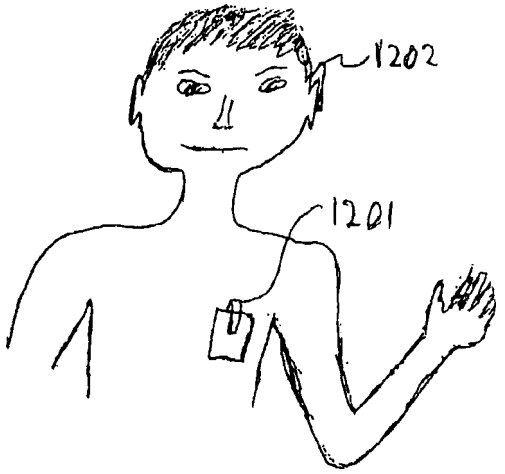
Depending on how much voltage and current the external device requires, either the battery voltage or a regulated voltage is automatically made available to the external device.—DAP

7,336,796

43.66.Ts HEARING ASSISTIVE APPARATUS HAVING SOUND REPLAY CAPABILITY AND SPATIALLY SEPARATED COMPONENTS

Trevor I. Blumenau, San Francisco, California
26 February 2008 (Class 381/315); filed 26 March 2003

Sounds in the vicinity of a person using hearing assistance are acquired and stored for replay at a later time. At least part of the data acquisition and replay system is separated from the sound production apparatus.



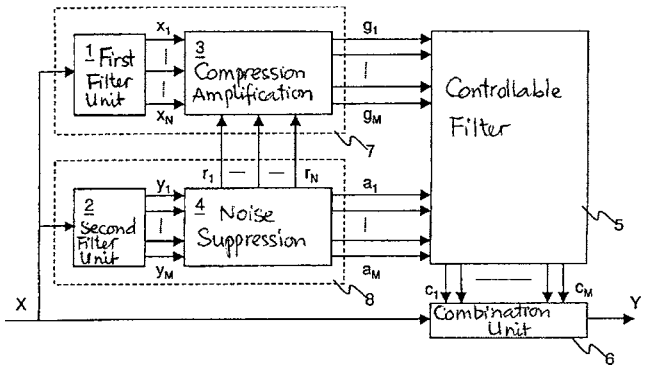
Sounds to be replayed may be transmitted from the data replay unit to the hearing assistance device via a wired or wireless link. The sound production section may or may not be adapted to be wearable and ear mounted. The person receiving the hearing assistance controls the duration of the sounds to be replayed.—DAP

7,340,072

43.66.Ts SIGNAL PROCESSING IN A HEARING AID

Arthur Schaub, assignor to Bernafon AG
4 March 2008 (Class 381/312); filed in the European Patent Office 26 February 2003

Methodology is given for determining, with little processing delay, coefficients for noise suppression from modulation depths in the input signal in a first set of frequency ranges, and in parallel determining frequency-dependent coefficients for compression amplification dependent on the noise suppression coefficients and input signal power levels in a second set of larger frequency ranges. Coefficients for adaptively filtering the input signal with a controllable filter are determined from combining the compression coefficients and the noise suppression coefficients. Adaptation takes place only if changes in the coefficient values exceed a predefined threshold.—DAP

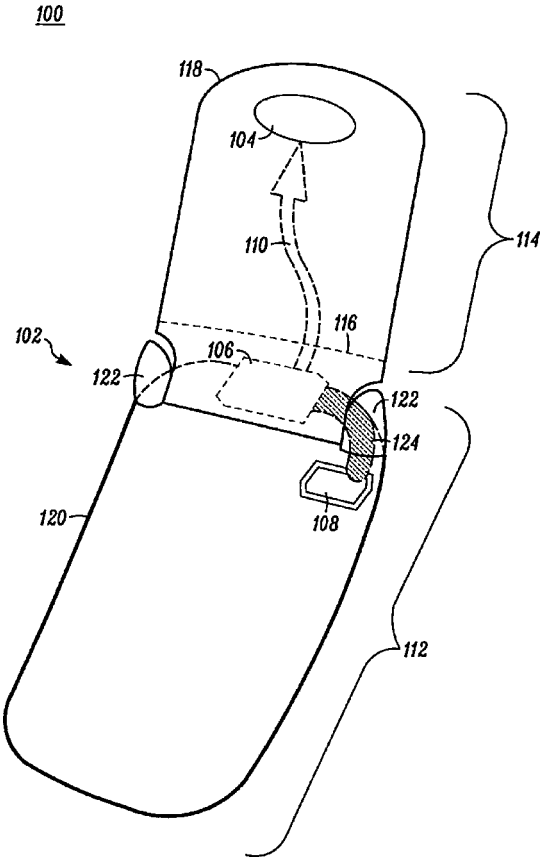


7,343,181

43.66.Ts WIRELESS COMMUNICATION DEVICE HAVING ELECTROMAGNETIC COMPATIBILITY FOR HEARING AID DEVICES

Yiu K. Chan *et al.*, assignors to Motorola Incorporated
11 March 2008 (Class 455/575.3); filed 8 August 2005

RF interference in hearing aids from cellular phones is reduced with a two-piece "clam-shell" phone design having the wireless transceiver, antenna, and speaker in the bottom portion furthest away from the hearing aid,



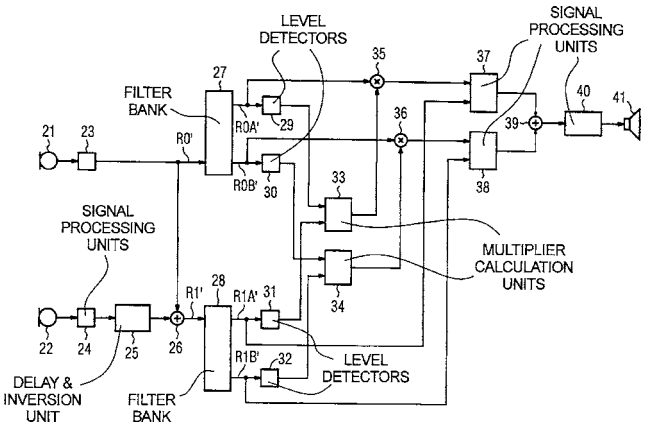
a sound passage for the speaker output, an earpiece and other nonelectromagnetic conductive materials in the upper portion. The minimum distance between the speaker in the lower housing and the earpiece in the upper housing is about a quarter wavelength of the wireless transceiver operating frequency, such as 800 MHz.—DAP

7,340,073

43.66.Ts HEARING AID AND OPERATING METHOD WITH SWITCHING AMONG DIFFERENT DIRECTIONAL CHARACTERISTICS

Eghart Fischer, assignor to Siemens Audiologische Technik GmbH
4 March 2008 (Class 381/313); filed in Germany 20 June 2003

Audible switching artifacts are avoided in a directional hearing aid that uses at least two microphones to produce different order directionality and



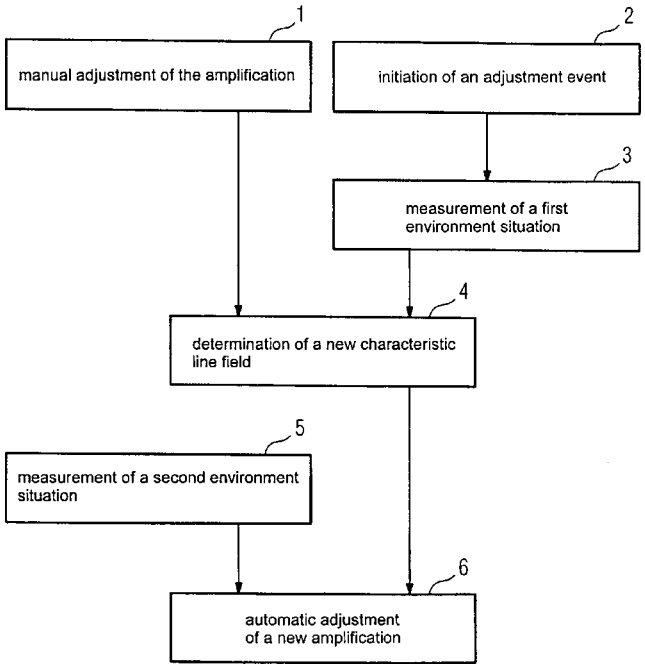
adaptive polar patterns. The microphone output levels are equalized to avoid sudden level changes while rapidly switching between or superimposing microphone signals.—DAP

7,340,074

43.66.Ts DEVICE AND METHOD TO ADJUST A HEARING DEVICE

Josef Chalupper, assignor to Siemens Audiologische Technik GmbH
4 March 2008 (Class 381/315); filed in Germany 27 February 2003

Methodology is presented for the wearer to fine-tune a hearing aid in his or her own listening environments without the intervention of a hearing professional. The wearer manually selects a preferred parameter set for a



first listening environment via an input mechanism that could be on the hearing aid, or via a wired or wireless remote device. Thereafter, when a similar environment is detected, the hearing aid automatically switches to those parameter settings.—DAP

7,340,308

43.66.Ts METHOD FOR ELECTRICALLY STIMULATING THE COCHLEA

Ben M. Clopton *et al.*, assignors to Advanced Cochlear Systems, Incorporated
4 March 2008 (Class 607/57); filed 8 June 2005

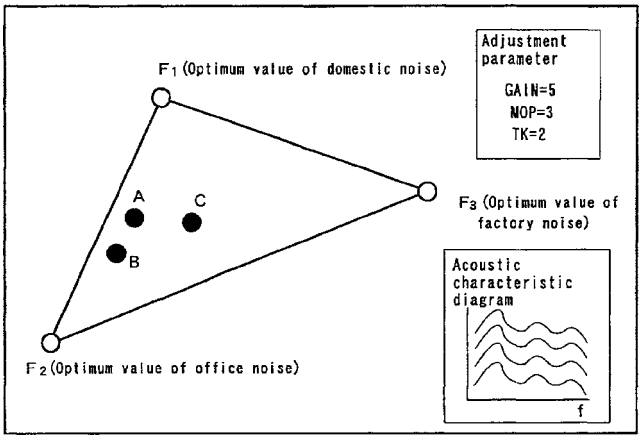
Goals include increasing the dynamic range of stimulation for loud sounds relative to soft sounds, lowering both threshold stimulus current intensity and excitation spread. Charge is injected through a first electrode and an opposite charge is injected on either a relatively distant electrode, or a set of adjacent electrodes, or a combination of the distant and adjacent electrodes. Electrodes are chosen from a look-up table on the basis of instantaneous frequency and magnitude of the sounds. Lower and higher sound levels result in more opposite charge being injected through the distant electrode and adjacent electrodes, respectively.—DAP

7,343,021

43.66.Ts OPTIMUM SOLUTION METHOD, HEARING AID FITTING APPARATUS UTILIZING THE OPTIMUM SOLUTION METHOD, AND SYSTEM OPTIMIZATION ADJUSTING METHOD AND APPARATUS

Hideyuki Takagi *et al.*, assignors to Rion Company, Limited
11 March 2008 (Class 381/313); filed in Japan 15 December 1999

A single set of signal processing parameters, in the form of one n-dimensional solution vector that is displayed visually, is determined for



several environmental sounds by using a genetic algorithm and subjective responses from a particular hearing aid wearer.—DAP

7,343,022

43.66.Ts SPECTRAL ENHANCEMENT USING DIGITAL FREQUENCY WARPING

James M. Kates, assignor to GN ReSound A/S
11 March 2008 (Class 381/316); filed 13 September 2005

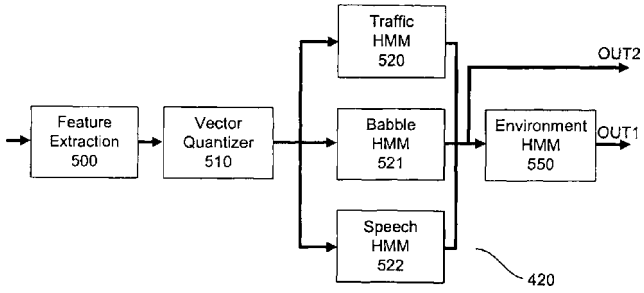
A filter bank for a hearing aid has cascaded all-pass sections to provide, without temporal aliasing, frequency warping that corresponds better to auditory filter bandwidths than fixed bandwidth FFT-based filters. Power spectra are calculated from the warped sequence of delayed samples, and multichannel compression and spectral enhancement gains are determined. An inverse frequency domain transform then produces a set of time-domain filter coefficients. The delayed speech segments are convolved with the compressed-spectrally enhanced filter coefficients in the warped time domain to produce the digital output signal.—DAP

7,343,023

43.66.Ts HEARING PROSTHESIS WITH AUTOMATIC CLASSIFICATION OF THE LISTENING ENVIRONMENT

Nils Peter Nordqvist and Arne Leijon, assignors to GN ReSound A/S
11 March 2008 (Class 381/321); filed in Denmark 4 April 2000

This patent covers automatic detection and classification of the wearer's acoustic listening environment and the resulting automatic adjustment of hearing aid signal processing parameters. Feature vectors associated with symbols are identified from recorded real-world sound signals in off-line



training. The feature vectors and symbols generated by vector quantization are processed with Hidden Markov Models to calculate the probability of sound sources occurring in an environment.—DAP

7,342,168

43.72.Ar SOUND EFFECTER, FUNDAMENTAL TONE EXTRACTION METHOD, AND COMPUTER PROGRAM

Masaru Setoguchi, assignor to Casio Computer Company, Limited
11 March 2008 (Class 84/619); filed in Japan 28 February 2005

The patent describes a signal processing mechanism for altering the pitch of a speech waveform to a rational multiple of the original. This is done by manipulating the frame-by-frame phase components of a short-time Fourier transform (STFT) of the signal. A greatest common divisor (GCD) method is used to determine the phase alteration algorithm. Because the phase adjustments proceed according to the GCD relationship, it is not necessary to determine the pitch of the original signal prior to performing the pitch-changing computations. In an alternative application of the same phase computations, the pitch of a signal may be determined based on the STFT phase data.—DLR

7,337,113

43.72.Ne SPEECH RECOGNITION APPARATUS AND METHOD

Kenichiro Nakagawa *et al.*, assignors to Canon Kabushiki Kaisha
26 February 2008 (Class 704/233); filed in Japan 28 June 2002

The goal is to improve speech recognition performance when the noise produced by a device changes. The present operating mode of a device is

OPERATING MODE	ACOUSTIC MODEL	POWER MODEL
FAX DATA TRANSMIT MODE	ACOUSTIC MODEL A	NOISE POWER A
FAX DATA RECEIVE MODE	ACOUSTIC MODEL B	NOISE POWER B
STANDBY MODE	ACOUSTIC MODEL C	NOISE POWER C

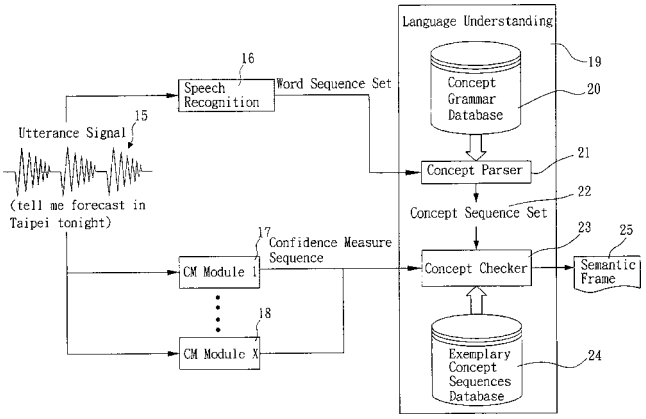
detected by a microphone monitoring the noise produced and memory is searched for stored information, including an acoustic model and noise power, that relate data for speech recognition to that operating mode. An instruction for speech recognition is issued based on this data. The sensor continues to monitor the operating noise made by the device for changes, which, if detected, trigger an update of the data.—DAP

7,333,928

43.72.Ne ERROR-TOLERANT LANGUAGE UNDERSTANDING SYSTEM AND METHOD

Huei-Ming Wang and Yi-Chung Lin, assignors to Industrial Technology Research Institute
19 February 2008 (Class 704/9); filed in Taiwan 31 May 2002

This speech recognition system uses three levels of processing, converting from the audio signal to word strings, from words to grammar structures, and from grammars to concept structures. At each level, the result is compared to a database of known valid patterns. As a way of measuring the closeness of fit, in each case, an edit sequence is produced that will edit the



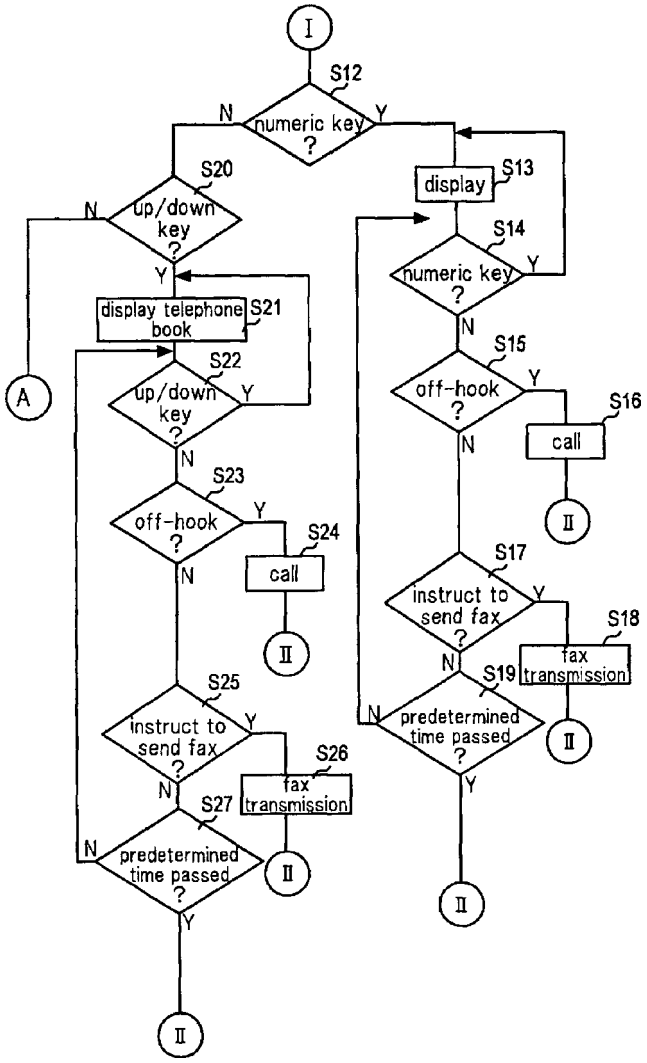
trial pattern so as to match the closest reference pattern. The edit sequences are then scored to determine the accuracy of the recognition process. Overall performance is good because any errors are corrected at each stage of the process.—DLR

7,340,041

43.72.Ne TELEPHONE DEVICE HAVING OPERATION FUNCTION BY VOICE INPUT

Hiroyuki Otagaki *et al.*, assignors to Sanyo Electric Company, Limited
4 March 2008 (Class 379/88.01);filed in Japan 17 November 2000

A user can also initiate either a telephone book search or a call simply by taking the handset off the hook and speaking. When an "off-hook" condition is detected for a handset, a first flag is set and the user's voice is recognized via voice recognition from voice input sent by the handset. The system retrieves and displays the stored users's name and telephone number from the telephone book memory. With the first flag still ON, when a second



flag is set, the same is done with a voice input from a different party. When a third voice input is detected, a call is originated using stored information for dialing.—DAP

7,334,478

43.80.Vj DEVICE FOR GUIDING THE MOVEMENT OF A TRANSDUCER OF AN ULTRASONIC PROBE

Won Soon Hwang, assignor to Medison Company, Limited
26 February 2008 (Class 73/618); filed in Republic of Korea 15 July 2005

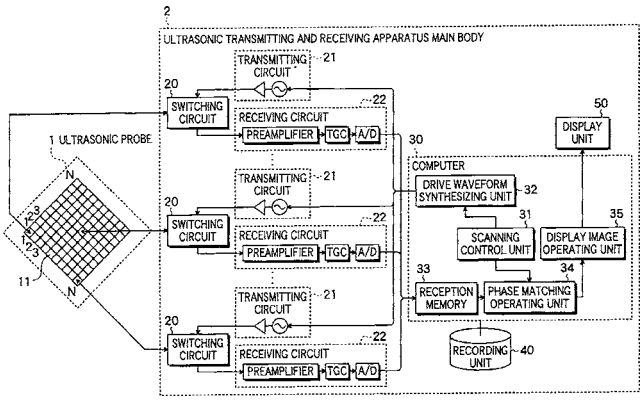
Ultrasonic transducer motion is guided to acquire a sequence of images that span three dimensions by using a pair of rails with a slot in each rail into which fit a pair of bearings that project from each of two opposite sides of the transducer.—RCW

7,335,160

43.80.Vj ULTRASONIC TRANSMITTING AND RECEIVING APPARATUS

Tomoo Satoh, assignor to FUJIFILM Corporation
26 February 2008 (Class 600/437); filed in Japan 6 November 2003

This apparatus includes an ultrasonic probe, transmitting circuitry, transmit-receive switches, echo receiving circuitry, display and recording units, and a computer that performs transmitting and receiving functions.



Multiple beams are simultaneously transmitted in different directions. Multiple beams are also formed from the received echo data. The architecture and operation of the apparatus facilitate imaging a volume in which structures are moving.—RCW

7,335,169

43.80.Vj SYSTEMS AND METHODS FOR DELIVERING ULTRASOUND ENERGY AT AN OUTPUT POWER LEVEL THAT REMAINS ESSENTIALLY CONSTANT DESPITE VARIATIONS IN TRANSDUCER IMPEDANCE

Todd A Thompson *et al.*, assignors to Timi 3 Systems, Incorporated
26 February 2008 (Class 601/2); filed 24 July 2002

Rules are used to deliver ultrasound energy at a constant level of power when the impedance or center frequency of the transducer in the system varies.—RCW

7,338,448

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND COMPOUND IMAGING WITH COMBINED FUNDAMENTAL AND HARMONIC SIGNALS

Xiaohui Hao *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
4 March 2008 (Class 600/443); filed 7 November 2003

Images are compounded using echo data obtained from different operating modes. The modes may be fundamental frequency, harmonic frequency, coded harmonic frequency, or variable frequency operation of the imaging system. The imaging mode may also be based on different steering angles that permit spatial compounding and help reduce grating lobe artifacts.—RCW

7,338,449

43.80.Vj THREE DIMENSIONAL LOCATOR FOR DIAGNOSTIC ULTRASOUND OR MEDICAL IMAGING

Wayne J. Gueck and John C. Lazenby, assignors to Siemens Medical Solutions USA, Incorporated
4 March 2008 (Class 600/447); filed 25 May 2004

A point is examined from two different viewing directions. The position of the point within a volume is determined from the intersection of two

lines each parallel to the viewing direction from which the image is obtained and passing through the selected point in each image.—RCW

7,338,450

43.80.Vj METHOD AND APPARATUS FOR PERFORMING CW DOPPLER ULTRASOUND UTILIZING A 2D MATRIX ARRAY

Kjell Kristoffersen and Glenn Reidar Lie, assignors to General Electric Company
4 March 2008 (Class 600/447); filed 27 August 2004

A continuous wave signal with a dithered component, such as a periodically varied amplitude or phase, is transmitted. Corresponding echo signals are received. The continuous wave received signal is processed to extract the Doppler shift for use during imaging.—RCW

7,338,451

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

Hirohiko Tsuzuki, assignor to Fujifilm Corporation
4 March 2008 (Class 600/458); filed in Japan 17 January 2001

Ultrasonic scatterers having an average particle size of 0.01–10 micrometers are injected into a body. An ultrasonic wave containing ten or more cycles is transmitted into a region of the body containing the scatterers. An ultrasonic wave containing four or more but less than ten cycles is then transmitted into the same region after a delay. The ten-cycle burst produces subharmonic echos that are detected by the shorter tone burst. Subharmonic processing in the system is designed to improve spatial resolution.—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.

Flexible cue use in nonnative phonetic categorization (L)

Mirjam Broersma[a)]
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, Radboud University, Nijmegen, 6500 HE The Netherlands

(Received 1 August 2007; revised 8 May 2008; accepted 9 May 2008)

Native and nonnative listeners categorized final /v/ versus /f/ in English nonwords. Fricatives followed phonetically long (originally /v/-preceding) or short (originally /f/-preceding) vowels. Vowel duration was constant for each participant and sometimes mismatched other voicing cues. Previous results showed that English but not Dutch listeners (whose L1 has no final voicing contrast) nevertheless used the misleading vowel duration for /v/-/f/ categorization. New analyses showed that Dutch listeners did use vowel duration initially, but quickly reduced its use, whereas the English listeners used it consistently throughout the experiment. Thus, nonnative listeners adapted to the stimuli more flexibly than native listeners did. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2940578]

PACS number(s): 43.71.Hw, 43.71.Es, 43.71.Sy [AJ] Pages: 712–715

I. INTRODUCTION

This paper investigates whether nonnative listeners can be more flexible than native listeners in their use of perceptual cues for making phoneme distinctions. In a previous study (Broersma, 2005), Dutch and English listeners categorized English obstruents at the end of nonwords as voiced or voiceless. Like English, Dutch has an obstruent voicing distinction, but unlike English, it has no voiced obstruents at the end of words in isolation (Booij, 1995). Thus, although Dutch listeners are familiar with the voicing distinction from their native language, they are not familiar with making this distinction in final position, or with the use of preceding vowel duration as a cue for this distinction. Preceding vowel duration is a perceptual cue for word-medial voicing distinctions in Dutch (Van den Berg, 1989), but a less prominent one than for the final voicing distinction in English (e.g., Raphael, 1972). Dutch listeners seem to be able to generalize their knowledge about vowel duration and word-medial obstruent voicing to word-final obstruents, as they categorize the duration of vowels differently depending on the underlying voicing of Dutch word-final obstruents (Jongman *et al.*, 1992).

Previous results (Broersma, 2005) showed that Dutch listeners accurately distinguished English final voicing contrasts, but that they did not use vowel duration in the same way English listeners did. In an experiment where vowel duration was made uninformative and even misleading, English listeners persisted in using this cue for making final /v/-/f/ distinctions, whereas Dutch listeners did not use the cue significantly. Vowel duration was made uninformative to preclude the possibility that the Dutch listeners used vowel duration only as a result of the nature of the stimulus materials. This leaves open the possibility, however, that Dutch listeners did *not* use vowel duration only as a result of the nature of the materials. Thus, the Dutch listeners might have more flexibly adapted to the information in the stimulus materials than the English listeners.

In the Broersma (2005) experiment, an 11-step /v/ to /f/ continuum was combined with a phonetically long and a phonetically short vowel. However, each participant heard only one of the vowels throughout the experiment. Thus, vowel duration was not informative and, moreover, mismatched the voicing information in the fricatives for some of the items, as v-like fricatives were sometimes preceded by a short vowel and f-like fricatives by a long vowel. In order to assess the use of vowel duration as a cue for final fricative voicing, 50-percent crossover points in the conditions with long and short preceding vowels were compared. For the English listeners, there was a significant shift in the 50-percent crossover point, with more "voiced" responses in the Long Vowel condition than in the Short Vowel condition. For the Dutch listeners, there was a small trend in the same direction, but this was not significant. Further, the Dutch listeners' categorization curve was steeper than the English listeners' curve in the Long Vowel condition. (There was no difference in the Short Vowel condition.)

In the same experiment, for a final /z/ to /s/ continuum, no shift in the 50-percent crossover points in the Long Vowel versus the Short Vowel condition was found, for the English

a)Electronic mail: mirjam.broersma@mpi.nl.

 0001-4966/2008/124(2)/712/4/$23.00

TABLE I. The 50-percent crossover point and steepness of the categorization curve at that point (with lower values indicating steeper slopes), for Dutch and English listeners, in the Long Vowel and Short Vowel conditions.

	50-percent crossover point				Steepness			
	Dutch		English		Dutch		English	
Part	Long	Short	Long	Short	Long	Short	Long	Short
1	6.7	5.8	6.9	5.0	0.4	0.4	0.5	0.2
2	6.4	5.9	7.7	5.9	0.3	0.4	0.4	0.3
3	6.3	5.5	6.9	5.8	0.4	0.3	0.5	0.3
4	6.4	5.8	7.2	5.3	0.3	0.3	0.5	0.3
5	6.4	6.0	7.0	5.4	0.3	0.4	0.5	0.3
6	6.2	6.0	6.0	5.0	0.4	0.3	0.4	0.3
1–6	6.4	5.8	7.0	5.4	0.4	0.3	0.5	0.3

or for the Dutch listeners. There are spectral differences between the alveolar and the labiodental stimuli that may explain the lack of such a shift for the /z/ to /s/ continuum. First, the difference in intensity between voiced and voiceless fricatives in the frequency range important for voicing was larger for the alveolar than for the labiodental fricatives (intensity of the first spectral peak below 500 Hz, measured from 10 ms at the center of the fricative, with means of fast Fourier transform using a Gaussian window: /z/-/s/: 29.5 −9.5=20.0 dB; /v/-/f/: 24.7−10.9=13.8 dB). Further, the overall intensity of the alveolar fricatives was higher than that of the labiodental fricatives (/z/: 67.7; /s/: 64.0; /v/: 65.8; /f/: 57.9 dB) (cf. Jongman *et al.*, 2000), and that of the preceding vowels lower (alveolars, long vowel: 76.1, short vowel: 66.5 dB; labiodentals, long vowel: 77.9, short vowel: 69.0 dB). Thus, as the alveolar fricatives contained larger amplitude differences related to voicing, and their frication was (both absolutely and relatively to the vowel) more easily perceptible than that of the labiodental fricatives, listeners may have been less inclined to exploit vowel duration as a cue to voicing for the alveolar than for the labiodental fricatives.

For the /v/ to /f/ continuum, the Dutch listeners might have discovered during the experiment that vowel duration was not a helpful cue to final fricative voicing here and learned to ignore it. This paper attempts to test this explanation by establishing whether the listeners did not use vowel duration from the start of the experiment, or tried to use vowel duration as a voicing cue initially, but stopped doing so at some point in the experiment because of the nature of the stimulus materials. Relevant data could be obtained from the results of the practice part of the study, which were not taken into account in Broersma (2005).

II. METHOD

In a Two-Alternative Forced Choice experiment, listeners categorized sounds as "v" or "f." Participants were 28 native speakers of English, and 28 native speakers of Dutch who were proficient in English as a second language.

The materials consisted of 11 fricatives spliced onto two carriers. The carriers were one token of /ku:/ with a phonetically long vowel (extracted from a recording of /ku:v/), and one with a phonetically short vowel (extracted from a recording of /ku:f/). The fricatives formed a continuum from a natural /v/ to a natural /f/ (extracted from another recording of /ku:v/ and /ku:f/) with nine intermediate steps, created following the procedure of Stevenson (1979) and Repp (1981) by adding up the waveforms of the /v/ and the /f/ in varying, equally spaced proportions.

In the practice part, participants categorized each step of the continuum three times, then there was a break during which they could ask questions, and finally they categorized each step one more time. After that, the main part of the experiment, consisting of 20 presentations of each step, started without any further demarcation. All items, in the practice part as well as the main part of the experiment, were presented in semirandomized order. Items in the Long Vowel condition contained the phonetically long vowel and those in the Short Vowel condition the phonetically short vowel. Participants were randomly assigned to a condition, with equal numbers in both conditions. Each participant thus heard only the long vowel or only the short vowel throughout the entire experiment. For more details about the method, see Broersma (2005).

III. RESULTS

Each subject's categorization curve was fitted with logistic regression to determine the 50-percent crossover point and the steepness of the curve at that point. First, the results of the practice part and the main part of the experiment were analyzed together (Table I, part 1-6), with an analysis of variance (ANOVA) with 50-percent crossover point as the dependent variable and vowel duration and native language as independent variables. There was a significant interaction between vowel duration and native language ($F(1,41)=5.67, p<0.05$), showing that the effect of vowel duration was larger for the English listeners than for the Dutch listeners. However, the main effect of vowel duration was significant not only for the English listeners ($F(1,18)=32.51, p<0.001$), but also for the Dutch listeners ($F(1,23)=4.74, p<0.05$). Thus, in contrast to Broersma (2005), when the practice part was included, Dutch listeners showed significant evidence of the use of vowel duration for final fricative voicing decisions.

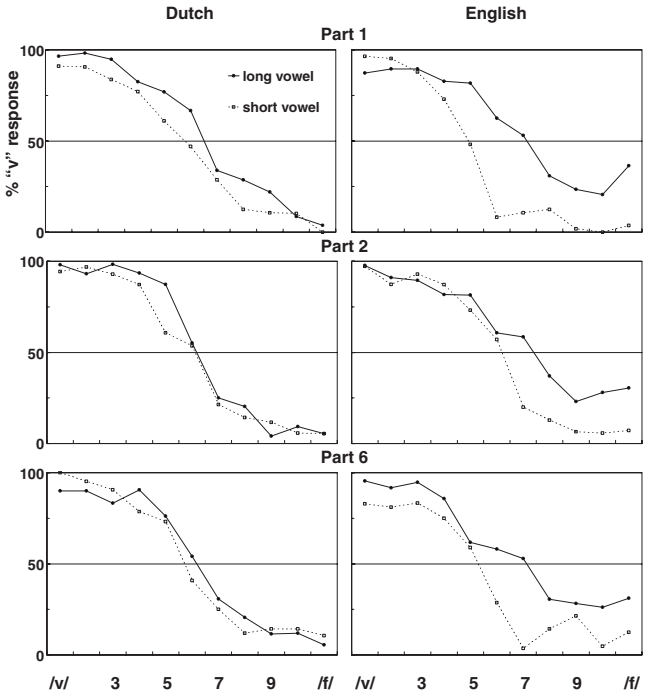


FIG. 1. Mean percentage of "v" responses as a function of the place on an 11-step /v/ to /f/ continuum and preceding vowel duration, per experimental part (1, 2, and 6) and participants' native language.

Next, the practice part was analyzed separately (Fig. 1 and Table I, part 1). There was a main effect of vowel duration ($F(1,47)=15.15, p<0.001$), no main effect of native language ($F(1,47)<1$), and no interaction between native language and vowel duration ($F(1,47)=1.84, p>0.1$). Thus, in the practice part, Dutch and English listeners' responses depended similarly on the duration of the preceding vowel.

Finally, the change in the use of vowel duration over time, in particular from the practice part to the main part of the experiment, was assessed. The main part of the experiment was split into parts of the same length as the practice part (44 trials). The practice part was called Part 1, and the main part was divided into Parts 2–6, with Part 2 containing the first 44 trials of the main part, Part 3 the next 44 trials, and so on. Part 6 contained only 33 trials. There was a three-way interaction among Part, vowel duration, and language ($F(5,205)=2.40, p<0.05$). Crucially, comparing the results of the practice (Part 1) to those of Part 2 (Fig. 1 and Table I, part 2), for the Dutch listeners, there was an interaction between Part and vowel duration ($F(1,25)=4.46, p<0.05$). For these listeners, the effect of vowel duration was significant in Part 1 ($F(1,25)=5.63, p<0.05$), but not in Part 2 ($F(1,26)=1.2, p>0.1$). For the English listeners, there was no interaction between Part and vowel duration for Part 1 versus 2 ($F(1,21)=1.97, p>0.1$). Comparing all consecutive Parts in a similar, pairwise manner (i.e., Part 2 with 3, 3 with 4, 4 with 5, and 5 with 6), there were no other interactions between Part and vowel duration for the Dutch or for the English listeners. Thus, for the Dutch listeners, the use of vowel duration decreased from Part 1 to Part 2, and then stayed the same until the end of the experiment, while for the English listeners the use of vowel duration did not change throughout the experiment.

To assess how categorical the listeners' perceptual performance was, ANOVAs on the steepness of the categorization curves were performed. Recall that in the main part of the experiment, the Dutch listeners' categorization curve was steeper than the English listeners' curve in the Long Vowel condition, and there was no difference in the Short Vowel condition. In the practice part, there was also an interaction between native language and vowel duration ($F(1,47)=5.88, p<0.05$). Here, however, the English listeners' categorization curve was steeper than the Dutch listeners' curve in the Short Vowel condition ($F(1,24)=6.54, p<0.05$), and there was no difference in the Long Vowel condition ($F(1,23)<1$). Comparing Part 1 and Part 2, there was an interaction between Part and native language ($F(1,46)=5.40, p<0.05$), and there were no such interactions for any other set of consecutive Parts. Thus, whereas the Dutch listeners' perception was less categorical than the English listeners' perception in the practice part, it was more categorical than the English listeners' responses in the main part of the experiment.

IV. DISCUSSION

The analysis of the practice part shows that the Dutch listeners initially used vowel duration as a cue for final /v/-/f/ categorization to the same extent as English listeners did. However, their use of it rapidly diminished over time. The English listeners, on the other hand, did not change their use of vowel duration as a voicing cue throughout the experiment. The English listeners, who initially categorized (some of) the fricatives more categorically, later categorized (some of) them less categorically than the Dutch listeners: In the practice part they perceived the fricatives in the Short Vowel condition more categorically than the Dutch listeners did, whereas in the rest of the experiment they perceived the fricatives in the Long Vowel condition less categorically than the Dutch listeners did. This change occurred at the same time that the Dutch listeners changed their use of vowel duration (between the practice part and the first 44 trials of the main part of the experiment), and may be related to the Dutch listeners' decreased use of this misleading cue and the English listeners' persistent use of it. This confirms the conclusion from Broersma (2005) that the nonnative listeners did not use vowel duration as a voicing cue in the same way the English listeners did. It seems, however, that they were no less capable of using it, but that they adapted to the nature of the stimulus materials better than the native listeners did. Generally, nonnative listeners might be less certain about which perceptual cues to use than native listeners are. Further, the Dutch listeners' limited experience with the use of vowel duration as a voicing cue might have made it easier for them to ignore this cue than for the native listeners, with their extensive experience with it.

While insecurity about the perceptual relevance of phonetic information and less practice with a perceptual cue might sometimes hinder speech perception, in this study, the nonnative listeners' greater flexibility proved to be an advantage over the native listeners' firmer and presumably more secure cue weighting strategies. For nonnative listeners, who

are trying to work out how phonetic information indicates phonological distinctions in a particular second language, which may be very different from their native language, a large degree of flexibility indeed seems useful.

ACKNOWLEDGMENTS

Many thanks to Sarah Schimke for suggesting these analyses, to Anita Wagner for help with spectral measurements, and to two anonymous reviewers for helpful comments.

Booij, G. (**1995**). *The Phonology of Dutch* (Oxford University Press, Oxford).

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Jongman, A., Sereno, J. A., Raaijmakers, M., and Lahiri, A. (**1992**). "The phonological representation of [voice] in speech perception," Lang Speech **35**, 137–152.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Repp, B. H. (**1981**). "Perceptual equivalence of two kinds of ambiguous speech stimuli," Bull. Psychon. Soc. **18**, 12–14.

Stevenson, D. C. (**1979**). *Categorical Perception and Selective Adaptation Phenomena in Speech*. Doctoral dissertation, University of Alberta, Edmonton, Canada.

Van den Berg, R. J. H. (**1989**). "Perception of voicing in Dutch two-obstruent sequences: Covariation of voicing cues," Speech Commun. **8**, 17–25.

Simultaneous production of low- and high-frequency sounds by neonatal finless porpoises (L)

Songhai Li, Kexiong Wang, and Ding Wang[a)]
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, 430072, People's Republic of China

Shouyue Dong
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, 430072, People's Republic of China and Graduate School of the Chinese Academy of Sciences, Beijing, 100039, People's Republic of China

Tomonari Akamatsu
NRIFE, Fisheries Research Agency, Hasaki, Kamisu, Ibaraki 314-0408, Japan

(Received 4 December 2007; revised 6 May 2008; accepted 19 May 2008)

Phocoenids are generally considered to be nonwhistling species that produce only high-frequency pulsed sounds. Here our results show that neonatal finless porpoises (*Neophocaena phocaenoides*) frequently produce clear low-frequency (2–3 kHz) pulsed signals, without distinct high-frequency energy, just after birth and can produce both low- (2–3 kHz) and high-frequency (>100 kHz) pulsed signals simultaneously until about 20 days postnatal. The results indicate that low-frequency signals of neonatal finless porpoises are not an early form of high-frequency signals and suggest that low- and high-frequency signals may be produced by different sound production mechanisms.
 [DOI: 10.1121/1.2945152]

PACS number(s): 43.80.Ka [WWA] Pages: 716–718

I. INTRODUCTION

Most studies of odontocete vocalization have focused on tonal whistles, often less than 20 kHz in energy, and pulsed sounds, with energy extending into the ultrasonic region (Popper, 1980; Au, 1993; Matthews *et al.*, 1999). Whistles are thought to be used exclusively for communication, whereas pulsed signals are used for both echolocation and communication (Popper, 1980; Herman and Tavolga, 1980). Not all odontocetes whistle, but all produce some form of pulsed sound (Popper, 1980). Phocoenids, cephalorhynchids, and *Kogia* spp. generally produce only high-frequency pulsed sounds. Although some studies have investigated the two-component sonar of harbor porpoises, which has a low-frequency component of about 20 kHz (Kamminga and Wiersma, 1981) or 2 kHz (Andersen and Amundin, 1976), the low-frequency component generally overlies the high-frequency signal, and the underlying mechanism may be "a linear process of addition" (Kamminga and Wiersma, 1981). However, bottlenose dolphins emit whistles and high-frequency pulsed sounds simultaneously (Lilly and Miller, 1961), suggesting the presence of at least two different sound production mechanisms, one for whistles and the other for pulsed sounds. Neither an unattached low-frequency vocalization repertoire nor multimechanism sound production has been observed in phocoenids. However, our recent vocalization recordings described here indicate that neonatal Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*) produce unattached low-frequency pulsed signals and can emit both low- and high-frequency pulsed signals simultaneously, suggesting the possibility of separate sound production mechanisms.

II. MATERIALS AND METHODS

We recorded the vocalizations of two captive neonatal Yangtze finless porpoises: Terry, born on 5 July 2005 at night, and Tommy, born on 2 June 2007 in the afternoon. Both males shared the same mother. The recordings were made within a wide frequency range from 100 Hz to 147 kHz (Li *et al.*, 2007). The vocalizations of Terry were recorded while the neonate was living with his mother (who was ~8 years old) and an adult female tank-mate (~11 years old) in a 3 m deep, 25×7 m kidney-shaped pool. Tommy was recorded while living with his mother and older brother, Terry, in the same pool for the first 39 days postnatal, and then alone in a round pool about 2 m deep and 6.5 m in diameter.

We used sound recordings and ad lib notation with event-based sampling methods to document sound production and the behavioral context. A hydrophone (ST1020, Oki Electric Co. Ltd., Tokyo, Japan; sensitivity: −180 dB re: 1 V μPa^{-1} +3/−12 dB, to 150 kHz), input the porpoise vocalizations to an underwater sound level meter (SW1020, Oki) and a Sony PCHB244 digital data recorder, which had a flat frequency response from dc to 147 kHz within 3 dB. The hydrophone was located at 0.5 m depth and 0.5 m from the tank wall. A 100 Hz high-pass filter was incorporated into the vocalization recordings. Sound analysis was conducted using PC-based DTDisk™ (Direct-to-Disk Recorder, Version 1.11,

[a)]Author to whom correspondence should be addressed. Electronic mail: wangd@ihb.ac.cn

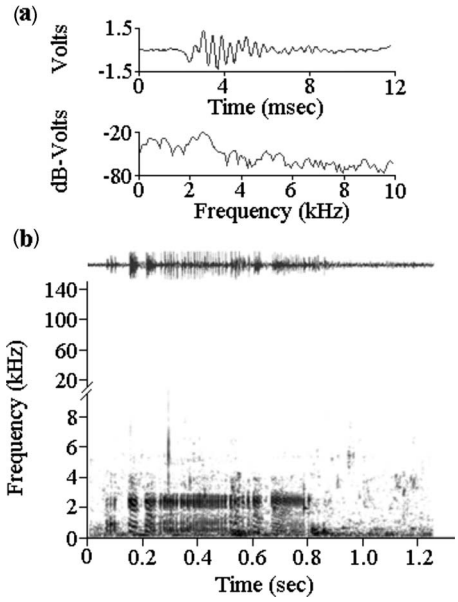


FIG. 1. Typical low-frequency pulsed sound produced by neonatal finless porpoises within several hours of birth. (a) Waveform (top panel) and spectrum (bottom panel) of a representative sound pulse extracted from the sound in (b). (b) Waveform (top panel) and spectrogram (bottom panel) of a representative low-frequency pulsed sound. No distinct energy extends over 10 kHz in the spectrogram. The sound was high-pass filtered at 100 Hz to reduce background noise. The spectrum and spectrogram were calculated at 8192 FFT (fast Fourier transport) (sampling rate=500 kHz, frequency resolution=61 Hz, Hanning windows).

November 2002; American Engineering Design) and SIGNAL™ software (Version 4.03, December 2005; American Engineering Design). Porpoise sound productions were noted on the recordings and subsequently located using the cursor option. To identify the vocalizing animal when the neonates resided with other individuals, we analyzed only those instances when we observed only one individual vocalizing within 2 m and directly at the hydrophone with head scanning motions (Li *et al.*, 2007) or bubbling.

III. RESULTS

During the first few hours after the porpoises were born, we frequently recorded low-frequency pulsed sounds without distinct high-frequency energy, lasting about 3–5 ms (Fig. 1). The peak frequencies were ~2–3 kHz, and the received sound pressure levels (SPLs) were ~130–134 dB re 1 μPa pp. During the vocalizations of Tommy, who was born in the daytime, we often noticed bubbling behavior associated with the presence of low-frequency sounds.

After about 20 days postnatal, both of the neonates could produce low- and high-frequency pulsed sounds simultaneously. The low-frequency sounds were similar to those produced by the neonates just after birth in both waveform characteristics and energy distribution (Figs. 1 and 2), but were much weaker in SPLs, which were between 116 and 122 dB re 1 μPa pp. The peak frequencies of the high-frequency pulsed sounds, which lasted less than 100 μs, were >100 kHz [Fig. 2(b)], and the SPLs were ~150 dB re 1 μPa pp. No overlay relationship existed between the low- and high-frequency pulsed signals, but the two sets of pulse trains showed different timing (Fig. 2).

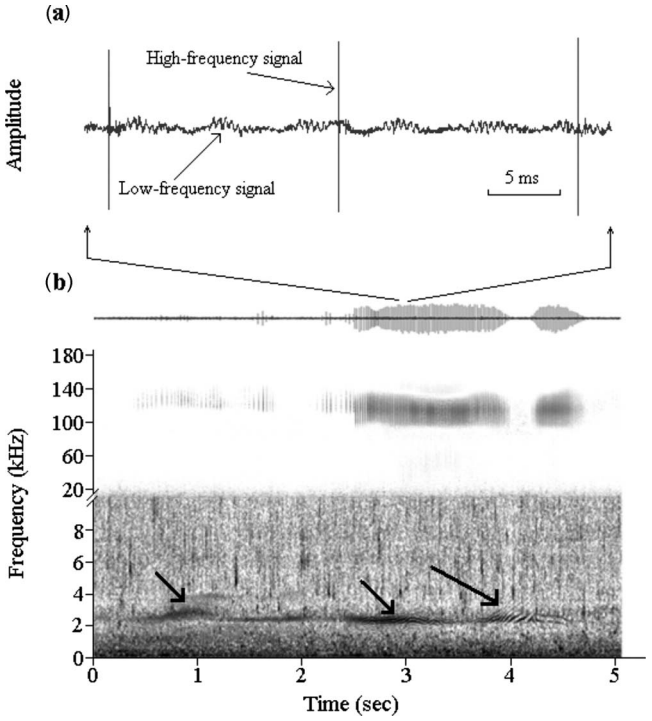


FIG. 2. Simultaneous production of low- and high-frequency pulsed sounds by neonatal finless porpoises until a maximum of about 20 days postnatal. (a) Zoom of a segment of the waveform in (b) (top panel), showing both low- and high-frequency signals. Note the much weaker (>20 dB) low-frequency signals, compared to high-frequency sounds, and the lack of an overlay relationship between low- and high-frequency pulsed signals, but the different timing between the two sets of pulse trains. The low- and high-frequency signals are denoted by black arrowheads. (b) Waveform (top panel) and spectrogram (bottom panel) of a representative sound produced by neonatal finless porpoises after about 20 days postnatal. Note the low-frequency energy indicated by black arrowheads in the spectrogram. The sound was high-pass filtered at 100 Hz to reduce background noise. The spectrogram was calculated at 8192 FFT (sampling rate=500 kHz, frequency resolution=61 Hz, Hanning windows).

IV. DISCUSSION

During the postnatal development of some bats (Brown and Grinnell, 1980; Moss, 1988; Moss *et al.*, 1997) and nonphocoenid odontocetes (Reiss, 1988; Madsen *et al.*, 2003), the vocalization of echolocation signals tends to rise in frequency and decrease in duration. Our data show that neonatal Yangtze finless porpoises emit low-frequency, long-duration pulsed sounds just after birth. However, by 20 days postnatal at the latest, they can produce both low- and high-frequency pulsed sounds simultaneously. The high-frequency pulsed sounds are indistinguishable from adult echolocation clicks in both temporal and frequency domains (Li *et al.*, 2007), while the simultaneous low-frequency signals are similar to those produced by the neonates just after birth both in waveform characteristics and energy distribution (Figs. 1 and 2). These results indicate that the low-frequency pulsed sounds are not early forms of the high-frequency sounds. As no inchoate form of the high-frequency pulsed sounds was found in the neonatal finless porpoises, the development of echolocation signals in this species may represent a pathway different from that of bats and nonphocoenid odontocetes. Because no overlay relationship exists between the low- and high-frequency pulsed signals, but the timing is different be-

tween the two pulse train sets (see the waveform and spectrogram in Fig. 2), we suggest that the low- and high-frequency signals may not originate in the same sound production mechanism and that low-frequency signals are not a "by-product" of high-frequency sounds, but rather are generated by a separate anatomical structure. Thus, neonatal finless porpoises may possess at least two different sound production mechanisms.

Unlike terrestrial mammals, including bats, odontocetes do not produce sounds in the larynx but in the nasal passages above the larynx (Ridgway *et al.*, 1980; Mackay and Liaw, 1981). The whistling odontocete species may have two different sound production mechanisms, one for whistles and the other one for pulsed sounds (Popper, 1980). The pulsed sounds are produced within the nasal system by manipulating airflow through the dorsal bursae/monkey lips complex (Cranford, 1988, 2000). Since neonatal finless porpoises can produce both low- and high-frequency pulsed sounds simultaneously, with a clearly observed timing difference between the two sets of pulse trains, they might also possess two different sound production mechanisms. After about 20 days postnatal, the SPLs of the low-frequency sounds broadcast simultaneously with the high-frequency sounds were much weaker ($>$10 dB) than those of the low-frequency sounds produced alone just after birth. Thus, the low-frequency pulsed sounds might be produced by an apparatus above the dorsal bursae/monkey lips complex. In the first developmental stage of neonatal finless porpoises, the dorsal bursae/monkey lips complex is most likely not fully developed and may not be able to hold up airflow to produce high-frequency pulsed sounds for echolocation and communication. Instead, it might produce relatively strong low-frequency sounds directly by the unabated airflow in a sound production site above the monkey lips complex. About 20 days after birth, the monkey lips complex may have developed sufficiently to obstruct the airflow, thus producing relatively strong high-frequency pulsed sounds with SPLs of about 150 dB re 1 μPa pp. The neonate would then only be able to produce relatively weak low-frequency sounds [Fig. 2(a)] by the prostrate airflow in the sound production site above the monkey lips complex. However, the exact site and mechanism for production of the low-frequency signals have yet to be investigated.

As the low-frequency sounds of neonatal finless porpoises are of substantially low amplitude and frequency but long duration, these sounds are unlikely to function in echolocation. However, they may serve as calls to the mother by the infants. Our qualitative observations support this idea, as the neonates emit low-frequency sounds more frequently when they are apart from the mother. Thus, one advantage of preserving this kind of low-frequency sound in infancy could be that neonatal finless porpoises are capable of calling their mothers without fully developed echolocation.

ACKNOWLEDGMENTS

We thank Y. Xian for data collection and Q. Zhao, B. Yu, and Y. Hao for animal management support. This work was supported by the National Basic Research Program of China (2007CB411606), the Chinese National Natural Science Foundation (30730018), the President's Fund of the Chinese Academy of Sciences, and Special Funds for Presidential Scholarships of the Chinese Academy of Sciences (082Z01).

Andersen, S. H., and Amundin, M. (**1976**). "Possible predator-related adaption of sound production and hearing in the Harbour Porpoise (*Phocoena phocoena*)," Aquat. Mamm. **4**, 56–58.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Brown, P. E., and Grinnell, A. D. (**1980**). "Echolocation ontogeny in bats," in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 355–377.

Cranford, T. (**1988**). "The anatomy of acoustic structures in the spinner dolphin forehead as shown by X-ray computed tomography and computer graphics," in *Animal Sonar Systems: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 67–77.

Cranford, T. (**2000**). "In search of impulse sound sources in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 109–155.

Herman, L. M., and Tavolga, W. N. (**1980**). "The communication systems of Cetaceans," in *Cetacean Behavior*, edited by L. M. Herman (Wiley, New York), pp. 149–209.

Kamminga, C., and Wiersma, H. (**1981**). "Investigations on cetacean sonar. II. Acoustical similarities of differences in odontocete sonar signals," Aquat. Mamm. **8**, 41–62.

Li, S., Wang, D., Wang, K., Xiao, J., and Akamatsu, T. (**2007**). "The ontogeny of echolocation in a Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*)," J. Acoust. Soc. Am. **122**, 715–718.

Lilly, J. C., and Miller, A. M. (**1961**). "Sounds emitted by the bottlenosed dolphin," Science **133**, 1689–1693.

Mackay, R. S., and Liaw, C. (**1981**). "Dolphin vocalization mechanisms," Science **212**, 676–678.

Madsen, P. T., Carder, D. A., Møhl, B., and Ridgway, S. H. (**2003**). "Sound production in neonate sperm whales," J. Acoust. Soc. Am. **113**, 2988–2991.

Matthews, J. N., Rendell, L. E., Gordon, J. C. D., and Macdonald, D. W. (**1999**). "A review of frequency and time parameters of cetacean tonal calls," Bioacoustics **10**, 47–71.

Moss, C. (**1988**). "Ontogeny of vocal signals in the big brown bats, *Eptesicus fuscus*," in *Animal Sonar Systems: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 115–120.

Moss, C. F., Redish, D., Gounden, C., and Kunz, T. H. (**1997**). "Ontogeny of vocal signals in the little brown bat, *Myotis lucifugus*," Anim. Behav. **54**, 131–141.

Popper, A. N. (**1980**). "Sound emission and detection by delphinids," in *Cetacean Behavior*, edited by L. M. Herman (Wiley, New York), pp. 1–52.

Reiss, D. (**1988**). "Observation on the development of echolocation in young bottlenose dolphins," in *Animal Sonar Systems: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 121–128.

Ridgway, S. H., Carder, D. A., Green, R. F., Gaunt, A. S., Gaunt, S. L. L., and Evans, W. E. (**1980**). "Electromyographic and pressure events in the nasolaryngeal system of dolphins during sound production," in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 239–250.

Application of Hamiltonian of ray motion to room acoustics (L)

Sin'ichiro Koyanagi, Takeru Nakano, and Tetsuji Kawabe[a)]
Physics Department, Department of Acoustic Design, Kyushu University, Shiobaru, Fukuoka, 815-8540, Japan

(Received 30 December 2007; revised 27 May 2008; accepted 28 May 2008)

Based on a standard Hamiltonian of acoustic ray, it is shown that a ray motion in a finite region can be treated as a particle motion inside a potential well. The boundary reflections of ray can be described by introducing a so-called confining potential to confine a ray motion in a closed domain. It is shown that the square well potential model for the ray motion can reproduce the reverberation time in a two-dimensional room with irregular walls which is consistent with the Norris–Eyring law. It is also shown that the sound reverberation relates the ray chaos of the billiards in polygons with smooth convex walls. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2946714]

PACS number(s): 43.55.Br, 43.20.Dk, 43.25.Rq, 43.30.Cq [NX] Pages: 719–722

For the propagation of sound wave with the high-frequency limit, the acoustic ray equation can be formulated by the Hamiltonian form[1] consistent with the Helmholtz wave equation. The advantage of the Hamiltonian form is that we can study the ray propagation under the various circumstances including ray chaos in a similar manner as done for many dynamical systems with chaotic properties.[2] For the study of the conservative dynamical systems, the Hamiltonian formalism is essential to understand the behavior of the trajectories in the phase space determined by the potential function V. Such an analysis is possible when the Hamiltonian for dynamical systems with N degrees of freedom is of the form

$$H = \frac{1}{2}(p_1^2 + \cdots + p_N^2) + V(q_1, \ldots, q_N), \tag{1}$$

where the q's and p's are, respectively, the coordinates and the conjugate momenta of the system satisfying the Hamilton equations of motion as $dq_i/dt = \partial H/\partial p_i$ and $dp_i/dt = -\partial H/\partial q_i = -\partial V/\partial q_i$. Henceforth, we refer to the form of the Hamiltonian (1) as the standard Hamiltonian.[3] Once such a standard Hamiltonian is given, it is possible to study the ray propagation in the finite regions by properly manipulating the potential function.

In this paper, we try to apply the standard Hamiltonian equivalent to the Helmholtz wave equation to the acoustic problems. Using the potential model for the ray motion derived from this standard Hamiltonian, we study the acoustic phenomena such as the ray propagation and the ray chaos in a finite region, in an analogous way as other dynamical systems. Especially, we apply this model to the room with an absorption uniformly distributed on the boundary and show the characteristic decay time of ray energies to be consistent with the Norris–Eyring law.

a)Author to whom correspondence should be addressed. Electronic mail: kawabe@design.kyushu-u.ac.jp

The standard Hamiltonian we consider[4] is

$$H = \frac{\boldsymbol{p}^2}{2} + V(\boldsymbol{x}) = \frac{\boldsymbol{p}^2}{2} + \frac{-1}{2c^2(\boldsymbol{x})}, \tag{2}$$

with the constraint $H=0$. Here, $\boldsymbol{p} \equiv \boldsymbol{k}/\omega$ is the momentum defined by the wave number vector $\boldsymbol{k}$ and the angular frequency ω, $\boldsymbol{x}$ is the position vector, and $c(\boldsymbol{x})$ is the sound speed. The Hamilton equations of motion are

$$\frac{d\boldsymbol{x}}{d\tau} = \frac{\partial H}{\partial \boldsymbol{p}}, \quad \frac{d\boldsymbol{p}}{d\tau} = -\frac{\partial H}{\partial \boldsymbol{x}}, \tag{3}$$

where the scaled time τ instead of t, $d\tau = c^2 dt$, is used to retain the standard form of Hamilton equations of motion as $d\boldsymbol{x}/d\tau = \boldsymbol{p}$ and $d\boldsymbol{p}/d\tau = -\partial V/\partial \boldsymbol{x}$.

In order to construct the dynamical system describing the ray motion in the finite region, we need to consider the meaning of the constraint $H=0$ in Eq. (2). This constraint raises the problem that the potential V of Eq. (2) cannot confine the ray trajectories in the finite region. To cope with this problem, we have to add a so-called confining potential $U(\boldsymbol{x})$ to H of Eq. (2) to confine them. With the total potential function $V+U$, the ray motion can be completely determined by the standard Hamiltonian (2), once the sound speed $c(\boldsymbol{x})$ in V is given.

Since the confining potential $U(\boldsymbol{x})$ needs to describe the reflection of ray trajectories at the boundary, its functional form should ensure the behavior of a square well-type potential. As a concrete example of $U(\boldsymbol{x})$ for the ray motion in the $L_x \times L_y$ rectangle region, we choose the expression $U_4(x,y) = U(\boldsymbol{x})$ as follows:

$$U_4 = a(e^{m(x-L_x)} + e^{-m(x+L_x)} + e^{m(y-L_y)} + e^{-m(y+L_y)}), \tag{4}$$

where a is a constant with the same dimension as $V(\boldsymbol{x})$ and m is a large enough number to ensure a square well-type behavior, i.e., $U \to 0$ for $|x| \le L_x$ and $|y| \le L_y$ while $U \to \infty$ for otherwise. Figure 1 shows the ray trajectory for a linearly y-dependent sound speed $c(x,y) = c_0 + c_1 y$, where c_0 is a reference sound speed, and c_1 is the coefficient of the y-dependent sound speed. As expected from physical viewpoint, the trajectory curves downward as y increases. This result is qualitatively the same as one calculated by the

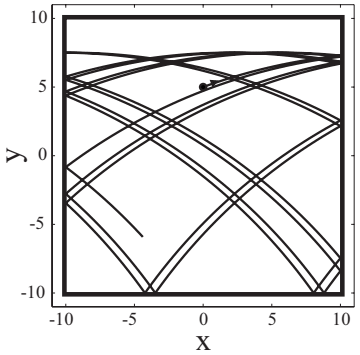


FIG. 1. Ray trajectory for a linearly increasing sound speed, with $c_0=340$ m/s, $c_1=10$. The initial conditions are $(x_0,y_0)=(0,5)$, $\theta_0=\tan^{-1}(p_{y0}/p_{x0})=20°$, and the parameters of Eq. (7) are $L_x=L_y=10$ m, $a=1/(2c_0^2)$, and $m=1000$.

Hamiltonian $H=cp$ with the background function[5] which is introduced into the sound-speed profile to confine all rays in a closed space. For the reflections on the boundary in the irregular and polygonal region more complicated than the rectangle, it is possible to realize them by adding appropriate expressions of U to V in Eq. (2).

Now let us study the ray chaos related to sound reverberation in the room acoustics. For this purpose, we consider two kinds of irregular shapes with five corners designed by U_{5a} and U_{5b}, respectively, as

$$U_{5a}=ae^{m(x+y-14)}+\hat{U}_4, \tag{5}$$

$$U_{5b}=ae^{m(-(x-10)^2-(y-10)^2+36)}+\hat{U}_4, \tag{6}$$

$$\hat{U}_4=a(e^{m(3x-y-25)}+e^{m(-10x-3y-100)}+e^{-m(y+10)}+e^{-m(x-\sqrt{2}y+10)}), \tag{7}$$

where U_{5a} describes the pentagonal domain with straight walls and U_{5b} is a modified version of U_{5a} with straight walls and a convex wall. Hereafter, we call this convex wall the diffusive wall. Figures 2(a) and 2(b) show the ray trajectories calculated from the same initial condition for the case of a constant sound speed $c(x,y)=c_0$, whose behavior seems quite irregular and complex. Indeed, we can confirm these complex behaviors from the Poincaré surface of section (PSS) on the (y,p_y) plane as shown in Figs. 2(c) and 2(d). Here, the PSS is defined by a projection of y and p_y values on the (y,p_y) plane, at whose values the condition that $x=0$ and $dx/d\tau>0$ is satisfied. From randomly 1.0×10^4 scattered points in the PSS of Figs. 2(c) and 2(d), which are calculated from 3.7×10^4 reflections of the trajectories on the walls, we see that both shapes produce the diffusive behavior of acoustic rays. This result seems to be consistent with the common idea for irregular shapes modeling the reverberation rooms.

The room acoustics are characterized by the existence of absorption on the boundaries of the room and thus of reverberation.[6–10] Since the ray trajectories of Fig. 2 are confined in the square well-type potential, they will describe the ray motion in the rooms with specular reflecting walls. In order to account for the decay of sound in the rooms due to absorption, we estimate the characteristic decay time T_{dec}

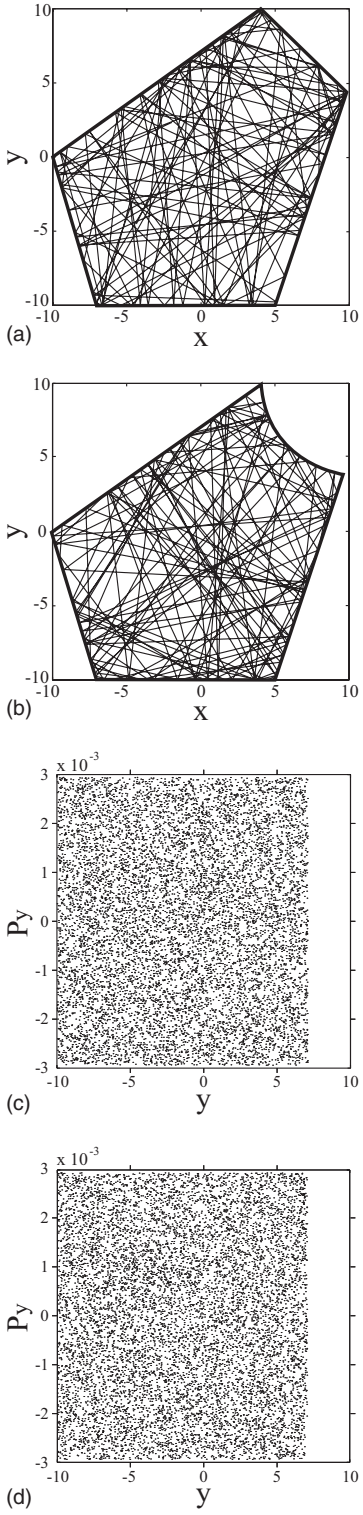


FIG. 2. Ray trajectories and their PSS for the case of the irregular domain: (a, c) for U_{5a}, and (b, d) for U_{5b}. The number of PSS is 1.0×10^4 confined in $y=[-10,5\sqrt{2}]$ and $p_y=[-3,3]\times10^{-3}$, and $N_{\mathrm{total}}=2.7\times10^9$.

defined by $\exp(-t/T_{\mathrm{dec}})$, relating to the reverberation time T_{rev} by $T_{\mathrm{rev}}=(6\log10)T_{\mathrm{dec}}\approx13.81T_{\mathrm{dec}}$.[8,9] For the calculation of T_{dec}, initially, we emit N_0 rays isotropically from a sound source at (x_0,y_0) in a room, and then allow them to propagate in absence of absorption on the wall for a long time T_{erg} to realize the uniform distribution of acoustic energy. After T_{erg}, we switch on the absorption on the wall, whose moment

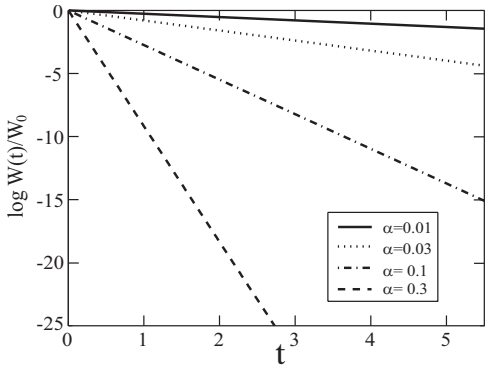


FIG. 3. Total acoustic energy $W(t)$ for several absorption coefficients α in the case of U_{5a}. The initial conditions are $N_0=3.0\times10^3$, $(x_0,y_0)=(0,0)$, $N_{\text{total}}=1.2\times10^7$. $T_{\text{erg}}=50\bar{l}/c_0=1.8$ is determined after 3.1×10^6 iterations.

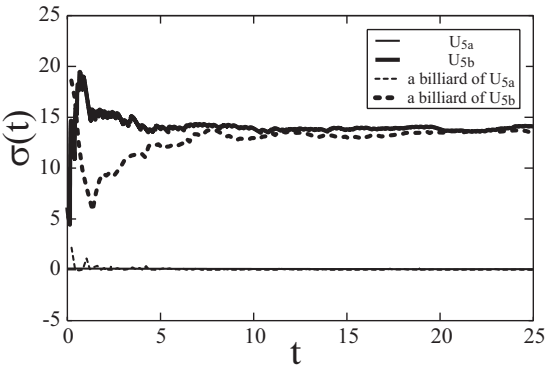


FIG. 4. Maximal Lyapunov exponents $\sigma(t)$ for the potential models (the solid lines) of U_{5a} and U_{5b}, and their billiard versions (the dotted lines). The values for the potential models are $\sigma\approx6.5\times10^{-2}$ for U_{5a} and $\sigma\approx14.1$ for U_{5b}, while those for the billiard versions are $\sigma\approx5.00\times10^{-3}$ for U_{5a} and $\sigma\approx13.6$ for U_{5b}. All values of σ are determined at $t=25.1$.

we assign to the origin of time, $t=0$, for the simulation for measuring the total acoustic energy $W(t)$. For the case of the absorption uniformly distributed on the wall with an absorption coefficient α, the energy of the ray is multiplied by a constant factor $(1-\alpha)$ at each reflection. Figure 3 shows the result for the case of U_{5a} for several values α with $N_0=3\times10^3$ and $T_{\text{erg}}=50\bar{l}/c_0$, where $\bar{l}$ is the experimental mean free path, and we evaluate $\bar{l}=12.95$ for this shape. From the semilog plot in Fig. 3, we obtain the numerical values of T_{dec} for U_{5a} as follows: $(\alpha,T_{\text{dec}})=(0.01,3.7999)$, $(0.03,1.2554)$, $(0.1,0.3645)$, and $(0.3,0.1090)$. From the similar simulation for U_{5b} with $\bar{l}=12.32$, we obtain as follows: $(\alpha,T_{\text{dec}})=(0.01,3.6108)$, $(0.03,1.1924)$, $(0.1,0.3458)$, and (0.3, 0.1032).

The measured value T_{dec} should be compared to the theoretical one $\bar{T}_{\text{dec fluc}}$, which is the improved version of the time $\bar{T}_{\text{dec}}$ in the Norris–Eyring law[11] by taking into account the correction due to fluctuations in the number of encounters with the absorbing walls, as follows:[12]

$$\bar{T}_{\text{dec fluc}}=\frac{1}{c}\frac{\sigma_\infty^2}{\sqrt{\bar{l}_0^2-2\sigma_\infty^2\log(1-\alpha)}-\bar{l}_0}, \tag{8}$$

where $\bar{l}_0=\pi S/P$ is the theoretical mean free path with a surface S and a perimeter P of two-dimensional domain, and σ_∞^2 is the asymptotic standard deviation of the fluctuations in length between two successive rebounds, whose value can be numerically determined from the relevant correlation. In the present cases, we obtained $\sigma_\infty^2=24.2$ for U_{5a}, and $\sigma_\infty^2=13.2$ for U_{5b}. As $\bar{l}_0=12.97$ (i.e., $S=258.9$, $P=62.73$) for U_{5a} and $\bar{l}_0=12.33$ (i.e., $S=248.7$, $P=63.35$) for U_{5b}, we obtain from Eq. (8) the values of $\bar{T}_{\text{dec fluc}}$ as follows: $(\alpha,\bar{T}_{\text{dec fluc}})=(0.01,3.7976)$, $(0.03,1.2549)$, $(0.1,0.3647)$, and $(0.3,0.1096)$ for U_{5a}, and $(\alpha,\bar{T}_{\text{dec fluc}})=(0.01,3.6107)$, $(0.03,1.1924)$, $(0.1,0.3458)$, and (0.3, 0.1032) for U_{5b}. From the comparison between the theoretical values $\bar{T}_{\text{dec fluc}}$ and the measured ones T_{dec}, we see that all relative errors between them are within an accuracy of 0.6×10^{-2}. Thus, this agreement with the measured values T_{dec} will indicate that the potential model for the ray motion is applicable to the room acoustics.

Next let us quantitatively clarify the characteristics of the diffusive sound field produced by two different shapes of U_{5a} and U_{5b}. For this purpose, we estimate the maximal Lyapunov exponent (LE) σ because this quantity is a reliable indicator of chaos. By linearizing the equations of motion derived from Eq. (3) and then measuring the longtime evolution of their solutions up to the total iteration $N_{\text{total}}=2.3\times10^8$, i.e., $t=\tau/c_0^2=25.1$, we obtain the behavior of $\sigma(t)$ as shown in Fig. 4 from which the maximal LE are determined as follows: $\sigma\approx6.5\times10^{-2}$ for U_{5a} plotting with a thin solid line and $\sigma\approx14.1$ for U_{5b} plotting with a bold solid line. Since the σ value for U_{5a} will be regarded as zero, the maximal LE implies that the ray motion becomes chaotic for U_{5b} but does not for U_{5a}. At first glance, theseresults seem to be in conflict with the PSS of Figs. 2(c) and 2(d) indicating chaotic motions.

From the viewpoint of billiards in polygons,[13] however, the billiard with a boundary of Fig. 2(a) (we refer to the billiard of U_{5a}, for short) is categorized by the parabolic billiards whose typical systems are the billiards in polygons, while the billiard with a boundary of Fig. 2(b) (the billiard of U_{5b}) is by the hyperbolic ones whose typical systems are the dispersing billiards such as the polygons with smooth convex obstacles. The parabolic billiards are known to be regular, i.e., $\sigma=0$, while the hyperbolic ones are chaotic, $\sigma>0$ due to their strong mixing properties. As shown in Fig. 4, the billiard of U_{5a} plotting with a thin dotted line has $\sigma\approx5.00\times10^{-3}$, which is practically zero, and the billiard of U_{5b} plotting with a bold dotted line has $\sigma\approx13.6$. These results clearly indicate that the billiard of U_{5a} is regular while the billiard of U_{5b} is chaotic. As for the origin of the randomly scattered points of the PSS for U_{5a} in Fig. 2(c), it might be explained by the conjecture that the motion in the parabolic billiards is ergodic and weakly mixing,[14] or by the pseudochaotic systems with random trajectories but, at the same time, with zero LE.[15] Since it is shown that the potential system at high energy limit,i.e., the system with a perfectly square well potential, possesses the common features on chaos as its billiard version for some Hamiltonian systems,[16] the present model will also have the same result as the corresponding billiard model with respect to the LE.

Thus we understand that the diffusive wall like U_{5b} makes the acoustic field chaotic.

All calculations here were done by the fourth-order Runge–Kutta–Fehlberg method with step-size adjustment algorithm to demand high precision.[17] In order to maintain the constraint $H=0$, we have performed the Runge–Kutta integration so as to keep the accuracy of the integration within a desired accuracy, *eps*, per N iterations by monitoring of local truncation error and adjusting step size. In the present calculations we set $eps=10^{-14}$ and $N=10^4$, and calculated the integration until the total iterations N_{total}, e.g., $N_{\text{total}}=2.7\times10^9$ for the PSS in Figs. 2(c) and 2(d), and $N_{\text{total}}=1.2\times10^7$ for $W(t)$ in Fig. 3. In order to check our results, we also calculated the trajectories for different values of *eps*, N_{total}, and m of Eqs. (5) and (6), and for many different values of initial conditions, and then confirmed that the qualitative results remain almost the same.

In conclusion, we have presented the square well potential model describing a ray motion in a finite region. This model is shown to well reproduce the Norris–Eyring law, whose validity has been studied by mainly using the billiard systems.[8,9] In the billiard system, however, the noninteracting rays always propagate along straight lines between the boundary of the domain. On the other hand, our potential model can describe the curved rays in the domain, e.g., Fig. 1, whose curvature generally depends on the sound speed $c(\boldsymbol{x})$ taking account of the inhomogeneity of medium. This feature will be one of the advantages of the present model compared to ray tracing.[6–9] Thus, the square well potential model will convey more information on the room acoustics that goes beyond the conventional billiard with straight trajectories. Furthermore, the extension of Eq. (2) to a higher dimension is straightforward so that it will be possible to treat the room acoustics in a three-dimensional space.

Let us briefly comment on the standard Hamiltonian (2). As pointed out in Ref. 4, under the constraint that $cp=1$, there are many functional forms of Hamiltonian with constraint $H=0$ such as $H=p-1/c$ and $H=\ln(cp)$, which are equivalent to Eq. (2). From the viewpoint of dynamical systems, however, the standard Hamiltonian (2) is superior to other functional forms because the potential picture for the ray motion is naturally obtained.

Finally, we would like to comment on the Hamiltonian used for the underwater acoustics, which is a desirable standard Hamiltonian derived from the parabolic ray equation[18] under several assumptions. Due to the assumption of a one-way ray propagation, however, this standard Hamiltonian is not capable of describing a ray motion in a closed domain. Therefore, we expect that our present model will also serve for the underwater acoustics.

ACKNOWLEDGMENTS

We acknowledge the members of Department of Acoustic Design of Kyushu University for useful discussions.

[1] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959); *The Classical Theory of Fields* (Pergamon, MA, 1987).
[2] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics* (Springer, Berlin, 1992).
[3] L. Casetti, M. Pettini, and E. G. D. Cohen, "Geometric approach to Hamiltonian dynamics and statistical mechanics," Phys. Rep. **337**, 237–341 (2000).
[4] L. M. Brekhovskikh and O. A. Gordin, *Acoustics of Layered Media II: Point Sources and Bounded Beams*, Springer Series on Wave Phenomena (Springer, New York, 2006), Vol. **10**.
[5] T. Kawabe, K. Aono, and M. Shin-ya, "Acoustic ray chaos and billiard system in Hamiltonian formalism," J. Acoust. Soc. Am. **113**, 701–704 (2003).
[6] W. B. Joyce, "Sabine's reverberation time and ergodic auditoriums," J. Acoust. Soc. Am. **58**, 643–655 (1975).
[7] F. Mortessagne, O. Legrand, and D. Sornette, "Role of the absorption distribution and generalization of Sabine's reverberation law in chaotic rooms: Geometrical and wave theory," J. Acoust. Soc. Am. **93**, 2343–2344 (1993).
[8] F. Mortessagne, O. Legrand, and D. Sornette, "Role of the absorption distribution and generalization of exponential reverberation law in chaotic rooms," J. Acoust. Soc. Am. **94**, 154–161 (1993).
[9] O. Legrand and D. Sornette, "Test of Sabine's reverberation time in ergodic auditoriums within geometrical acoustics," J. Acoust. Soc. Am. **88**, 865–870 (1990).
[10] H. Kuttruff, *Room Acoustics*, 3rd ed. (E and FN Spon, London, 1991).
[11] C. F. Eyring, "Reverberation time in "dead" rooms," J. Acoust. Soc. Am. **1**, 217–241 (1930).
[12] C. W. Kosten, "The mean free path in room acoustics," J. Acoust. Soc. Am. **10**, 245–250 (1960).
[13] See, e.g., M. V. Berry, "Regularity and chaos in classical mechanics, illustrated by three deformations of a circular billiard," Eur. J. Phys. **2**, 91–102 (1981); E. Gutkin, "Billiards in polygons," Physica D **19**, 311–333 (1986); S. Tabachnikov, *Geometry and Billiards* (MASS (Mathematics Advanced Study Semesters) at Pennsylvania University, Phildelphia, PA, 2005).
[14] R. Artuso, G. Casati, and I. Guarneri, "Numerical study on ergodic properties of triangular billiards," Phys. Rev. E **55**, 6384–6390 (1997); G. Casati and T. Prosen, "Mixing property of triangular billiards," Phys. Rev. Lett. **83**, 4729–4732 (1999).
[15] G. M. Zaslavsky and M. A. Edelman, "Fractional kinetics: From pseudochaotic dynamics to Maxwell's Demon," Physica D **193**, 128 (2004).
[16] T. Kawabe and S. Ohta, "Chaos in a periodic three-particle system under Yukawa interaction," Phys. Rev. A **41**, 720–725 (1990).
[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran* 2nd ed. (Cambridge University Press, 1992).
[18] K. B. Smith, M. G. Brown, and F. D. Tappert, "Ray chaos in underwater acoustics," J. Acoust. Soc. Am. **91**, 1939–1949 (1992); K. B. Smith, M. G. Brown, and F. D. Tappert, "Acoustic ray chaos induced by mesoscale ocean structure," J. Acoust. Soc. Am. **91**, 1950–1959 (1992).

The Herschel–Quincke tube: The attenuation conditions and their sensitivity to mean flow

Mikael Karlsson[a] and Ragnar Glav[b]
KTH, The Marcus Wallenberg Laboratory for Sound and Vibration Research, Teknikringen 8, 100 44 Stockholm, Sweden

Mats Åbom[c]
KTH, Linné Flow Centre, 100 44 Stockholm, Sweden

(Received 30 April 2007; revised 26 March 2008; accepted 9 May 2008)

The classic Herschel–Quincke tube is a parallel connection of two ducts yielding multiple noise attenuation maxima via destructive interference. This problem has been discussed to different degrees by a number of authors over the years. This study returns to the basics of the system for the purpose of furthering the understanding of the conditions necessary for noise attenuation and especially their sensitivity to mean flow. First, the transmission loss for an *N*-duct system with mean flow and arbitrary conditions of state in the different ducts is derived. Next, the two types of conditions yielding the attenuation maxima are studied. In addition to a discussion of the underlying physics, generic expressions for frequencies at which maximum attenuation occur are presented. Experiments without mean flow generally show good agreement with theory based on straight duct elements. However, more detailed models may be required for accurate simulations in the presence of mean flow. A simple model compensating for the losses associated with bends is shown to improve the results significantly for the geometry studied.
 [DOI: 10.1121/1.2940580]

PACS number(s): 43.28.Py, 43.20.Mv [LLT] Pages: 723–732

I. INTRODUCTION

The Herschel–Quincke (HQ) tube, consisting in its original form of two parallel coupled one-dimensional (1D) waveguides, yields noise cancellation at frequencies corresponding to destructive interference between the acoustic waves in the alternative paths. The most common application published to date is rigid-walled pipes containing a gaseous media. The HQ arrangement is not often used but may be of interest for applications with mean flow since, correctly implemented, the HQ arrangement could provide noise attenuation with a small pressure loss penalty.

In a presentation to the Sections of Physics of the British Association at Cambridge University that was published in 1833, Herschel[1] first presented the idea of destructive wave interference in 1D waveguides. The topic was really the absorption of light and, in a rather heated debate, Herschel only used the destructive interference of sound waves as an illustrative example. Some 30 years later, Quincke[2] developed the idea further and validated the basic case experimentally. The next notable contribution to the subject was in 1928 when Stewart[3] pointed out that attenuation maxima are explained not only by destructive interference of out-of-phase waves but also by another phenomenon. In a later work, Stewart[4] derived a condition for the attenuation maxima that allowed for any area ratio between the two ducts. More recently, Selamet *et al.*[5] and Selamet and Easwaran[6] gave closed form expressions for the transmission loss for the two-duct and the *N*-duct (*N* ducts in parallel) configurations. The influence of mean flow, which is relevant for many practical applications of this arrangement, was investigated by Torregrosa *et al.*,[7] Zhichi *et al.*,[8] and Fuller and Bies.[9]

Various authors have developed this basic principle further. Desantes *et al.*[10] added an extra interconnecting pipe to extend the number of possible interference paths. Hwang *et al.*[11] implemented the concept in exhaust systems using two HQ tubes in series and actively changing the length of the two interference paths in order to adapt the attenuation to follow the harmonics of the engine with variations of engine speed. Trochon,[12] and later McLean,[13] reported on the use of quarter-wave resonators at the nodal points of the HQ tube arrangement, thereby achieving a broader and smoother attenuation curve for turbo and intake noise in internal combustion engines. Burdisso and Smith[14] applied a circumferential array of HQ tubes to turbofan engines. This work is interesting since it aims at higher order modes, whereas other studies are often restricted to the plane wave range. Finally, Griffin *et al.*[15] made a theoretical study of an actively controlled membrane introduced in the longer of the two ducts that allowed the characteristics of the system to adapt to changes in the incoming disturbance.

Although this is a classic problem, some phenomena regarding the occurrence and magnitude of the attenuation maxima, especially with mean flow present, have not been fully investigated. To the authors' knowledge, explicit expressions for the two attenuation conditions have been given for only one specific configuration: identical characteristic impedances in the ducts and no mean flow. Torregrossa

[a]Electronic mail: kmk@kth.se
[b]Electronic mail: glav@kth.se
[c]Electronic mail: matsabom@kth.se

et al.[7] include mean flow effects in an ideal two-duct configuration in which they derive an expression for the transmission loss. They note the system's sensitivity to mean flow but do not explain it.

This work focuses on understanding the attenuation conditions and their sensitivity to mean flow. The HQ arrangement is studied using the mobility matrix, an efficient and compact choice for parallel coupled systems. A general expression for the transmission loss of an N-duct system with mean flow and arbitrary conditions of state in the ducts in a manageable format is proposed. Next, the two types of conditions yielding attenuation maxima are discussed. Based on this understanding, generic expressions for the frequencies where maximum attenuation occurs are derived. Special attention is paid to explaining the influence of mean flow on the derived attenuation conditions. Also, a simple model is given to compensate for the inevitable use of curved ducts. Finally, the system is validated experimentally with and without mean flow using the two-microphone wave decomposition method.

II. THEORY

For a parallel coupled system such as the HQ arrangement, an efficient way of relating the upstream and downstream state variables of a two port, assuming plane wave theory, is the mobility matrix formulation:

$$\begin{bmatrix} \hat{q}_1 \\ \hat{q}_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \end{bmatrix}, \tag{1}$$

where the state variables $\hat{p}$ and $\hat{q}$ are temporal Fourier transforms of acoustic pressure and volume flow, respectively, at nodes one and two (Fig. 1). The total matrix for N two ports coupled in parallel is obtained by adding the mobility matrices:

$$\mathbf{Y}^{\mathrm{HQ}} = \begin{bmatrix} \sum_{n=1}^{N} Y_{11,n} & \sum_{n=1}^{N} Y_{12,n} \\ \sum_{n=1}^{N} Y_{21,n} & \sum_{n=1}^{N} Y_{22,n} \end{bmatrix}. \tag{2}$$

Relating incident to transmitted waves using the coefficients of reflection and transmission, R and T, and assuming a reflection free termination, the following set of equations is obtained:

$$\begin{aligned} (1+R) &= Z_{\mathrm{in}} Y_{11}^{\mathrm{HQ}} (1+R) + Z_{\mathrm{in}} Y_{12}^{\mathrm{HQ}} T, \\ T &= Z_{\mathrm{out}} Y_{21}^{\mathrm{HQ}} (1+R) + Z_{\mathrm{out}} Y_{22}^{\mathrm{HQ}} T, \end{aligned} \tag{3}$$

where Z is the characteristic impedance. Solving these equations yields

$$R = \frac{1 - Z_{\mathrm{in}} Y_{11}^{\mathrm{HQ}} + Z_{\mathrm{out}} Y_{22}^{\mathrm{HQ}} + Z_{\mathrm{in}} Z_{\mathrm{out}} \det(Y^{\mathrm{HQ}})}{1 + Z_{\mathrm{in}} Y_{11}^{\mathrm{HQ}} - Z_{\mathrm{out}} Y_{22}^{\mathrm{HQ}} - Z_{\mathrm{in}} Z_{\mathrm{out}} \det(Y^{\mathrm{HQ}})}, \tag{4a}$$

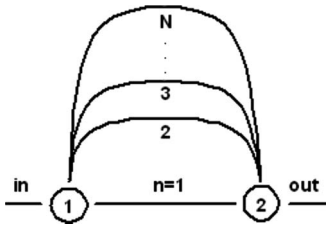


FIG. 1. An N-duct HQ arrangement.

$$T = \frac{2 Z_{\mathrm{out}} Y_{21}^{\mathrm{HQ}}}{1 + Z_{\mathrm{in}} Y_{11}^{\mathrm{HQ}} - Z_{\mathrm{out}} Y_{2}^{\mathrm{HQ}} - Z_{\mathrm{in}} Z_{\mathrm{out}} \det(Y^{\mathrm{HQ}})}, \tag{4b}$$

where $\det(Y) = Y_{11}Y_{22} - Y_{12}Y_{21}$. The transmission loss is now obtained by

$$\mathrm{TL} = 10 \log_{10}\left(\frac{Z_{\mathrm{out}} (1+M_{\mathrm{in}})^2}{|T|^2 Z_{\mathrm{in}} (1+M_{\mathrm{out}})^2} \right), \tag{5}$$

where M is the Mach number.

Using Eqs. (2), (4b), and (5), the transmission loss for any HQ arrangement can be derived if the mobility matrix components for each branch of the N-duct configuration are known. The mobility matrix components for a straight duct element of length L with mean flow are

$$Y_{\mathrm{duct}} = \begin{pmatrix} \dfrac{-i\cot(kL)}{Z} & \dfrac{-e^{ikML}}{iZ\sin(kL)} \\ \dfrac{1}{ie^{-ikML} Z \sin(kL)} & \dfrac{i\cot(kL)}{Z} \end{pmatrix}, \tag{6}$$

$$k = \frac{k_0 (1 - i\zeta)}{1 - M^2}, \tag{7}$$

where k_0 is the wave number and ζ represents losses (e.g., Dokumaci[16]).

For a phenomenological study, especially without mean flow, it is reasonable to use straight duct elements only. However, for more detailed studies, in particular with mean flow, a more complex model may be of interest. Any HQ arrangement includes at least one bend; in a duct bend, flow separation can potentially influence the transmission properties and may also give rise to an internal source. Here, only the passive part—that is, the influence of mean flow on the reflection and transmission properties of the bend—is accounted for. Basically, the bend is a local flow constriction that can be expressed by the contraction ratio.[17] This ratio, however, cannot be derived in a straightforward manner from the geometry; in practice it is more convenient to relate to the pressure loss over the bend.[18,19] Assuming low Helmholtz and Mach numbers, the mobility matrix representing the resistive part—that is, the losses of the bend—are represented by

$$Y_{\mathrm{bend}} = \frac{1}{MZC_L} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \tag{8}$$

where C_L is the pressure loss coefficient, which for most geometries can readily be found in handbooks.[20,21] Otherwise, the pressure loss coefficient can be measured or computed using, for example, any standard computational fluid

dynamics (CFD) code. The proposed model does not take the length of the duct into account. Instead, the reactive part is modeled using straight duct elements corresponding to the length of the centerline of the bend before and after the coordinate where the lumped resistive part is introduced. The resulting two port is then fed into the global mobility matrix for the HQ tube in Eq. (2). It can be discussed which position in the bend is most representative for introducing the local flow restriction. As a first approximation, it is suggested that the restriction be located at the end of the bend, where, for a typical bend used in engineering problems, the vena-contracta effect would be most accentuated.

III. CONDITIONS FOR ATTENUATION MAXIMA

The transmission loss for any arbitrary N-duct configuration can be obtained from the above-derived Eqs. (2) and (4b)–(6). However, for brevity, this section limits the study to the classic two-duct configuration. The extension to an N-duct case, however interesting from an academic or an engineering point of view, does not take the analysis any further.

In the test case, the two pipes have the same constant cross-sectional area, but one pipe is twice the length of the other. The inlet and outlet duct cross-sectional areas equal the sum of the areas of the two HQ ducts. When a Mach number is given without index, it is assumed equal in the two ducts. In this section an ideal case is considered—that is, straight duct elements are used and losses are neglected.

In studying transmission loss, there is an assumption of a reflection-free termination. Thus, for complete reflection of incoming sound, no wave can be transmitted downstream of node two. This implies that both the acoustic pressure and the volume velocity at node two are zero at complete reflection. To distinguish between the two different types of attenuation condition, node one is studied. In a reactive-type silencer, such as the HQ tube, attenuation is obtained at frequencies where the acoustic pressure and velocity are out of phase. Within the system there will be positions where either the acoustic pressure or the volume velocity is zero. Fixed observation points, at nodes one and two in this application, can, but need not, coincide with the positions where a state variable is zero. However, from the above-presented argument, it is known that the state variables are zero at node two. If node one also coincides with a nodal point of the acoustic pressure, the original wave interference condition predicted by Herschel is obtained. For the second type of attenuation peak, originally observed by Stewart,[3] with one exception the state variables do not have a nodal point at node one. For the case of identical impedances in the N ducts and no mean flow, the acoustic volume velocity is zero at node one. Described previously by Stewart,[3] this case is commonly derived in later publications as a straightforward analytical solution.[5,7] As shown later in this paper, for this special case the two different attenuation conditions actually coincide and can be described by a single expression. However, for a general description that also covers the case discussed by Stewart,[4] where the impedance ratio of the two ducts differs from unity, it is necessary to refer to the basic

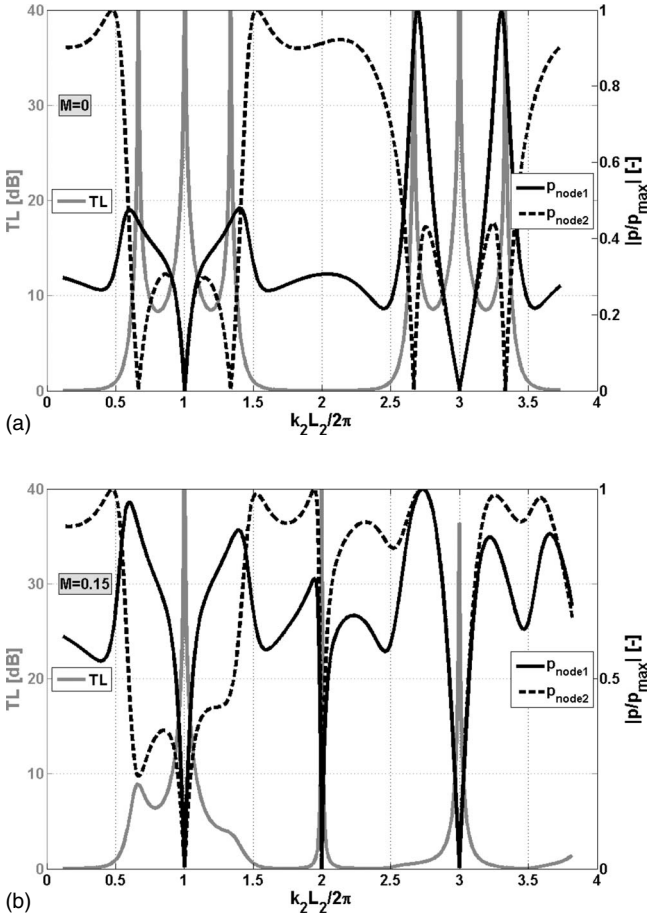


FIG. 2. Acoustic pressure amplitude in the two nodes. (a) $M=0$ and (b) $M=0.15$.

condition: At node one, both the acoustic pressure and the volume velocity are nonzero but are out of phase at odd multiple of $\pi/2$.

To derive expressions for the two conditions yielding attenuation maxima, it is sufficient to study the acoustic pressure. In Fig. 2, the acoustic pressure in the two nodes is shown with and without mean flow. The two cases can now be distinguished, and are denoted as: *Type I*—The acoustic pressure is zero at the downstream node but nonzero at the upstream node and *Type II*—The acoustic pressure is zero at both nodes.

In Fig. 2, it can be seen that Type I is especially sensitive to mean flow. A new attenuation maximum of Type II appears with mean flow at a frequency corresponding to a whole wavelength difference between the two ducts.

A. Type I attenuation condition

Using Eqs. (1), (2), and (6), Type I condition yields

$$Y_{21}^{\mathrm{HQ}} = 0 \rightarrow \frac{1}{Z_1 e^{-ik_1 M_1 L_1} \sin(k_1 L_1)} + \frac{1}{Z_2 e^{-ik_2 M_2 L_2} \sin(k_2 L_2)} = 0, \tag{9}$$

which is dependent upon the characteristic impedances in the two ducts. Consequently, in changing the ratios Z_2/Z_1 or A_2/A_1, if the same condition of state is assumed in the two

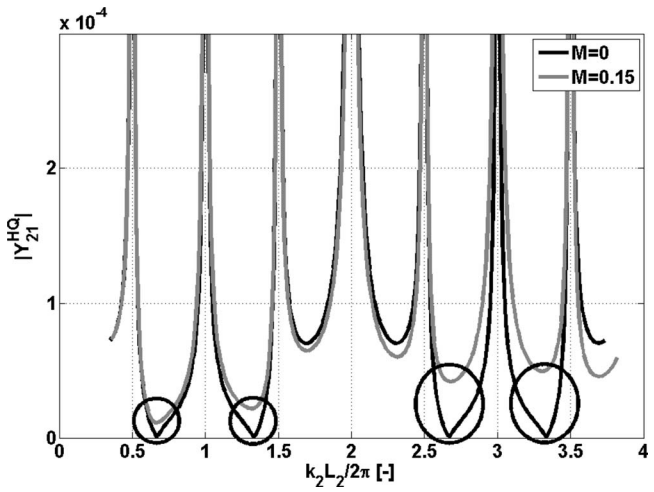


FIG. 3. The Type I attenuation condition at $M=0$ and $M=0.15$. Equal area ducts.

ducts, the frequencies at which the attenuation maxima occur shift.

Assuming identical characteristic impedances and no mean flow, the condition given by Eq. (9) equals the "resonance condition" as described in previous publications.[5,7] This condition holds for that special case since the acoustic volume velocity then is zero at both nodes. Both attenuation conditions are then derived from the Type I expression alone and are reduced to functions only of the duct lengths:

$$\text{Type I condition } (M=0, Z_1=Z_2): \quad k(L_1+L_2) = 2m\pi, \quad m=1,2,\ldots, \tag{10a}$$

$$\text{Type II condition } (M=0, Z_1=Z_2): \quad k(L_2-L_1) = (2n-1)\pi, \quad n=1,2,\ldots. \tag{10b}$$

However, for deviations from this special case, the Type II condition is not described by the Type I expression but must be derived separately.

In Fig. 3, the absolute value of the Type I condition is given for Mach=0 and Mach=0.15. The influence of mean flow is strong and increases with frequency. As shown in Fig. 2, this influence is reflected in the transmission loss. As illustrated later in this paper, the attenuation given by the Type I condition deteriorates considerably, even at moderate Mach numbers.

B. Type II attenuation condition

Using the mobility matrix formulation derived in Sec. II, the Type II attenuation condition is expressed as

$$\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \end{bmatrix} = [Y^{\mathrm{HQ}}]^{-1} \begin{bmatrix} \hat{q}_1 \\ \hat{q}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{11}$$

As noted in deriving the Type I condition, the acoustic volume velocity is zero at both nodes for the special case of equal impedances in the ducts and no mean flow, thus fulfilling the Type II condition. However, for the general case, the mobility matrix must be studied. The attenuation condition is

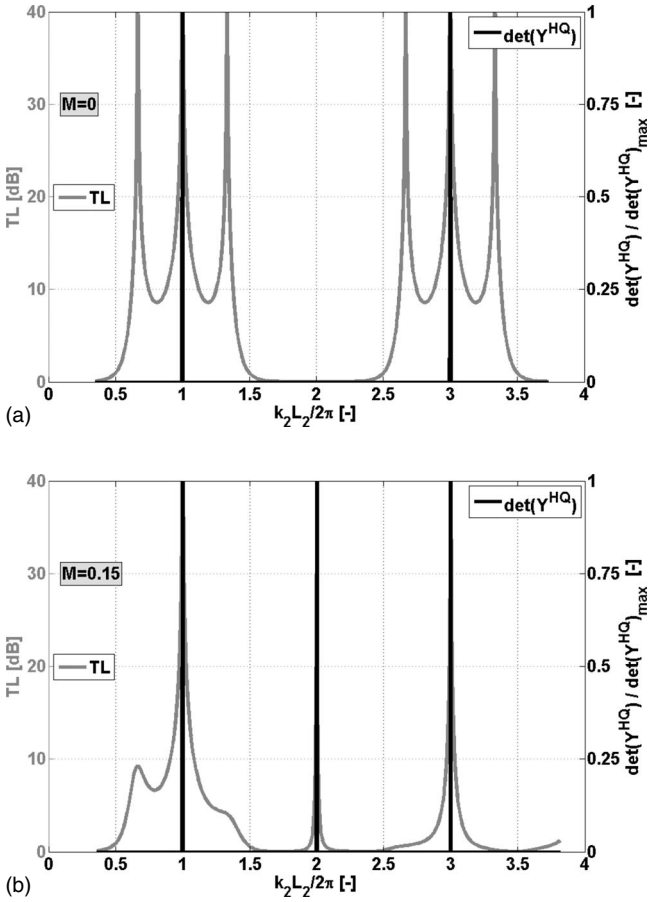


FIG. 4. The Type II attenuation condition at (a) $M=0$ and (b) $M=0.15$. Equal area ducts.

now given as: $\det(\mathbf{Y}^{\mathrm{HQ}}) \to \infty$ (see Fig. 4). Returning to the expression for the transmission, T, Eq. (4b), in Type II blocking frequencies both the numerator and denominator include terms of the type $1/\sin(kL) \to \infty$. But the denominator includes the product of two such terms, and therefore the ratio of the numerator and denominator goes to zero, yielding zero transmission.

Applying the Type II condition, using Eqs. (2) and (6),

$$\frac{\cos^2(k_1L_1)}{Z_1^2\sin^2(k_1L_1)} + \frac{2\cos(k_1L_1)\cos(k_2L_2)}{Z_1Z_2\sin(k_1L_1)\sin(k_2L_2)} + \frac{\cos^2(k_2L_2)}{Z_2^2\sin^2(k_2L_2)} - \frac{1}{Z_1^2\sin^2(k_1L_1)} - \frac{e^{-ik_1L_1M_1}}{e^{-ik_2L_2M_2}Z_1Z_2\sin(k_1L_1)\sin(k_2L_2)} - \frac{e^{-ik_2L_2M_2}}{e^{-ik_1L_1M_1}Z_1Z_2\sin(k_1L_1)\sin(k_2L_2)} - \frac{1}{Z_2^2\sinh^2(k_2L_2)} \to \infty \tag{12}$$

and the following expressions are derived:

$$k_1L_1 = m\pi, \quad m=1,2,3,\ldots, \tag{13a}$$

$$k_2L_2 = n\pi, \quad n=1,2,3,\ldots, \tag{13b}$$

$$\cos(k_1L_1)\cos(k_2L_2) - \cos(k_2L_2M_2 - k_1L_1M_1) \neq 0. \tag{13c}$$

The derived conditions are dependent on the lengths of the ducts and not on their characteristic impedances. For the special case of no mean flow and identical wave numbers in the two ducts—that is, the classic case studied by most authors—the derived condition stipulates the odd integral multiple of a half wavelength difference predicted by Herschel.[1] The Type II attenuation condition is then reduced to Eq. (10b). As noted by Torregrossa *et al.*,[7] Fig. 4(b) shows that with mean flow attenuation maxima also start to occur at even multiples of half a wavelength difference. This phenomenon is explained by Eq. (13c). Without mean flow, this condition includes only odd multiples of half a wavelength difference, while with mean flow, other combinations are possible.

In Eq. (13c), frequency and Mach number combinations exist where the attenuation maxima corresponding to any multiple (odd or even)of half a wavelength difference in duct length cancel. For our example, where one duct is twice as long as the other and where identical wave and Mach numbers in the two ducts are assumed, the canceling Mach numbers are given as

$$M_{\text{cancel}} = \begin{cases} \dfrac{2l-1}{m}, & l = 1,2,3\ldots, \quad m \text{ odd} \\ \dfrac{2l}{m}, & l = 0,1,2\ldots, \quad m \text{ even}, \end{cases} \tag{14}$$

where m is the counter defined in Eq. (13a), which here is the number of half wavelength differences between the ducts. In Fig. 5(a), the magnitude of Eq. (13c) at frequencies corresponding to the first four potential Type II reflection frequencies is given as a function of Mach number. For this example, m is interchangeable with $k_2L_2/2\pi$. As predicted, the odd instances cancel for Mach numbers that are odd multiples of one over their order number, while the even multiples cancel without mean flow and then for Mach numbers corresponding to multiples of two over their order number. Figure 5(b) shows the transmission loss at the corresponding blocking Mach numbers. The first multiple of half a wavelength difference does not cancel for the Mach number range chosen (its first canceling Mach number is unity). The second multiple only cancels at $M=0$ since its second canceling Mach number is also unity, and it is out of the chosen range. The third multiple cancels as predicted at $M=1/3$. Finally, $m=4$ cancels first at $M=0$ and then at $M=1/2$. The large Mach number range displayed is for illustrative purposes only. It should be noted that the mean flow and acoustic fields are assumed decoupled, thus limiting the valid range to approximately $M=0.3$. Furthermore, in any practical application, considerable flow noise levels are expected at these high Mach numbers.

Figure 6 illustrates the difference between the two attenuation conditions derived. The blocking frequencies in the Type I condition shift significantly with the characteristic impedance ratio, and the attenuation performance is more sensitive to mean flow effects than it is for the Type II condition. In addition to the sensitivity to mean flow, with the same Mach number in the two ducts, other problems arise when the Mach number differs (see Fig. 7). The Type I condition is more sensitive and the performance deteriorates significantly even at low Mach number differences ($\Delta M \sim 0.025$) between the pipes. The attenuation in the Type II condition also reduces at Mach number differences ($\Delta M \sim 0.1$) occurring in typical engineering problems. In practical applications with mean flow, it is likely to be the rule rather than the exception that the mass flow distribution between

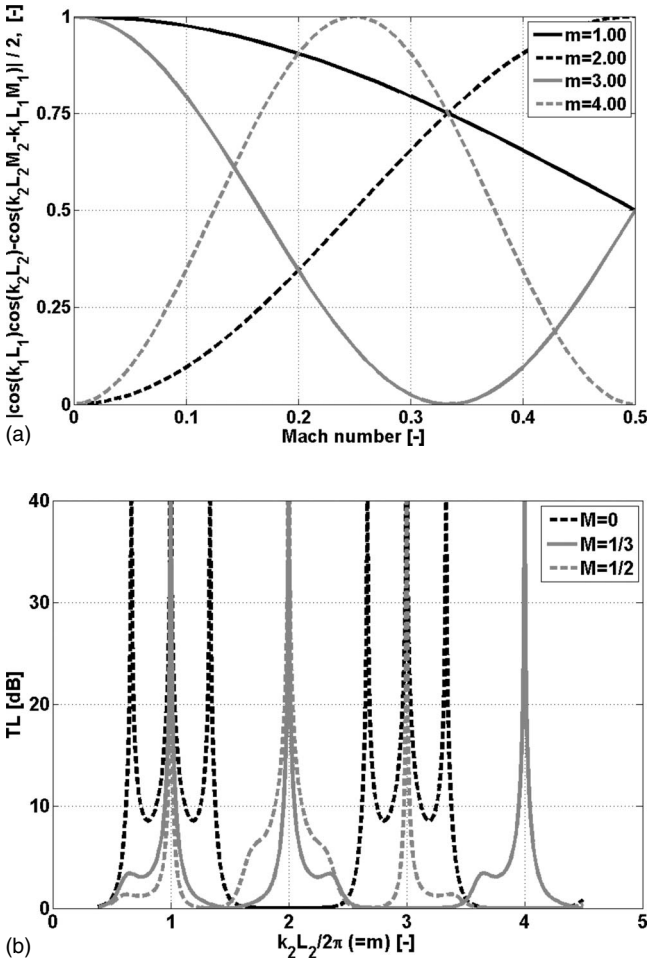


FIG. 5. Study of the Mach number dependence on the Type II condition assuming identical Mach numbers in the two ducts. (a) Equation (13c) for the first four possible instances of the Type II condition. (b) TL at corresponding Mach numbers.

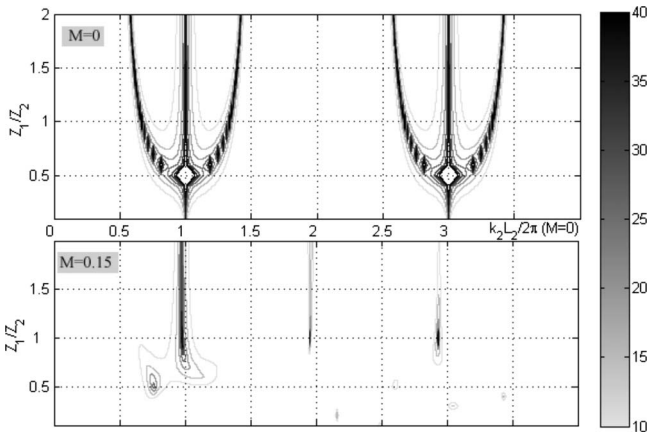


FIG. 6. Transmission loss as a function of dimensionless frequency and characteristic impedance ratio between the pipes.

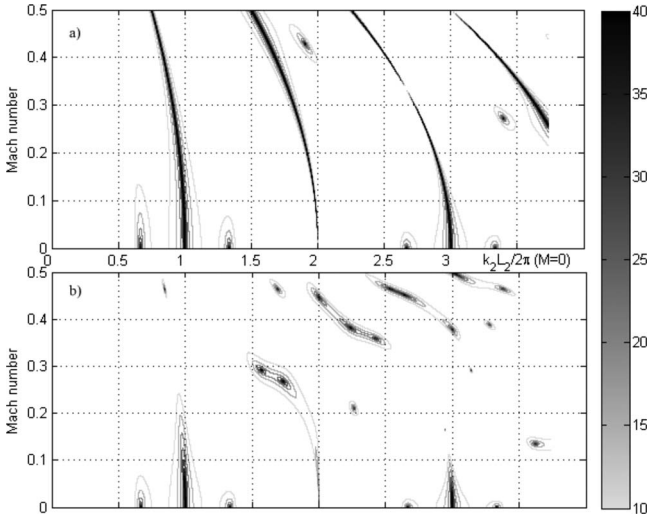


FIG. 7. Transmission loss as a function of dimensionless frequency and Mach number. Equal area ducts. (a) $M_1=M_2$ and (b) $M_1=0$ M_2 varies.

the ducts varies. This distribution is determined by, for example, the geometry of the junctions and bends or the friction factors of the ducts. For a stationary case this might be compensated for by the lengths of the ducts; however, for applications where the inlet mass flow rate varies, the HQ arrangement is likely to function as intended only in a limited range. In the work by Hwang *et al.*,[11] the "Quincke" tube was attached to the main straight duct with a T junction. The pressure loss coefficient of the Quincke tube in such a configuration clearly exceeds the one for the straight duct it is connected to. Hence, the Mach number in the two ducts must have differed significantly; the difference is reflected in the measurements where only attenuation corresponding to the Type II condition is seen.

This argument is not limited to the Mach number but is applicable to all parameters influencing the wave numbers in the ducts. It is possible to consider cases where the temperature differs between the ducts. Controlling such phenomena is one way of reducing the actual duct length needed, but such control is difficult to implement.

IV. EXPERIMENTS

A. Test object

For validation, the case of equal area ducts was chosen. The system consists of a number of rigid-walled steel pipes that are fastened together. Separation of parts allows measurement of the complete system and of the individual elements. The nodes are realized by y junctions that are coupled to two different length ducts. The shorter pipe includes only one bend section, while in the longer pipe straight duct segments double the total length. (See Fig. 8 for a photograph and the geometrical data of the test object.) Because of the large length-to-diameter ratio of the pipes, the effect of the acoustic end corrections is small. In conducting measurements with mean flow, the geometry of the system—especially at the junctions—must be considered carefully. As previously discussed (see Fig. 7), the attenuation conditions are sensitive to relative Mach number differences between the ducts. As the main interest of this study is mean flow

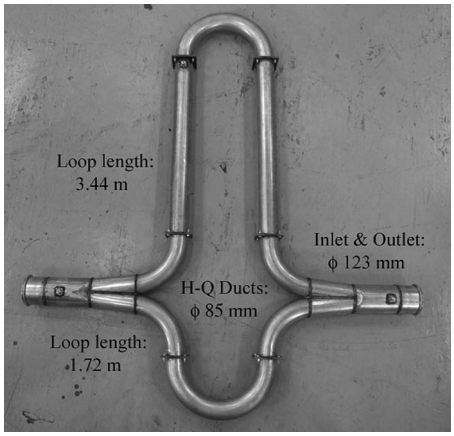


FIG. 8. Test object. Lengths are at the centerlines of the ducts.

effects, the design objective was to minimize the Mach number difference between the ducts, allowing both types of attenuation conditions to be seen at considerable mean flow velocities. The type of test object used by most authors,[5,8,11] a straight duct with another tube attached as a sidebranch at the nodes (T junction), is commonly considered a practical implementation of the system. However, for the purpose of this study the T-junction test object is not ideal because it has significantly different loss coefficients for the alternative paths, resulting in an uneven mean flow distribution. In addition, the T-junction setup may experience problems associated with grazing flow over the orifice (e.g., Cummings[22]). The y junction chosen for the test object in this study aims at reducing such effects. However, other problems may occur, for example, when the flow separates or recombines at the nodes. For reference, the object studied by Fuller and Bies,[9] a duct bend partitioned into two parallel paths, shows similarities to the design in this test. To reduce the relative influence of any pressure losses produced by flow effects in the bends, both pipes in the test object have the same curvature. The bends' radii are twice the diameter of the pipes, which, as a rule of thumb, should be sufficient to keep the pressure loss coefficient, C_L, reasonably low while still allowing for a manageable test object. However, for the shorter duct, where the bends are in line, the resulting bend-in-bend interaction inevitably produces higher losses. For this test case, with $M=0.15$ at the inlet, the relative mass flow distribution in the two ducts was 52% vs 48% in favor of the shorter pipe. The resulting Mach numbers were 0.157 and 0.143, respectively. The higher friction losses of the longer pipe are to some extent compensated for by the higher losses due to the bend-in-bend interaction in the shorter pipe.

B. Experimental procedure

The acoustic properties of the system were determined using the two-microphone wave decomposition method[23] with external sources.[24,25] Although predominantly used for two ports, as shown in Fig. 9, the technique is general and can be applied to any N port. In this instance, two- and three-port measurements were made. The selected state variables of an N port are related by an $N\times N$ matrix. Each independent measurement (realized by external sources)

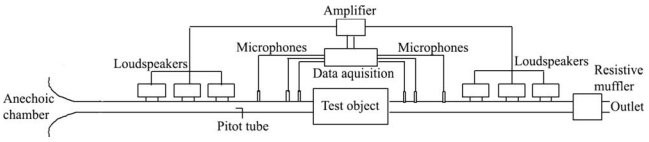


FIG. 9. Two-port configuration of the test rig.

yields N equations, and therefore N independent measurements are needed to solve for the full matrix. In practice, the implication is that at each port a microphone pair and an external source (loudspeaker) are required. To suppress the influence of flow noise, the transfer functions between the loudspeaker signal and the microphone signals are used rather than only the microphone signals (Åbom[26]). For the measurements without mean flow, random excitation was used, while swept sine excitation was used for the measurements with mean flow. The choice of excitation is not critical if sufficient averaging is done. However, this combination seems the most effective alternative for this setup. An array of three loudspeakers was installed at each port along the pipe to ensure that the excitation level was sufficient and that all frequencies were excited. The assumption of plane wave propagation restricted the frequency limit to about 1.6 kHz, which is more than sufficient for making comparisons with the previously simulated data. The actual frequency range was decided by the microphone separation distances. To cover the frequency range of interest, two separation distances were required, using three microphones at each port. The mean flow was generated by a fan and was fed into an anechoic stagnation chamber before entering the test setup, thus providing a stable and silent mean flow. The mean flow velocities were measured using standard Pitot tubes. In the inlet, a fully developed turbulent velocity profile was assumed, while in the HQ ducts a number of readings were taken to obtain the average velocity.

C. Results

In Fig. 10, the measured transmission loss is compared with simulations using a loss model[16] for the straight duct elements. For the zero mean flow case, the pipes are modeled using straight duct elements only, while results are given both with and without the suggested bend model for the case with mean flow. To illustrate the usefulness of such a simple model, the results shown are presented without any detailed knowledge of the system. The model is applied once for each of the three bends in the two ducts. In the longer duct the pressure loss coefficients used were C_L=0.2, 0.35, and 0.2, and in the shorter duct they were C_L=0.3, 0.5, and 0.3. The coefficients were estimated from handbook data.[20,21] Although not shown here, a more complex study simulating the mean flow field using a standard CFD code yielded a better fit to the measured data but not to an extent that made the increased computational effort worthwhile.

In the zero mean flow case, the simplification procedure of using straight duct elements is sufficient to locate all attenuation maxima in frequency and to capture the magnitude of the Type II condition. However, the attenuation maxima associated with the Type I condition are more damped than

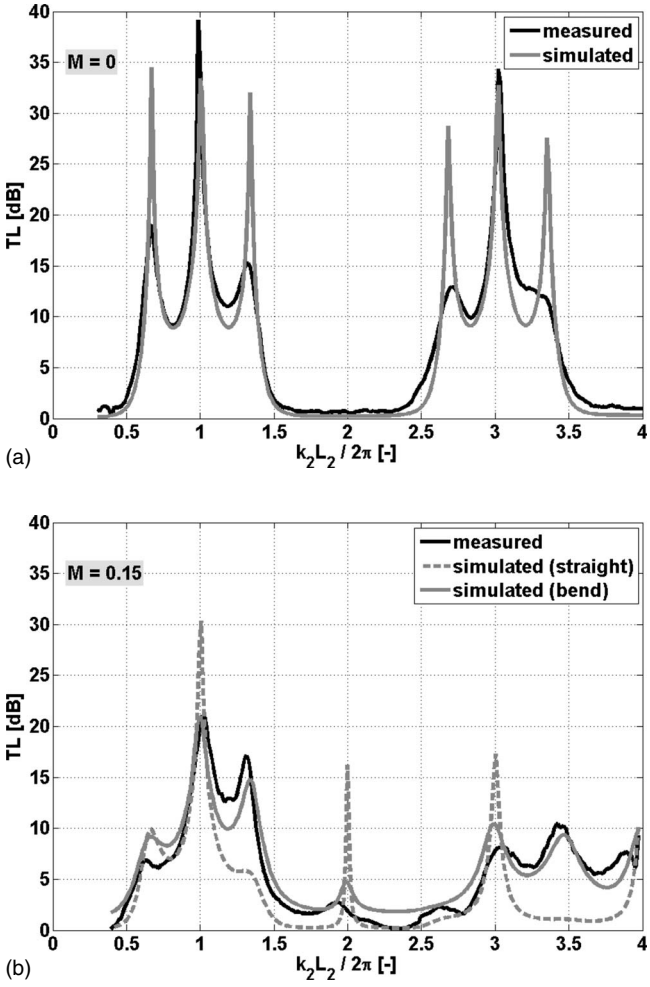


FIG. 10. Simulated and measured transmission loss at (a) M=0 and (b) M=0.15.

the model predicts. Authors conducting earlier experiments have not reported this result.[5,8,9] The observed phenomenon is interesting, especially since the damping of the Type II attenuation peaks is correctly estimated. Bends can affect the wave propagation in ducts even without mean flow. Keefe and Benade[27] have shown that the sheer losses in the bulk of the fluid are sometimes not negligible compared to the wall losses. This finding, however, does not explain the influence on the Type I condition alone.

To localize the cause of the observed phenomenon, each individual component of the system was measured by the same experimental procedure used for the complete system. The straight ducts (two port), the bends (two port) and the y junctions (three port) were measured and compared to modeled data. While the straight ducts and the bends showed only minor deviations from the models, the y junctions differed more significantly. Figure 11 illustrates this result, with the measured reflection coefficient of the node, seen from the inlet duct side, compared to modeled data. Since the area ratio going from the inlet duct to the two HQ ducts is close to unity, the reflection coefficient should be small in the plane wave range. Nevertheless, there are additional losses in the nodes resulting from imperfections in the geometry. The complete HQ tube was then reassembled analytically by connecting the measured nodes (y junctions) with ideal straight

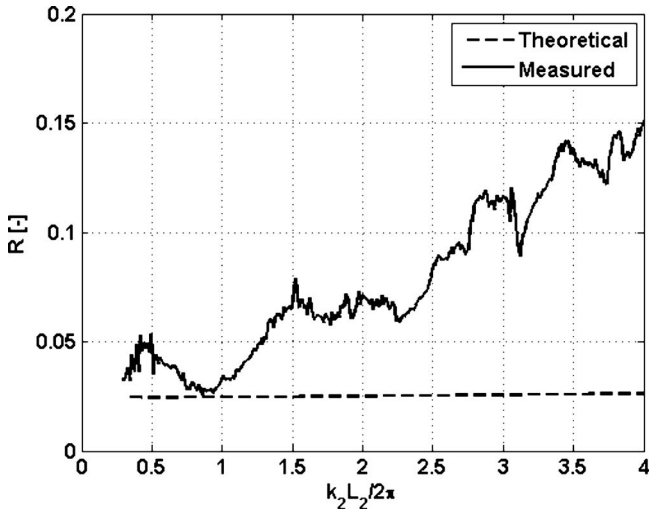


FIG. 11. Theoretical and measured reflection coefficient, R, of the y junction seen by the inlet duct side.

duct elements representing the alternative paths. The resulting transmission loss closely resembles the measured result for the whole system given in Fig. 10(a), thus indicating the imperfections of the nodes as the origin of the unexpected damping of the Type I attenuation maxima. The Type II condition stipulates the behavior of each branch individually as well as their interaction. In contrast the Type I condition specifies only the interaction, making the Type I condition more sensitive to relative changes between the branches.

With mean flow, the use of straight duct elements alone is less useful for this geometry; it captures the frequency but not the magnitude of the attenuation maxima. Again, the representation of the Type I condition is especially poor. Introducing the simple bend model improves the accuracy of the simulation significantly. The effect caused by losses in the nodes, as observed for the zero mean flow case, is small compared to the mean flow effects and cannot be distinguished at the high mean flow velocities used here. As discussed, with mean flow, Type II attenuation maxima begin to occur at frequencies where the difference in duct lengths corresponds to multiples of the wavelength (e.g., $k_2L_2/2\pi=2$).

V. CONCLUSION

The main contribution of this work, as summarized in Table I, is the derivation and understanding of the two different types of attenuation conditions, especially their sensitivity to mean flow. As both conditions require the acoustic pressure to be zero in the downstream node, the distinction between the two conditions is seen in the upstream node. The acoustic pressure is nonzero for the Type I condition and zero for the Type II condition. Once the underlying physics of the problem are understood, the mathematics can be handled by the compact expressions for the performance parameters obtained via the mobility matrix formulation.

For the special case of zero mean flow and identical characteristic impedances in the ducts, all solutions are given by the Type I condition alone. For the general case, the derived distinction between the Type I and Type II attenuation conditions is required. With mean flow, the original Type II wave interference condition yields sound attenuation at fre-

TABLE I. Summary of the main characteristics for the attenuation conditions.

	Type I	Type II
Condition given by	$\hat{p}_{\text{node},1}\neq 0,\ \hat{p}_{\text{node},2}=0$	$\hat{p}_{\text{node},1}=\hat{p}_{\text{node},2}=0$
Resulting equation(s)	$\frac{1}{Z_1 e^{-ik_1M_1L_1}\sin(k_1L_1)}+\frac{1}{Z_2 e^{-ik_2M_2L_2}\sin(k_2L_2)}=0$	$k_1L_1=m\pi,\quad m=1,2,3,\ldots$ $k_2L_2=n\pi,\quad n=1,2,3,\ldots$ $\cos(k_1L_1)\cos(k_2L_2)-\cos(k_2L_2M_2-k_1L_1M_1)\neq 0$
Impedance ratio dependence	Attenuation maxima shift in frequency with varying impedance ratio Z_1/Z_2.	Not dependent upon the impedance ratio between the ducts, only the length ratio.
General mean flow dependence	Strongly influenced by mean flow, especially if the mach number varies in between the ducts.	$M=0$ Only odd integral multiples of half a wavelength difference yields attenuation. $M\neq 0$ Even as well as odd integral multiples yields attenuation. There are always Mach number combinations for which a given multiple cancels. *Sensitivity* Negligible influence if $M_1=M_2$, moderate influence if $M1\neq M2$.
Equal characteristic impedances in the ducts and no mean flow	$\hat{q}_{\text{node},1}=\hat{q}_{\text{node},2}=0$ Solution reduced to a function of the lengths only.	Solution given by the Type I expression.
Modeling elements needed	With mean flow the use of straight duct elements only is not sufficient. The bend model improves the accuracy significantly.	Straight duct elements are sufficient for most practical applications. However, the simulation is improved using the bend model.

quencies where the difference in length between the two ducts corresponds to any multiple of half a wavelength, not just for the odd multiples in the special case of zero mean flow in the concept first presented by Herschel. For all multiples, there are Mach numbers where the attenuation condition cancels. As previously noted, for the even multiples the first canceling Mach number is zero. Generally the higher the multiple, the lower the Mach number where the attenuation condition cancels.

The Type II condition stipulates the behavior of each branch individually as well as the interaction between them, while the Type I condition specifies only the interaction. This difference makes the Type I condition more sensitive to disturbances. For example, even at a small Mach number difference between the two pipes, the magnitude of the Type I attenuation maxima is reduced considerably. In addition, in the experimental results these attenuation maxima were more damped than predicted, even without mean flow, due to imperfections in the nodes of the test geometry.

For a phenomenological study, especially one without mean flow, the use of straight duct elements only has proved sufficient. The blocking frequencies are identified and the attenuation conditions are well captured. However, depending on the geometry, more detailed models may be needed with mean flow present. The test geometry used in this work was chosen to minimize the difference in Mach number between the ducts, thus allowing the Type I condition to be seen with mean flow present. With this setup, the losses associated with the flow separation in the bends of the ducts become important. Applying a simple model for the bend losses improves the accuracy significantly. In other installations, for example with T junctions, the losses associated with the junction itself may have a dominant effect over the influence of the flow-related losses in bends. Then a straight duct model with the appropriate flow distribution might be sufficient.

This study focuses on the basic understanding of the system. However, various alternatives to the work could be studied. For example, varying the areas of the inlet and outlet relative to the HQ ducts changes the attenuation bandwidth but not the blocking frequencies. If the area of the inlet and outlet sections is reduced compared to one HQ duct, the attenuation given by the Type I condition is enhanced while the attenuation given by the Type II condition deteriorates, and vice versa. Here, only uniform geometry ducts have been studied, but each parallel path could certainly have more or less arbitrary geometry. This is easily handled by the general expression for the mobility matrix derived. Another potential way of altering the performance of the HQ tube is to allow for varying conditions of state in the different ducts.

While sensitive to mean flow, the arrangement actually yields sound attenuation over a wide frequency range at Mach numbers relevant for practical applications. Compared to a sidebranch resonator, such as the quarter-wave resonator or the Helmholtz resonator, the attenuation bandwidth is substantial and the sensitivity to mean flow is comparable. Just as with sidebranch resonators, the HQ tube can be implemented with a small pressure loss penalty. For applications where noise attenuation is required at very low frequencies and possibly also at high temperatures, the duct lengths become unrealistically long. However, for applications with mean flow and stationary source conditions in the medium frequency range—say, a few hundred hertz upwards—the HQ tube may prove an interesting option.

ACKNOWLEDGMENTS

The financial support of the Swedish Emission Research Program, Contact No. 310 10 1726, is acknowledged. Prototype material and laboratory facilities were provided by Swenox AB.

[1] J. F. W. Herschel, "On the absorption of light by coloured media, viewed in connection with the undulatory theory," Philos. Mag. **3**, 401–412 (1833).
[2] G. Quincke, "Über Interferenzapparate für Schallwellen," (On an interference device for sound waves), Ann. Phys. Chem. **128**, 177–192 (1866).
[3] G. W. Stewart, "The theory of the Herschel-Quincke tube," Phys. Rev. **31**, 696–698 (1928).
[4] G. W. Stewart, "The theory of the Herschel-Quincke tube," J. Acoust. Soc. Am. **17**, 107–108 (1945).
[5] A. Selamet, N. S. Dickey, and J. M. Novak, "The Herschel-Quincke tube: A theoretical, computational and experimental investigation," J. Acoust. Soc. Am. **96**, 3177–3185 (1994).
[6] A. Selamet and V. Easwaran, "Modified Herschel-Quincke tube: Attenuation and resonance for *n*-duct configuration," J. Acoust. Soc. Am. **102**, 164–169 (1997).
[7] A. J. Torregrosa, A. Broatch, and R. Payri, "A study of the influence of mean flow on the acoustic performance of Herschel-Quincke tubes," J. Acoust. Soc. Am. **107**, 1874–1879 (2000).
[8] Z. Zhichi, L. Song, T. Rui, G. Rui, D. Genhua, and L. Peizi, "Application of Quincke tubes to flow ducts as a sound attenuation device," Noise Control Eng. J. **46**, 245–255 (1998).
[9] C. R. Fuller and D. A. Bies, "The effects of flow on the performance of a reactive acoustic attenuator," J. Sound Vib. **62**, 73–92 (1979).
[10] J. M. Desantes, A. J. Torregrosa, H. Climent, and D. Moya, "Acoustic performance of Herschel-Quincke tube modified with an interconnecting pipe," J. Sound Vib. **284**, 283–298 (2005).
[11] Y. Hwang, J. M. Lee, and S.-J. Kim, "New active muffler system utilizing destructive interference by difference of transmission paths," J. Sound Vib. **262**, 175–186 (2003).
[12] E. P. Trochon, "A new type of silencer for turbocharger noise control," SAE Conference on Noise and Vibration Control, Traverse City, MI, April 2001, 2001-01-1436.
[13] I. McLean, "Optimized Herschel-Quincke acoustic filter," SAE Conference on Noise and Vibration, Traverse City, MI, May 2005, 2005-01-2360.
[14] R. A. Burdisso and J. P. Smith, "Control of inlet noise from turbofan engines using Herschel-Quincke waveguides," 21st AIAA Aeroacoustics Conference, Lahaina, HI, 12-14 June 2000, AIAA-2000-1994.
[15] S. Griffin, S. Huybrechts, and S. A. Lane, "An adaptive Herschel-Quincke tube," J. Intell. Mater. Syst. Struct. **10**, 956–961 (1999).
[16] E. Dokumaci, "A note on transmission of sound in a wide pipe with mean flow and viscothermal attenuation," J. Sound Vib. **208**, 653–655 (1997).
[17] G. C. J. Hofmans, "Vortex sound in confined flows," Doctoral thesis, Technical University of Eindhoven, Eindhoven, The Netherlands, 1998.
[18] S. Nygård, "Low frequency sound in duct systems," Technical Report, KTH MWL, Stockholm, Sweden, 2001.
[19] H. Gijrath, S. Nygård, and M. Åbom, "Modelling of flow generated sound in ducts," in Proceedings of the Eighth International Congress on Sound and Vibration (ICSV8), Hong Kong, July 2001.
[20] D. S. Miller, *Internal Flow Systems*, 2nd ed. (BHR Group, Cranfield, UK, 1990).
[21] R. D. Blevins, *Applied Fluid Dynamics Handbook* (Van Nostrand Reinhold, New York, 1984).
[22] A. Cummings, "The effects of grazing turbulent pipe-flow on the impedance of an Orifice," Acustica **61**, 233–242 (1986).
[23] A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using two-microphone random excitation technique," J. Acoust. Soc. Am. **61**, 1362–1370 (1977).
[24] M. L. Munjal and A. G. Doige, "Theory of a two source-location method

for direct experimental evaluation of the four-pole parameters of an aeroacoustic element," J. Sound Vib. **141**, 323–333 (1990).
[25]M. Åbom, "A note on the experimental determination of acoustical two-port matrices," J. Sound Vib. **155**, 185–188 (1992).
[26]M. Åbom, "Measurement of the scattering-matrix of acoustical two-ports," Mech. Syst. Signal Process. **5**, 89–104 (1991).
[27]D. H. Keefe and A. H. Benade, "Wave propagation in strongly curved ducts," J. Acoust. Soc. Am. **74**, 320–332 (1983).

Sound propagation in the vicinity of an isolated building: An experimental investigation

W. C. Kirkpatrick Alberts II,[a] John M. Noble, and Mark A. Coleman
U.S. Army Research Laboratory, Attn: AMSRD-ARL-CI-ES, 2800 Powder Mill Road, Adelphi, Maryland 20783

(Received 18 September 2007; revised 19 February 2008; accepted 17 May 2008)

Recently, the study of acoustics in urban terrain has been concerned with the propagation of sound through street canyons typical of residential areas in large cities, while sparsely built suburban and rural areas have received little attention. An isolated building's effect on propagating sound is a fundamental case of suburban acoustics and urban acoustics in general. Its study is a necessity in order to determine the processes that might be required to model the sound field in the building's vicinity, e.g., diffraction and wind effects. The work herein presents the results of an experimental effort to characterize the interaction between propagating sound and a single story, gabled-roof building typical of some North American suburban and rural areas. Recorded data are found to reasonably compare to a common diffraction model in some instances.
 [DOI: 10.1121/1.2945151]

PACS number(s): 43.28.En, 43.20.El [KA] Pages: 733–742

I. INTRODUCTION

From the point of view of someone living in a city or a suburban area, the background noise level, due to a range of mechanisms from multilane highways to a neighbor's gasoline engine powered leaf blower, might be considered as one of the largest annoyances in an urban setting. Militarily, how sound propagates in a city is becoming more relevant as the battlefield is more often moving into the streets. The problems of urban noise and sound field in and around streets have been and continue to be topics of interest in the literature.[1–4] Localizing sound sources in the complex urban environment has also been of recent interest.[5] Many models that attempt to describe the sound field in and around urban street canyons exist, some with supporting experiments and some without.[6–9] Experimental efforts in urban acoustics have also often been concerned with the sound field in street canyons and the impact of road and train noise.[10–12] While some experiments have actually been performed in streets, success has also been achieved by using scale models in experiments.[10,13] Somewhat overlooked, however, are cases that might be considered fundamental to urban acoustics, e.g., sound propagation near an isolated building.

Two recent examples of efforts to characterize fundamental cases of urban acoustics are the work of Liu and Albert[14] and the work of Heimann.[15] Liu and Albert described the study of a right-angle wall of basic construction and its interaction with acoustic impulses. To model their observations, they use a two-dimensional finite-difference time-domain (FDTD) calculation of the coupled first-order velocity-pressure equations. The authors' comparison between experimental data and the FDTD model demonstrates that neglecting the third spatial dimension leads to very little error in the calculated wave forms and verifies the low-pass filter effect of diffraction around the wall. Heimann used a FDTD implementation of a three-dimensional linearized Euler model and, while considering turbulence, wind flow, and absorbing ground, demonstrated numerically the three-dimensional interaction between a 250 Hz tone and two finite length flat- or hipped-roof buildings separated by an asphalt street. Each process' effect on the sound level behind the building was studied, and the following observations were made: the hipped-roof building shields less than the flat-roof building; the wind flow over and around the hipped-roof building further reduces the shielding in the downwind propagation case; absorbing ground tends to lower the overall sound levels around the building; and turbulence had little effect on the shielding over the short propagation ranges investigated.

The literature has many examples of efforts to describe the pressure field on the shadowed side of noise barriers, typically seen along large highways.[16,17] Much like the wall in Ref. 14, barriers represent a set of fundamental cases in urban acoustics, and several theories exist for predicting the pressure behind barriers of varying heights, widths, and lengths with varying ground conditions.[18–21] The frequency domain model of Pierce[17] is one such diffraction model that is well suited to calculating the pressure behind a building, considering that it can readily support single and double diffractions.

The effort reported here presents the experimental study of sound propagation near an isolated building. This fundamental case of urban acoustics, as a necessary building block between the study of barriers and the study of street canyons, has been relatively untreated in the literature. While propagation effects have been studied extensively around finite length thin and wide barriers, it is not clear whether the complex geometry of an isolated gabled-roof building can be equated to a barrier in all instances, specifically when oblique angles of incidence are encountered. Experimental observations reported here are compared qualitatively to the

[a]Author to whom correspondence should be addressed. Electronic mail: kirk.alberts@arl.army.mil

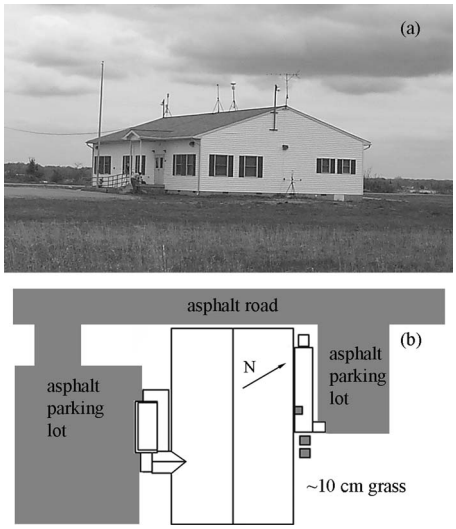


FIG. 1. (a) Photograph of the building used during the study. (b) Plan view showing the terrain around the building.

results of Liu and Albert and Heimann and quantitatively to Pierce's diffraction theory to explore the applicability of a common diffraction model to this problem.

The following section describes the apparatus and experimental procedures used in studying the propagation of sound in the vicinity of an isolated building and the procedures used to characterize the impedance of the ground around the building. The third section describes the diffraction model used during later comparisons with experimental results and the approximations made to facilitate the usage of the model. The fourth section presents data taken during the experiment and offers some qualitative and quantitative analyses of the results.

II. EXPERIMENTAL PROCEDURES

The configuration used during the experiment hinges upon four identical data acquisition boxes (boxes 0–3), each containing a single board computer and a 32 channel, 16 bit analog-to-digital converter board to collect the signals of four microphones and one temperature probe. Ultrasonic anemometer data and Global Positioning System (GPS) data, used for synchronization, were collected via serial port. The acquisitions were controlled through an Ethernet connection with each box, and acquired data were saved to a compact flash drive for later retrieval and processing. The Network Time Protocol (found at http://ntp.isc.org) coupled with the GPS units was used to discipline the system clocks of the boxes to an offset of less than 100 ms from the GPS time, thus acceptably synchronizing the boxes considering the chosen sample rate of 10 kHz. Synchronization issues due to loss of GPS signal were addressed, when possible, by time shifting the recordings of an errant box by comparing direct path lengths to recorded events.

Figure 1(a) is a photograph of the building used for this experimental work. The building is of a construction typical to that of suburban North American housing in that it is built on a concrete block foundation with wood stud walls consisting of exterior plywood sheathing under vinyl siding and

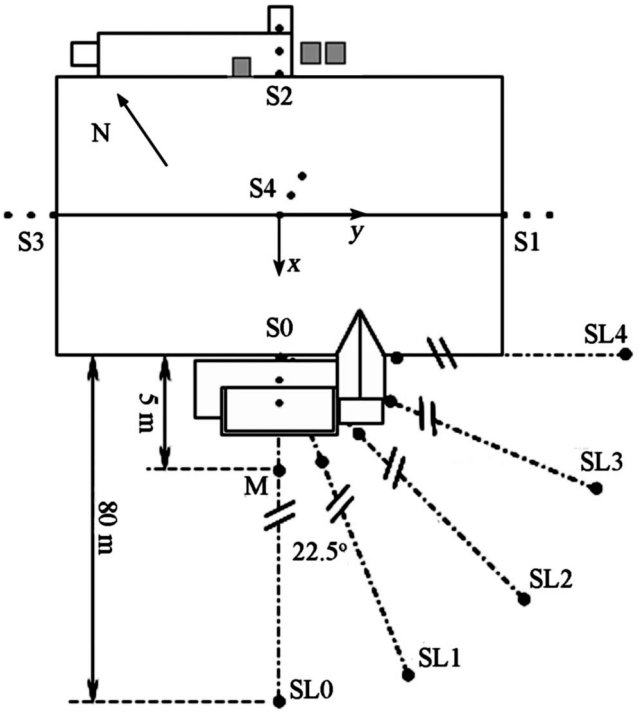


FIG. 2. Plan view of the experimental configuration.

interior gypsum board. In the cavities created by the studs is fiber glass insulation. The footprint of the building is nominally rectangular, and the roof is gabled such that the ridge runs along the long axis of the building's footprint and is covered in asphalt shingles. The uniformity of the roof is interrupted by a small covered entryway whose gable faces perpendicular to the long axis of the building. Significant obstacles in the environs of the building include an accessibility ramp and a planting bed at the southwest façade, and two large heating/air-conditioning units (HVAC) and a raised wooden deck holding a vending machine at the northeast façade (see Fig. 2). The ramp and the deck necessitate that the microphones at those positions are higher than those at the southeast and northwest façades. The line drawing in Fig. 1(b) approximates the terrain features of note around the building, specifically the asphalt parking lots and road. The larger lot, to the southwest, has an area roughly twice the footprint of the building, while the smaller lot has a footprint of roughly one-quarter the area of the building. Areas of the terrain to the southeast of the road are covered by grass ranging from approximately 10 cm to 1 m in height. To the northwest of the road, approximately 15 m from the building, is a creek that is nearly 1 km in width. The normal to the building's northeast façade points 34° clockwise from the north.

Figure 2 shows an overhead view of the experimental configuration used to monitor sound and meteorological conditions in the vicinity of the building. At the horizontal center of each façade of the building, three microphones were oriented vertically and mounted at a height of 1.3 m with separations from the façade of 5 cm, 1.05 m, and 2.05 m (the 5 cm offset is due to the diameter of the windscreens) measured perpendicular to the wall. At a distance of 1 m from each façade, an ultrasonic anemometer was placed at a height

TABLE I. Microphone coordinates relative to the horizontal center of the building at the ground level. The x and y coordinates of microphone 16 vary to stay in line between the source and the center of the façade at a distance of 75 m from the source.

Sensor position	Microphone ID	x (m)	y (m)	z (m)
S0	1	6.17	0	1.62
	2	7.17	0	1.62
	3	8.17	0	1.62
S1	4	0	9.88	1.3
	5	0	10.88	1.3
	6	0	11.88	1.3
S2	7	−6.17	0	2.05
	8	−7.17	0	2.05
	9	−8.17	0	2.05
S3	10	0	−9.88	1.30
	11	0	−10.88	1.30
	12	0	−11.88	1.30
S4	13	0	0	6.6
	14	0	0	7.6
	15	0	0	8.6
M	16	Variable	Variable	1

of approximately 1.8 m measured to the center of the anemometer's measurement area. Mounted just below the microphones was a temperature probe. To measure the sound that goes over the roof a set of three microphones was oriented horizontally and placed, measuring vertically from the ridge of the roof, at heights of 5 cm, 1.05 m, and 2.05 m. Sensor sets at each façade were numbered such that S0 was located at the southwest elevation, S1 at the southeast, S2 at the northwest, S3 at the northeast, and S4 on the roof. The 16 pressure-field microphones, half-inch B&K model 4192, used were identified as follows: microphones 1–3 were at S0, 4–6 at S1, 7–9 at S2, 10–12 at S3, and 13–15 at S4. At each sensor position, the lowest numbered microphone was closest to the wall, and the microphone numbers increased with increasing separation from the wall. Microphone 16, labeled M in Fig. 2, was always located at a height of 1 m and a distance of 5 m from microphone 1 during southwest firings or 5 m from microphone 7 during northeast firings and was allowed to move such that it was always between the source and microphone 1 or 7. Figure 2 also shows the different source locations (SL0–SL4) at 80 m from the center of the southwest façade. After SL0, the normal incidence case, SL1–SL4 were located at angles of incidence increasing counter clockwise in increments of 22.5°. This configuration was repeated by measuring the normal incidence case from the northeast façade and increasing the angle of incidence clockwise. All of the microphone coordinates are tabulated in Table I for easy reference. Coordinates in Table I are measured from the horizontal center of the building with positive x-values in the southwest direction, positive y-values in the southeast direction, and positive z-values measured vertically from the ground. All measurements assume a flat ground, and the height of the microphones at S0 includes the 32 cm height of the ramp and the height of the microphones at S2 includes the 75 cm height of the deck.

Sound sources used during the test were a propane cannon with a muzzle height of 42 cm and a 46 cm diameter loudspeaker broadcasting a 50 Hz square wave from the back of a truck at a height of approximately 1.3 m. The square wave was chosen in an attempt to simultaneously broadcast many frequencies while avoiding 60 Hz (and harmonics) electrical noise and while gaining an amplitude advantage over broadband noise. At 80 m the propane cannon impulse had peak positive pressures ranging from less than 20 to approximately 45 Pa. The center frequency of the impulse in the absence of the building was roughly 120 Hz. The square wave at 80 m exhibited a level of 75 dB at 20 μPa for the primary frequency. If the propane cannon is considered to act as a pipe, its 10 cm diameter muzzle approximates a point source to nearly 1100 Hz. Considering the loudspeaker as an unbaffled piston and calculating its directivity pattern shows that the loudspeaker can be approximated as a point source until nearly 500 Hz.[22]

In order to characterize the ground over which the sound was propagating, the procedures set forth by the ANSI template method for ground impedance were followed.[23] Therefore, the overall details of the experiment will not be repeated here, but specifics such as filters used and driver used will be described. The sound source used during the level difference measurements consisted of a compression driver feeding into a 3.175 cm diameter pipe 50 cm in length. This configuration acts as a point source until approximately 2.7 kHz. In order to protect the compression driver, a 150 Hz high-pass filter was placed between the amplifier and the driver. The wave form used during the experiments was white noise band limited from 0 to 2.7 kHz.

III. DIFFRACTION MODEL AND APPROXIMATIONS

Because sound propagation around a building can involve both single and double diffractions, a model for this application should support both types. The Pierce model is capable of handling single and double diffractions and is readily applied, hence the reason for its selection. The Pierce model will be used extensively, so the equations used in calculating the diffracted pressures are reproduced below. Details regarding geometrical approximations and other related information follow the description of the model.

The pressure due to a single edge can be determined from the frequency domain model of Pierce by applying the following equations, reproduced from Ref. 17, to the problem at hand:

$$p_{\text{diffr}} = \frac{e^{ikL}}{L}\frac{e^{i\pi/4}}{\sqrt{2}}[A_D(X_+) + A_D(X_-)]. \quad (1)$$

In Eq. (1), L is the total diffracted path length between source and receiver, $X_+ = X(\theta + \theta_0)$, $X_- = X(\theta - \theta_0)$, k is the wave number, $X(\theta)$ is defined as

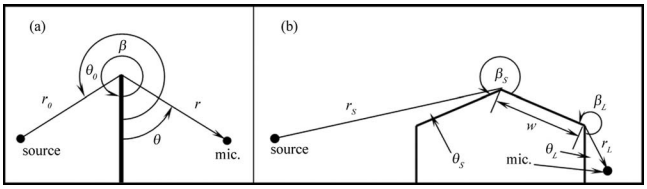


FIG. 3. Geometries used in modeling: (a) single diffraction and (b) double diffraction.

$$X(\theta) = \left[\frac{2rr_0}{\lambda L}\right]^{1/2} M_\upsilon(\theta) \tag{2}$$

for a point source with λ being the wavelength, and (in terms of auxiliary Fresnel functions) $A_D(X)$ is given by

$$A_D(X) = \mathrm{sgn}(X)[f(|X|) - ig(|X|)]. \tag{3}$$

$M_\upsilon(\theta)$, with $\upsilon = \pi/\beta$, is defined as

$$M_\upsilon(\theta) = \frac{\cos \upsilon\pi - \cos \upsilon\theta}{\upsilon \sin \upsilon\pi}. \tag{4}$$

The functions $f(X)$ and $g(X)$ depend on the Fresnel integrals,

$$C(X) = \int_0^X \cos\left(\frac{\pi t^2}{2}\right) dt,$$

$$S(X) = \int_0^X \sin\left(\frac{\pi t^2}{2}\right) dt \tag{5}$$

in the following manner:

$$f(X) = \left(\frac{1}{2} - S\right)\cos\left(\frac{\pi X^2}{2}\right) - \left(\frac{1}{2} - C\right)\sin\left(\frac{\pi X^2}{2}\right),$$

$$g(X) = \left(\frac{1}{2} - C\right)\cos\left(\frac{\pi X^2}{2}\right) + \left(\frac{1}{2} - S\right)\sin\left(\frac{\pi X^2}{2}\right). \tag{6}$$

Pressures arising from double diffractions can be calculated in a manner similar to the above by applying the following expression to appropriate paths:

$$p_{D\text{-diff}} = \frac{e^{ikL}}{L}[f(Y_>) - ig(Y_>)][f(BY_<) - ig(BY_<)]. \tag{7}$$

In Eq. (7), f and g are as defined in Eq. (6) and $Y_>$ and $Y_<$ are the greater and smaller values, respectively, of the following quantities Y_S and Y_L:

$$Y_S = \left[\frac{2r_S(w + r_L)}{\lambda L}\right]^{1/2} M_{\upsilon_S}(\beta_S - \theta_S),$$

$$Y_L = \left[\frac{2r_L(w + r_S)}{\lambda L}\right]^{1/2} M_{\upsilon_L}(\beta_L - \theta_L), \tag{8}$$

where $M_\upsilon(\theta)$ is defined in Eq. (4). All angles and distances in Eqs. (2), (4), and (8) are defined in Figs. 3(a) and 3(b), where Fig. 3(b) depicts the application of Pierce's wide barrier geometry to the building.

Typically, when calculating the pressure behind a thin barrier of infinite length on the ground, there are four propagation paths to be considered. These are source-crest-receiver, source-ground-crest-receiver, source-crest-ground-receiver, and source-ground-crest-ground-receiver. Making the barrier of finite length adds additional four paths: source-edge-receiver, source-ground-edge-receiver, and similarly for the other barrier edge. If the barrier is made wide and finite, the same eight paths exist, but double diffractions will occur. The situation in this work is very similar to that of a wide barrier of finite extent, with the exception that the gabled roof and the experimental configuration create propagation paths that obliquely impinge on edges of the roof not parallel to the ground, thus requiring some approximation.

Throughout the calculations the building has been considered rigid and has been modeled with a width of 12.24 m, a length of 19.66 m, a ridge height of 6.55 m, a roof pitch of 24.6°, an outside ridge angle of 229.3° (β_S), and an effective wall height of 3.75 m. The height of the wall was found by neglecting the eave overhang as a simplification of the building's geometry, which also yields an outside eave angle of 245.4° (β_L). This removes from the calculations two diffracted paths that reflect from the wall on the microphone side. The covered entry on the southwest façade of the building has also not been considered in the calculations. In all cases where the propagation path encountered, but did not intersect, an eave before reaching the ridge or a gable, any diffracted pressure at that eave was neglected as it would introduce a third diffraction.

For source locations with paths that intersect the ridge of the roof, SL0–SL3, the eight paths above were used in calculating the pressure. The diffracting edges modeled are the ridge of the roof, the eave closest to the microphone, and all vertical edges of the building. In those instances where the propagation path interacts with the gable end of the building, an edge not parallel to the ground, certain approximations have been made. First the point of intersection between the gable and the most direct diffracted path is determined. That is then used to calculate the outside angle from the roof to the wall. That angle is then used for a wedge of the same height as the intersection point, but of infinite length and with a crest parallel to the ground. This removes the complication of the roof pitch, thereby simplifying the calculations. In the case where both gables are crossed, SL4 propagating to S3, the intersection point between the path and the incident façade is found as above. Subsequently the intersection point with the second gable is found, and then the roof is approximated as a wide barrier perpendicular to the propagation path with edge heights equal to those of the intersection points.

The lengths of paths that include a ground reflection are calculated using images and the reflected pressure is then modified by the spherical wave reflection coefficient incorporating the level difference results reported in the following section.[24] Ground-reflected paths on the microphone side of the building are calculated by determining the microphone image position about the plane of the wooden deck or of the concrete accessibility ramp, depending on the source location. The pressure along each path is also multiplied by an atmospheric absorption term.[24] The total pressure at a non-line-of-sight (NLOS) microphone is then the sum of the pressures due to the possible propagation paths as in Ref. 24, i.e.,

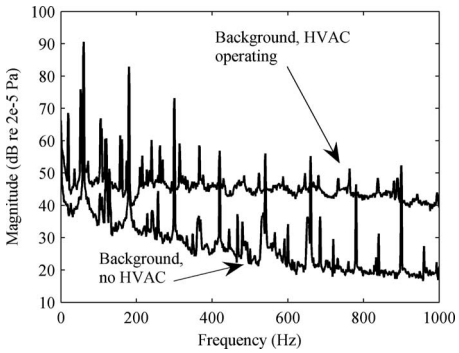


FIG. 4. Background noise levels at S2, microphone 8 with and without HVAC operating.

$$p_{\text{total}} = p_{\text{srm}}e^{-\alpha R_{\text{srm}}} + Q_s p_{\text{sgrm}}e^{-\alpha R_{\text{sgrm}}} + Q_m p_{\text{srgm}}e^{-\alpha R_{\text{srgm}}} + Q_s Q_m p_{\text{sgrgm}}e^{-\alpha R_{\text{sgrgm}}} + \cdots, \quad (9)$$

where the subscripts on the pressures, p, and on the distances, R, describe the different paths from source to microphone; i.e., R_{sgrm} represents the path length from source to ground to roof to microphone. In Eq. (9), α is the atmospheric absorption coefficient calculated as described in Appendix B.5 of Ref. 24, and Q_s and Q_m are the spherical wave reflection coefficients for the ground on the source side and on the microphone side of the building, respectively. In instances where diffraction over the roof of the building is considered, the angle of incidence on the ground at the microphone side of the building was very near the normal to the surface. Further, the reflecting surface below the microphones was often hard. Thus, Q_m is taken as equal to 1. Ground reflections on the source side of the building occurred in the grass, so its flow resistivity was used for all source-side ground reflections. An effective speed of sound was determined for each path by using the measured temperatures and winds listed below in $c_{\text{eff}}=c(T)+u$, where $c(T)$ is the sound speed assuming dry air and u is the horizontal wind speed component in the direction of the leg of the path between the source and the first diffracting edge.[24] The determination of the wind parameters is described further in Sec. IV B.

IV. RESULTS AND ANALYSIS

A. Background noise levels

Figure 4 shows averaged background sound levels recorded by microphone 8 at S2. The upper curve in the figure is the noise level when both HVAC units were operating. The lower curve is typical of background levels at the site. Of note is the large feature in each curve of 60 Hz electrical noise and its harmonics. Also of note is the large increase in broadband level due to the operation of the HVAC units. The averaged recording of the vending machine's cooling system has been left out of the figure because its contribution to the measured level amounts to a 5–10 dB increase in broadband level over the lower curve in the figure. Comparisons between Fourier transformed spectra square wave spectra and the figure show that the frequency content apart from the 50 Hz primary and its harmonics closely resembles the content in the lower curve in Fig. 4. Although the HVAC units and the vending machine were often operating during propane cannon tests, the time-averaging process, based on each arrival of the propane cannon impulse (events not equally separated in time), forced an incoherent addition of the noise. Thus, its effect on the propane cannon results was reduced.

TABLE II. Mean magnitudes and directions with accompanying standard deviations, σ, for all source positions during SL4 firing from the southwest using the propane cannon.

Sensor position		Mean	σ
S0	Magnitude (m/s)	2.2	0.5
	Direction (°)	124.7	9.3
S1	Magnitude (m/s)	1.8	0.5
	Direction (°)	204.6	15.7
S2	Magnitude (m/s)	0.5	0.3
	Direction (°)	228.9	103.8
S3	Magnitude (m/s)	0.4	0.3
	Direction (°)	238.2	97.2

B. Ground and meteorological conditions

Level difference measurements were performed at five positions on the grass and five positions on the asphalt. Using the one-parameter model of Delany and Bazley[25] to fit, values of 2.4×10^5 and 38×10^6 Pa s/m^2 were obtained for the effective flow resistivity for grass and for asphalt, respectively. Both values appear reasonable when compared to listings for similar grounds in the ANSI standard.[23]

The experiment was performed over three noncontiguous days. Over these three days the weather was somewhat variable. The overall conditions on the first day (southwest propane cannon positions) were a high temperature of near 30 °C with a light breeze varying from the south-southeast to east-southeast and clear skies. The second day (all square wave positions) exhibited a high temperature of 16 °C with a breeze from the north-northwest to east-northeast and overcast skies. The final day of testing (northeast propane cannon positions) had a high temperature of approximately 25 °C with a light breeze from the west-northwest to north-northeast and clear skies. Wind measurements at each façade demonstrate directional characteristics (significant vertical wind components at all façades and widely varying horizontal wind directions on leeward façades) similar to those observed in Refs. 15, 26, and 27.

To account for the moving medium in the model, an effective sound speed was calculated from the measured wind information by time averaging the wind magnitude and direction in a plane parallel to the ground at each of the four sensor locations. These averaged values were then used to determine an approximate wind magnitude and direction incident on the building. Table II lists the mean values and standard deviations of the magnitudes and directions, clockwise from north, measured at each sensor location during SL4 firing from the southwest with propane cannon excitation. In Table II, it is apparent that the wind field is incident on the building from the area adjacent to S0 and S1 since the measured magnitudes and directions are the largest and most

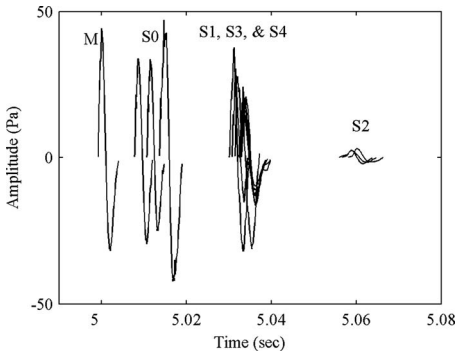


FIG. 5. Representative train of propane cannon impulses for SL0 firing from southwest.

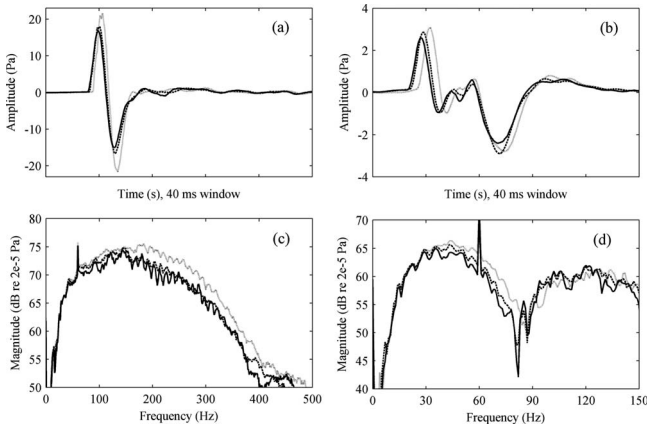


FIG. 6. Microphones 10 (solid), 11 (dashed), and 12 (gray) at S3 during (a) SL0 and (b) SL4 and [(c) and (d)] their respective Fourier transforms firing from the northeast using the propane cannon.

stable over the 5 min data collection when compared to the magnitudes and directions measured at S2 and S3. In examining the mean directions, it is noted that the wind at S0 traveled along the wall and the wind at S1 arrived at 10° angle to the wall. It can be deduced that the wind direction in the area adjacent to the S0-S1 corner was likely between the east-southeast and the south-southeast. For this source location, the wind direction is assumed to be from the southeast as it is very near the mean direction at S0. The mean speed at S0 is assumed to be the wind speed between the source and the building. Wind speeds and directions were determined in a similar manner for other source locations.

C. Representative propane cannon results

Figure 5 shows a train of impulses, in this case only first arrivals to avoid confusion, recorded during SL0 firing from the southwest. Microphone 16 (M) is the first impulse. It is followed closely by the microphones at S0 (3, 2, and 1, sequentially). The next set of three impulses reached is that corresponding to the roof microphones (13, 14, and 15, sequentially). The following six impulses correspond to the microphones on the northwest and southeast façades of the building. The last are the microphones located at the northeast façade of the building. All of the microphones that are in or near an acoustic shadow (4, 5 and 7–11) recorded impulses that demonstrate effects that can be attributed to diffraction, i.e., a decrease in peak amplitude and pulse broadening.[14] The impulse recorded by microphone 1 shows an increase in pressure compared to microphones 2 and 3, which is due to its proximity to the wall. By calculating the time delay between the first arrival at microphone 16 and the first arrival at microphone 7 and comparing that delay to possible paths, it is apparent, at least in this normal incidence case, that the dominant diffracted path is one over the roof of the building. This is in contrast to the results in Ref. 14, which showed good agreement between a two-dimensional approximation and measurements, implying that the dominant diffracted paths were around rather than over the wall under study. The differing results here are due to the scale and shape of the structure under study, which amounts to a longer path difference for around-building versus over-building propagation. In addition, the angle through which sound must diffract in this over-building case is significantly less than that diffracted over the wall in Ref. 14 in their normal incidence case. The propagation distance in the work here is also more than twice the distance of that in Ref. 14.

Three impulses recorded at S3 during SL0-northeast are shown in Fig. 6(a). In the configuration of SL0, regardless of firing direction, microphone 10 is NLOS, microphone 11 is just outside the shadow boundary, and microphone 12 is in the line of sight (LOS). Figures 6(a) and 6(c) depict, to some extent, the transition between the LOS and NLOS regions as evidenced by some features in the plots. The first feature of note is the change in amplitude between the microphones; microphone 11 exhibits peak pressures roughly 5 Pa below microphone 12 and microphone 10 is roughly 2 Pa below microphone 11. The second feature is the change in frequency content from LOS to NLOS as depicted in Fig. 6(c), Fourier transforms of the time traces in Fig. 6(a). Of note is that the peak frequency in microphone 12 shifts from roughly 200 to nearly 160 Hz in microphone 11, while only subtle changes are noted between microphones 10 and 11. In Fig. 6(b) the impulses as measured at S3 during SL4-northeast are shown. Beyond the changes in the impulse for reasons similar to those for Fig. 6(a), interference between two impulses is apparent. The ray paths, considering only paths directly from source to microphone in the absence of atmospheric effects, go from the source to the northwest corner of the building to the microphones and from the source to the southeast corner of the building to the southwest corner of the building to the microphones. Calculating these two ray paths yields a path difference of roughly 1 m, the path around the southerly side being the longest, corresponding to a time delay of approximately 3 ms, which is slightly shorter than half of the pulse duration. Similar frequency and amplitude changes are noted in Fig. 6(d), as were described for Fig. 6(c). This evidence for diffraction effects confirms expectations and lends confidence to the application to the building of the model described in Sec. III.

Figures 7(a)–7(c), each show plots of the Fourier transforms of the latest pulses to arrive in Fig. 5, which correspond to microphones 7, 8, and 9, respectively. The noise in the spectra stems from keeping a full second of data centered on each impulse to obtain 1 Hz frequency resolution in the resulting spectra and to capture all diffracted arrivals. The

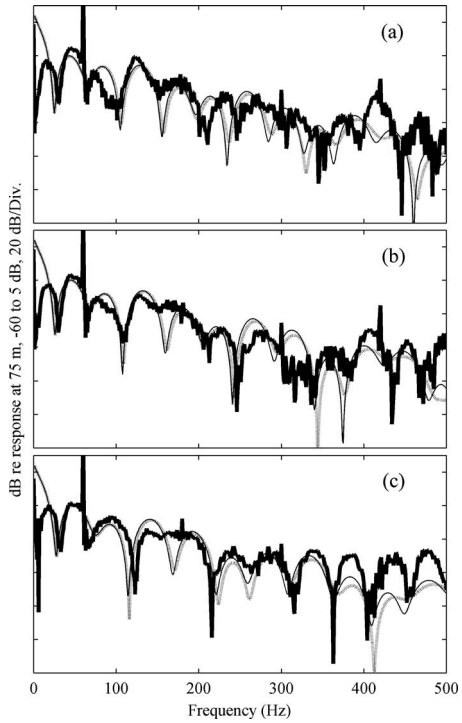


FIG. 7. Frequency content of microphones (a) 7, (b) 8, and (c) 9 (thick solid) during SL0-southwest due to propane cannon excitation and the pressure for each microphone as predicted by diffraction theory without (thin solid) and with (gray) wind (2.4 m/s at 135°) included in the sound speed.

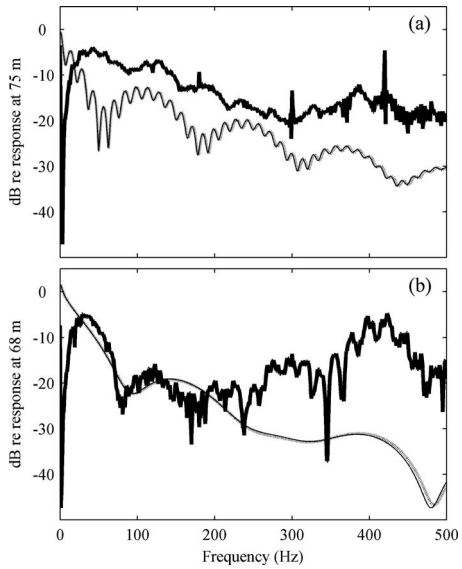


FIG. 8. Measured (thick solid) and calculated responses with (gray) and without (thin solid) wind in sound speed (a) at microphone 9 during SL3-southwest (wind: 1.8 m/s at 158°) and (b) at microphone 12 during SL4-southwest (wind: 2.2 m/s at 135°), both due to propane cannon excitation.

smooth grayed and solid curves in Fig. 7 are calculations of the pressure using the double diffraction calculation described above for all eight paths with and without wind in the effective sound speed for each microphone. In the figure, the measured and calculated curves are each normalized by the measured and calculated, respectively, initial impulse at microphone 16. This normalizing factor includes a direct and a ground-reflected path, the latter being modified by the spherical wave reflection coefficient including the measured flow resistivity for grassland. There is reasonable agreement between the curves to roughly 250 Hz, discounting the calculated minimum at 175 Hz. Above 250 Hz the agreement degrades, in part, due to imperfections in the experimental configuration and to simplifications made to the geometry of the building such as raising the wall height in order to neglect the overhang at the eaves. It is also noted that inclusion of the wind in the sound speed has only a minor effect in the areas where the calculations agree with the measurements.

As the incident angle becomes greater, the agreement between the Pierce calculation and the measurements at microphones 7–9 gets progressively worse as exemplified by Fig. 8(a), which shows the measured and calculated spectra at microphone 9 during SL2 firing from the southwest. While some of the features of the measurement are loosely recreated, the overall agreement is poor, with or without wind in the sound speed. The agreement between the calculation and the measurements at microphones 10–12 improves somewhat at frequencies below 200 Hz [see Fig. 8(b)]. The worst agreement between the model and measurements tends to occur in instances where the diffracting path crosses a single gable or a gable followed by an eave. This is likely due to the increasing severity of the geometrical approximations when the angle of incidence on a diffracting edge is large.

One benefit of the impulse excitation is that it can allow reflected and diffracted impulses to be separated and, thus, aid in determining where interference minima might occur during subsequent continuous wave excitation experiments. An example of the latter is plotted in Fig. 6(b). Neglecting possible paths over the roof and noting that only four other propagation paths (source–around southwest side–microphone, source–ground–around southwest side–microphone, source–around northeast side–microphone, and source–ground–around Northeast side–microphone) exist between the source and microphones 10, 11, and 12 during SL4, the double diffraction calculation can again be implemented. Doing so predicts a 15 dB drop directly from the magnitude at 50 Hz to the first minimum at roughly 140 Hz (subsequent minima occur at intervals of nearly 325 Hz). The calculation also shows a net decrease of over 25 dB at the highest frequencies in the range of interest.

As another example, Fig. 9(a) shows a longer time capture of the impulse recorded by microphone 16 during SL0-southwest. At 5.009 s there is a reflection that appears to arrive from a border around a planting bed in front of the building. That reflection is followed closely by a reflection off an accessibility ramp at 5.02 s. The last major reflection is that off the wall at 5.028 s. Currently, the causes of the fluctuations that occur after the wall reflection are not known. Figure 9(b) shows the response of microphone 6 recorded during SL4-northeast. The first arrival at 5 s is followed closely by the reflection off the wall. By assuming a range dependence of the form $p=\exp(ikr)/r$, where k is the wave number and r is the range from the source, then summing as in Eq. (9) over the four possible paths from the source to the microphone (source-microphone, source-ground-microphone, source-wall-microphone, and source-ground-wall-microphone), taking the reflection coefficient of

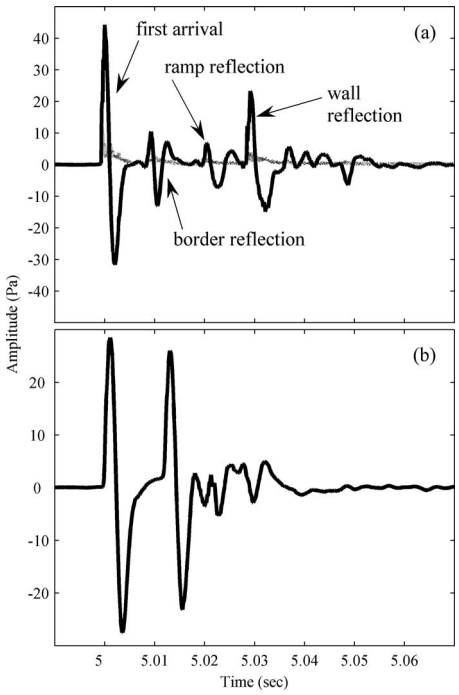


FIG. 9. (a) Reflections from various surfaces apparent in microphone 16 propane cannon recording during SL0-southwest. The dotted line in (a) represents the standard deviation of 16 impulses. (b) Microphone 6 propane cannon recording taken during SL4-northeast showing first arrival and wall reflection.

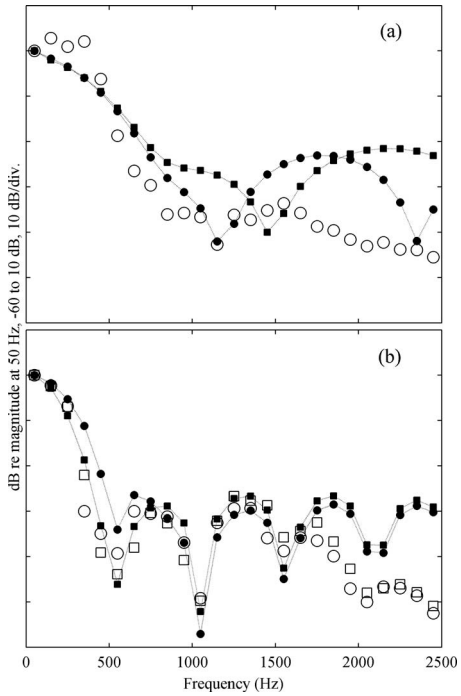


FIG. 11. (a) Response to the square wave of microphone 16, represented by open circles, recorded during SL0-southwest and calculated responses with (closed connected circles) and without (closed connected squares) wind (2 m/s, 225°) in the sound speed. (b) Square wave recording at microphone 6 during SL4-northeast (open circles) and SL4-southwest (open squares) and calculated responses with (closed connected circles) and without (closed connected squares) wind (2.1 m/s, 90°).

the wall to be 1, and utilizing the measured value of the effective flow resistivity for grassland in a calculation of the spherical wave reflection coefficient of the ground, it is found that interference minima might be observed at approximately every 85 Hz starting from roughly 40 Hz.[28]

D. Representative square wave results

The normalized spectra in Fig. 10 depict the responses of microphones 10–12 during SL4-southwest and the results of a double diffraction calculation considering all eight previously described paths. The lowest minimum, at 140 Hz, calculated previously for SL4-northeast, which should be very similar to SL4-southwest since there is a symmetry about the building's long axis, is not apparent in the measured curves, and subsequent higher frequency minima are

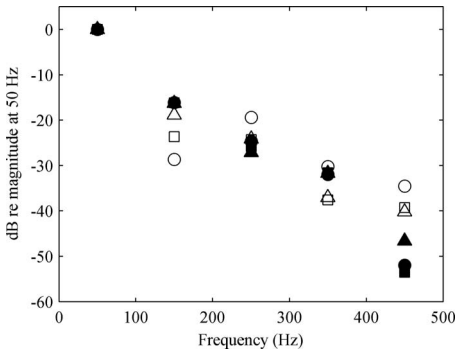


FIG. 10. Normalized microphone responses to square wave for SL4-southwest, microphones 10 (open circles), 11 (open squares), 12 (open triangles), and diffraction calculations (closed circles, squares, and triangles, respectively) including the wind (1.3 m/s at 45°) in the sound speed.

not observed in the data because the level at the microphones above 500 Hz was, unfortunately, in the noise of the acquisition system. The rather coarse sampling of the sound field by the 50 Hz square wave may be a contributing factor in the missing low-frequency minimum. It is observed in Fig. 10 that the diffraction calculation over the frequencies of the square wave agrees reasonably well with the magnitudes of the measured data until a frequency of 450 Hz. This tends to reinforce the assertion that the minimum is simply missed by the square wave.

Figure 11(a) shows the response of microphone 16 recorded during SL0-southwest normalized by its magnitude at 50 Hz. Also shown are curves calculated by assuming reflections from the ground and the wall. The calculated pressure at microphone 16 assumes that the only reflection of significance, beyond those from the ground, is that from the wall. It is apparent in Fig. 11(a) that, with the exception of the frequency locations of the minima, the agreement between the measurement and calculation is poor; however the inclusion of wind in the sound speed tends to slightly increase the agreement. Although not pictured, curves calculated by including the reflections from the border and the ramp [see Fig. 9(a)] with appropriate reflection coefficients do not improve the agreement with the measured curve. The poor agreement in Fig. 11(a) can be attributed to uncertainty in which obstacles actually cause the reflections, multiple reflections within those shown in Fig. 9(a), and the grass-asphalt impedance discontinuity. Figure 11(b) shows the responses of microphone 6 during SL4-northeast and southwest and calculated responses considering reflections from the ground and

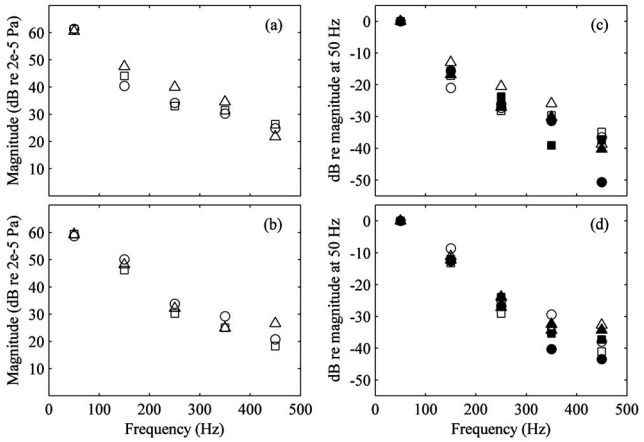


FIG. 12. NLOS microphone responses during square wave measurements for [(a) and (c)] SL0-southwest and [(b) and (d)] SL0-northeast. Microphones 7, 8, and 9 in (a) and (c) and 1, 2, and 3 in (b) and (d) are represented by unconnected open circles, open squares, and open triangles, respectively. Filled shapes represent calculated responses including wind (2 m/s at 45° for southwest and 1 m/s at 67.5° for northeast) in the sound speed.

the wall. All curves are normalized as in Fig. 11(a). The agreement between the three curves is good until a frequency of 1750 Hz where the measured curves rapidly decrease in magnitude presumably because of speaker response at higher frequency. In Fig. 11(b), both recordings of microphone 6 are shown because the propagation paths during SL4-northeast and southwest are essentially symmetric about S1 and the microphone responses are expected to be very similar. All pressures in Fig. 11 are calculated as described above and include atmospheric absorption, and the total calculated pressure is further modified by multiplying by the magnitude of the Fourier transform of an ideal 50 Hz square wave. As a first approximation, the reflection coefficient of the walls is set equal to 1.

Figure 12 shows plots of the NLOS microphones in the normal incidence cases of (a) SL0-southwest and (b) SL0-northeast and, in parts (c) and (d) of the figure, plots of normalized curves of the measurements in (a) and (b) along with curves calculated by diffraction theory. The curves in Figs. 12(c) and 12(d) have each been normalized by their respective magnitudes at 50 Hz. Figure 12 is only plotted to 500 Hz to better illustrate the observed response before the NLOS microphones reach the noise floor of about 20 dB at 20 μPa. The propagation direction during SL0-northeast was approximately downwind. Thus, considering Heimann's results,[15] the NLOS microphones in the downwind case could see an increase in level over those in the upwind case. Comparing the NLOS microphone responses in Figs. 12(a) and 12(b), it is evident that there is no consistent increase or decrease in level between the upwind and downwind propagation directions. The propagation path in the work here is roughly four times the path in Heimann's work, and the ground propagated over is almost of 10 cm tall grass. While Heimann's buildings are isolated, he considers a line source located between two identical buildings separated by a short distance. Therefore, when Heimann's configuration includes wind, the sound source is actually embedded within a vortex created by the upwind building.[15] The source in this work, at 80 m from the incident façade, is presumably located somewhat downstream or upstream from the eddies created by the building. In Figs. 12(c) and 12(d) the agreement between each of the microphones and its corresponding diffraction calculation is reasonable at some frequencies, with the exception of several outliers, which may be related to the inclusion of wind in the sound speed. In order to achieve the agreement shown, however, it was necessary to increase the effective flow resistivity of the grassland by a factor of 3. Similar results are found with increasing incident angle as were observed during comparisons to the propane cannon results.

V. CONCLUDING REMARKS

Acoustic impulse and square wave propagation in the vicinity of an isolated gabled-roof building demonstrate the complexity of this fundamental case of urban acoustics. Expected diffraction effects, specifically a strong low-pass filter effect of the multiple diffractions encountered by the propagating sound, are prevalent throughout the impulse and continuous wave excitation results. Trains of impulses show that propagation paths over the roof are significant, implying the need for inclusion of the third dimension in future models. Coarse sampling of the frequency spectrum by the square wave led to missed features in the resulting spectra.

Calculations utilizing a common diffraction model coupled to an approximated set of propagation paths reasonably agree with measurements at frequencies below 250 Hz in cases where the incident angle on the diffracting edge is nearly normal. Higher angle-of-incidence calculations and higher frequencies poorly agree with measured data presumably due to invalidation of geometrical approximations and experimental imperfections. However, the limited agreement with experimental observation in some cases implies that treating an isolated building as an effective wide barrier of finite extent is a valid low-frequency, near-normal incidence approximation. Qualitative comparisons between the experimental results presented here and existing experimental and numerical results demonstrate differences therein that can be expected considering the variation in scenarios discussed. The limitations of the model used and the complex nature of the problem necessitate further study of the measurements by more comprehensive theoretical or numerical models.

ACKNOWLEDGMENTS

This research was supported in part by an appointment to the U.S. Army Research Laboratory Postdoctoral Fellowship Program administered by the Oak Ridge Associated Universities through a contract with the U.S. Army Research Laboratory. The authors would like to thank Jack Kaiser for allowing us to repeatedly upset his routine during our experiments. One of the authors (W.C.K.A.) expresses his gratitude to Roger Waxler and Sandra Collier for many helpful conversations throughout the duration of this effort.

[1]F. M. Wiener, C. I. Malme, and C. M. Gogos, "Sound propagation in urban areas," J. Acoust. Soc. Am. **37**, 738–747 (1965).
[2]D. Aylor, J.-Y. Parlange, and C. Chapman, "Reverberation in a city street (L)," J. Acoust. Soc. Am. **54**, 1754–1757 (1973).

[3]J. P. Chambers, H. Saurenman, R. Bronsdon, L. Sutherland, R. Waxler, K. Gilbert, and C. Talmadge, "Effects of temperature induced inversion conditions on suburban highway noise levels," Acta. Acust. Acust. **92**, 1060–1070 (2006).
[4]J. Picaut, T. Le Pollès, P. L'Hermite, and V. Gary, "Experimental study of sound in a street," Appl. Acoust. **66**, 149–173 (2005).
[5]D. G. Albert, L. Liu, and M. L. Moran, "Time reversal processing for source location in an urban environment (L)," J. Acoust. Soc. Am. **118**, 616–619 (2005).
[6]T. Le Pollès, J. Picaut, and M. Bérengier, "Sound field modeling in a street canyon with partially diffusely reflecting boundaries by the transport theory," J. Acoust. Soc. Am. **116**, 2969–2983 (2004).
[7]J. Kang, "Numerical modeling of the sound fields in urban squares," J. Acoust. Soc. Am. **117**, 3695–3706 (2005).
[8]H. Onaga and J. H. Rindel, "Acoustic characteristics of urban streets in relation to scattering caused by building façades," Appl. Acoust. **68**, 310–325 (2007).
[9]J. Kang, "Sound propagation in street canyons: comparison between diffusely and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[10]K. K. Iu and K. M. Li, "The propagation of sound in narrow street canyons." J. Acoust. Soc. Am. **112**, 537–550 (2002).
[11]C. Lim, J. Kim, J. Hong, and S. Lee, "The relationship between railway noise and community annoyance in Korea," J. Acoust. Soc. Am. **120**, 2037–2042 (2006).
[12]B. De Coensel, T. De Muer, I. Yperman, and D. Bottledooren, "The influence of traffic flow dynamics on urban soundscapes," Appl. Acoust. **66**, 175–194 (2005).
[13]J. Picaut and L. Simon, "A scale model experiment for the study of sound propagation in urban areas (technical note)," Appl. Acoust. **62**, 327–340 (2001).
[14]L. Liu and D. G. Albert, "Acoustic pulse propagation near a right-angle wall," J. Acoust. Soc. Am. **119**, 2073–2083 (2006).
[15]D. Heimann, "Three-dimensional linearised euler model simulations of sound propagation in idealized urban situations with wind effects," Appl. Acoust. **68**, 217–237 (2007).
[16]K. M. Li and H. Y. Wong, "A review of commonly used analytical and empirical formulae for predicting sound diffracted by a thin screen," Appl. Acoust. **66**, 45–76 (2005).
[17]A. D. Pierce, "Diffraction of sound around corners and over wide barriers," J. Acoust. Soc. Am. **55**, 941–955 (1974).
[18]D. J. Saunders and R. D. Ford, "A study of the reduction of explosive impulses by finite sized barriers," J. Acoust. Soc. Am. **94**, 2859–2875 (1992).
[19]A. L'Espérance, "The insertion loss of finite length barriers on the ground," J. Acoust. Soc. Am. **86**, 179–183 (1989).
[20]M. Buret, K. M. Li, and K. Attenborough, "Diffraction of sound from a dipole source near to a barrier or an impedance discontinuity," J. Acoust. Soc. Am. **113**, 2480–2494 (2003).
[21]D. K. Wilson, V. E. Ostashev, S. L. Collier, N. P. Symons, D. F. Aldridge, and D. H. Marlin, "Time-domain calculations of sound interactions with outdoor ground surfaces," Appl. Acoust. **68**, 173–200 (2007).
[22]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[23]ANSI Report No. S1.18-1999, American National Standard Template Method for Ground Impedance.
[24]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, The Netherlands, 2001).
[25]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[26]J.-J. Kim and J. J. Baik, "A numerical study of the effects of ambient wind direction on flow and dispersion in urban street canyons using the RNG k-ε turbulence model," Atmos. Environ. **38**, 3039–3048 (2004).
[27]M. Skote, M. Sandberg, U. Westerberg, L. Claesson, and A. V. Johansson, "Numerical and experimental studies of wind environment in an urban morphology," Atmos. Environ. **39**, 6147–6158 (2005).
[28]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Melville, New York, 1989).

Sound-wave coherence in atmospheric turbulence with intrinsic and global intermittency[a)]

D. Keith Wilson[b)]
U.S. Army Engineer Research and Development Center, 72 Lyme Road, Hanover, New Hampshire 03755

Vladimir E. Ostashev[c)]
NOAA/Earth System Research Laboratory, Boulder, Colorado 80303 and Physics Department, New Mexico State University, Las Cruces, New Mexico 88003

George H. Goedecke[d)]
Physics Department, New Mexico State University, Las Cruces, New Mexico 88003

(Received 28 January 2008; revised 21 May 2008; accepted 21 May 2008)

The coherence function of sound waves propagating through an intermittently turbulent atmosphere is calculated theoretically. Intermittency mechanisms due to both the turbulent energy cascade (intrinsic intermittency) and spatially uneven production (global intermittency) are modeled using ensembles of quasiwavelets (QWs), which are analogous to turbulent eddies. The intrinsic intermittency is associated with decreasing spatial density (packing fraction) of the QWs with decreasing size. Global intermittency is introduced by allowing the local strength of the turbulence, as manifested by the amplitudes of the QWs, to vary in space according to superimposed Markov processes. The resulting turbulence spectrum is then used to evaluate the coherence function of a plane sound wave undergoing line-of-sight propagation. Predictions are made by a general simulation method and by an analytical derivation valid in the limit of Gaussian fluctuations in signal phase. It is shown that the average coherence function increases as a result of both intrinsic and global intermittency. When global intermittency is very strong, signal phase fluctuations become highly non-Gaussian and the average coherence is dominated by episodes with weak turbulence.
 [DOI: 10.1121/1.2945162]

PACS number(s): 43.28.Gq, 43.20.Fn, 43.28.Lv [RMW] Pages: 743–757

I. INTRODUCTION

The main purpose of this article is to examine how turbulent intermittency affects the coherence of propagating sound waves. Coherence of sound in a turbulent atmosphere is important for remote sensing [1] and assessment of performance of acoustic arrays for source localization.[2] Theoretical treatments of the coherence function (which describes the dependence of coherence on spatial separation between the observation points) for waves propagating in a random medium are well developed in many regards.[3–6] The calculated coherence function depends significantly on the model for the turbulence. In recent years, turbulence models have become more realistic in contrast to early studies that dealt with homogeneous, isotropic random scalar fields possessing a Gaussian correlation function. For example, the coherence function has now been calculated for the case of inhomogeneous, anisotropic turbulence with realistic spectra of temperature and wind velocity fluctuations.[7,8] Here, we consider the introduction of an additional realism, namely, intermittency.

Intermittency refers to the tendency of turbulence to occur in spatial and temporal bursts of activity. This property of turbulence[9,10] plays an important role in many practical problems. Mahrt[11] has proposed classifying intermittency as *intrinsic* or *global*. The former occurs on scales less than the outer scale of turbulence (the scale of largest eddies) as the turbulent cascade process progressively concentrates turbulent energy dissipation into smaller regions of space.[12] Global intermittency occurs on scales larger than the outer scale and may have several causes related to uneven production of turbulence in a particular environment. Possible causes include wind gusts from large-scale convective systems or topographic flow, irregular episodes of turbulent mixing in stably stratified (night time) boundary layers, large organized coherent structures, and uneven heating of the ground due to clouds and variations in ground-surface properties. The terms *small scale* and *large scale* are also used to describe intrinsic and global intermittency, respectively. [11,13]

Figure 1 is a conceptual illustration of intermittency and its effects on wave propagation. The eddies occur in "clouds," each of which represents a global intermittency event. Each turbulent cloud can be regarded as having its own outer scale and turbulent strength. Within each cloud, the smaller eddies organize into intermittent patches as a result of the turbulent cascade process (intrinsic intermittency). The scattering experienced by sound waves propagating along various paths through this region varies greatly,

[a)]Portions of this work were presented in V. E. Ostashev, D. K. Wilson, and G. H. Goedecke "Intermittent scalar QW model and sound propagation through intermittent turbulence," in *Proceedings of the 12th Long Range Sound Propagation Symposium*, New Orleans, LA, 2006, pp. 429–442.
[b)]Electronic mail: d.keith.wilson@usace.army.mil
[c)]Electronic mail: vladimir.ostashev@noaa.gov
[d)]Electronic mail: ggoedeck@nmsu.edu

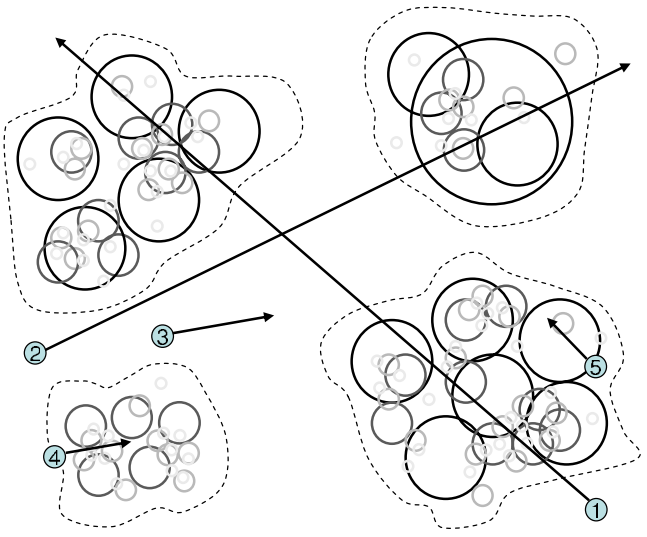


FIG. 1. (Color online) Illustration of intrinsic and global intermittency, and their effects on wave propagation. The solid circles represent turbulent eddies and the dashed lines represent global regions of active turbulence. The numbers and arrows indicate illustrative propagation paths as described in the text.

according to the number and intensity of clouds encountered. Propagation paths that are longer than or comparable to the size of the clouds, such as paths 1 and 2, experience regions of weak and strong turbulence due to global, and to a lesser extent due to intrinsic, intermittency. Paths 3–5 are short compared to the size of the clouds and therefore experience either strong or weak globally intermittent patches. Since path 3 does not actually pass through a cloud, sound waves propagating along this path experience very little scattering. Paths 4 and 5 are through similar globally strong regions, but scattering is stronger along path 4 because of differences in intrinsic intermittency.

Several previous studies have dealt with the effect of intrinsic intermittency on electromagnetic[14,15] and acoustic[16–19] propagation. The underlying idea of these studies is that if the propagating waves sample a region smaller than the outer scale, the index-of-refraction structure-function parameter C_n^2 for the scattering region (which describes the strength of the inertial-subrange spectrum in this region) becomes a random variable. The statistics of the varying C_n^2 can be described by the well-known log-normal model.[20] According to this model, the smaller the sample region, the greater the fluctuations in C_n^2.

Other studies have explicitly considered global intermittency effects on wave propagation, or provide treatments for intermittency effects that are not specifically global or intrinsic.[21–27] In these studies, it is generally assumed that the length of the propagation path is large compared to the outer scale. Hence, the fluctuations in C_n^2 resulting from intrinsic intermittency are unimportant. However, statistical parameters for the turbulence, including C_n^2, may still vary as a result of globally intermittent processes. Hentschel and Procaccia[22] considered an effect of intrinsic intermittency, namely, the modification of the well-known $-11/3$ power law for inertial-subrange turbulence, over global-scale propagation paths. Petenko and Shurygin[27] combined a two-regime model for global intermittency with a log-normal model for intrinsic intermittency.

In comparison to previous works, in this article we study the effects of combined intrinsic and global intermittency on coherence for line-of-sight, plane-wave sound propagation through a turbulent atmosphere. Of course, in order to model the propagation in intermittent turbulence, one must first model the intermittency. A sizable portion of the article is concerned with a quasiwavelet (QW) model of intermittency. A QW is a localized perturbation, analogous to a turbulent eddy, to the temperature and/or velocity field.[28,29] Several QW models of atmospheric turbulence have been developed recently, in which turbulence is represented as a collection of self-similar QWs of many different sizes.[28,30–32] These and similar models have been used for studies of wave propagation and scattering in a turbulent atmosphere.[33–37] Due to their spatially localized nature, QWs can be a useful tool for describing turbulence and other random phenomena with intrinsic and global intermittency features. QWs are also especially well suited to modeling inhomogeneous, anisotropic turbulence because, unlike customary wavelets, they may be distributed and oriented nonuniformly in space according to any desired joint probability distributions.

In this paper, we first obtain formulas for the three-dimensional (3D) spectrum, correlation function, variance, and kurtosis predicted by a QW model of temperature fluctuations with intrinsic intermittency. Then, the QW model is generalized to account for global intermittency. Although both temperature and velocity fluctuations affect the coherence of a sound wave propagating in a turbulent atmosphere, their contributions are additive and hence can reasonably be studied separately.[6,28,30] We therefore simplify the development of a QW model of turbulence with intrinsic and global intermittency by considering only temperature fluctuations. Reference 32 reports on preliminary results in developing a QW model of intermittent velocity fluctuations. Also, although we apply the model here to sound-wave propagation through turbulence, it potentially has broader applications, e.g., to electromagnetic propagation, geological heterogeneities close to Earth's surface and within the Earth,[38] boundary-layer meteorology, and eddy-based simulation of turbulence.

This paper is organized as follows. In Sec. II, a QW model of temperature fluctuations with intrinsic intermittency is developed. The model is generalized in Sec. III to include global intermittency along a one-dimensional path. In Sec. IV, this statistical framework for describing intrinsic and global intermittency is used as the basis for a theory of the coherence function of a sound wave propagating in a turbulent atmosphere. Predictions of the theory are studied with numerical simulation methods, and analytical approximations are derived and compared to the simulations. In Sec. V, the obtained results are summarized.

II. QW MODEL FOR INTERMITTENT SCALAR FLUCTUATIONS

In this section, basic formulas describing a QW model for fluctuations in scalar quantities are briefly discussed.

(The reader may refer to articles cited in the Introduction for more details.) Then, these formulas are used to develop QW models of temperature fluctuations with intrinsic and global intermittency.

A. Basic formulas

The underlying idea of the QW model is to represent a random field as a collection of randomly placed and oriented objects. The objects, like customary wavelets, are spatially localized and based on rescaling and translation of a parent function. However, some properties of customary wavelets, such as zero mean, can be relaxed. Their sizes and amplitudes can be selected according to any desired scaling laws. Lovejoy and Mandelbrot[39] have called this sort of self-similar representation a "fractal sum of pulses." Although most of the discussion to follow explicitly deals with turbulent temperature fields, the same modeling approach can be applied to other types of scalar fluctuations. For a QW model of turbulence, the individual QWs are roughly analogous to turbulent eddies.

A single temperature QW is a perturbation $\Delta T^{\alpha n}(\mathbf{R})$ to the mean field given by

$$\Delta T^{\alpha n}(\mathbf{R}) = \tau^{\alpha n} \Delta T_\alpha f(|\mathbf{R} - \mathbf{b}^{\alpha n}|/a_\alpha). \tag{1}$$

Here, $\mathbf{R}=(x,y,z)$ are the Cartesian coordinates, and $\alpha = 1,2,\ldots,N$ and $n=1,2,\ldots,N_\alpha$ are two indices, which determine a particular QW. The index α determines the size a_α of a QW: There are N QW sizes arranged in a diminishing order: $a_1 > a_2 > \cdots > a_N$. The largest a_1 and smallest a_N sizes are of the order of outer and inner scales of turbulence. In a QW model, there are N_α QWs with the same size a_α; individual QWs in a size class are distinguished by the index n. Furthermore, in Eq. (1) $\mathbf{b}^{\alpha n}$ are the coordinates of the center of the αnth QW, ΔT_α is its amplitude, $\tau^{\alpha n}$ is a random sign factor with zero mean and unit variance,[40] and f is the parent function describing the shape of the QW. Note that $\mathbf{b}^{\alpha n}$ is a random vector within the volume V where turbulence is modeled and f is the same for all QWs. Different parent functions f can be used in QW models. In what follows, a theoretical development will be done for an arbitrary parent function f. Specific results will subsequently be obtained for the Gaussian parent function, given by $f(\chi) = (2\pi)^{3/2} \exp(-\chi^2/2)$, where χ is a nondimensional argument. The overall field of temperature fluctuations $\tilde{T}(\mathbf{R})$ is the sum of contributions from the individual QWs as follows:

$$\tilde{T}(\mathbf{R}) = \sum_{\alpha=1}^{N} \sum_{n=1}^{N_\alpha} \Delta T^{\alpha n}(\mathbf{R}). \tag{2}$$

Here, the sums are taken over all QW sizes used and over all QWs within a particular size class.

A correlation function $B(\mathbf{R})$ of temperature fluctuations is defined as follows:

$$B(\mathbf{R}_1 - \mathbf{R}_2) = \langle \tilde{T}(\mathbf{R}_1) \tilde{T}(\mathbf{R}_2) \rangle, \tag{3}$$

where the brackets $\langle\ \rangle$ denote ensemble average and it is assumed that temperature fluctuations are statistically homogeneous. The 3D spectrum $\Phi(\kappa)$ of temperature fluctuations, where κ is the turbulence wave vector, is a Fourier transform of $B(\mathbf{R})$. For the case of isotropic, homogeneous turbulence, $\Phi(\kappa)$ depends only on the magnitude of the vector κ. Substituting Eqs. (1) and (2) into Eq. (3), and assuming that the $\mathbf{b}^{\alpha n}$ and $\tau^{\alpha n}$ are mutually independent, the following formula for $\Phi(\kappa)$ can be obtained by averaging over an ensemble of random realizations:[30]

$$\Phi(\kappa) = 8\pi^3 \sum_{\alpha=1}^{N} \phi_\alpha a_\alpha^3 \Delta T_\alpha^2 F^2(\kappa a_\alpha). \tag{4}$$

Here, $\phi_\alpha = N_\alpha a_\alpha^3 / V$ is the packing fraction of the QWs with the same size a_α, which is proportional to the ratio of the volume occupied by these QWs to the total volume $\mathcal{V}$ where temperature fluctuations are simulated. Furthermore, in Eq. (4) $F(\xi)$ is the spectral parent function; that is, the Fourier transform of the parent function $f(\chi)$. For the Gaussian parent function, $F(\xi) = \exp(-\xi^2/2)$.

The correlation function $B(R)$ is the inverse Fourier transform of $\Phi(\kappa)$. By the convolution theorem, the inverse transform of $F^2(\xi)$ is the convolution of $f(\chi)$ with itself. Hence, the correlation function is

$$B(R) = \sum_{\alpha=1}^{N} \phi_\alpha \Delta T_\alpha^2 f_2(R/a_\alpha), \tag{5}$$

where $f_2(\chi)$ is defined as

$$f_2(\chi) = f(\chi) * f(\chi) = \int d^3\chi' f(|\chi - \chi'|) f(\chi'). \tag{6}$$

The preceding formulas are quite general and can be used to model random fields having a range of properties. To capture the properties of turbulence, certain scaling relationships should be imposed on the QW sizes a_α, the packing fractions ϕ_α, and the temperature amplitudes ΔT_α. For example, to model statistically isotropic, homogeneous, and nonintermittent temperature fluctuations, the scaling relationships can be chosen as follows:[30]

$$a_\alpha = a_1 e^{-\mu(\alpha-1)}, \quad \phi_\alpha = \text{const}, \quad \Delta T_\alpha = \Delta T_1 (a_\alpha/a_1)^{1/3}, \tag{7}$$

where μ is a positive parameter usually much smaller than 1. According to the first of these scaling relationships, the neighboring sizes have a constant ratio ($e^{-\mu}$), which mimics a self-similar turbulent cascade of eddies with different sizes. The second relationship in Eq. (7) ensures that QWs with different sizes occupy the same total volume, which is expected for nonintermittent turbulence. Finally, the third relationship produces the classical Kolmogorov spectrum of temperature fluctuations in the inertial subrange as should be the case for isotropic, homogeneous turbulence. After substitution of Eq. (7) into Eq. (4) and some algebra, a spectrum $\Phi(\kappa)$ qualitatively similar to von Kármán's is obtained.[30]

Figure 2(a) shows an example, random QW field. This realization was made from 362 size classes ranging from $a_N = 0.2$ m to $a_1 = 10$ m. (QWs smaller than 0.2 m would not be apparent in the image.) The packing fraction ϕ is 0.0023 for all size classes and $\Delta T_1 = 0.5$ K. The QWs are distributed

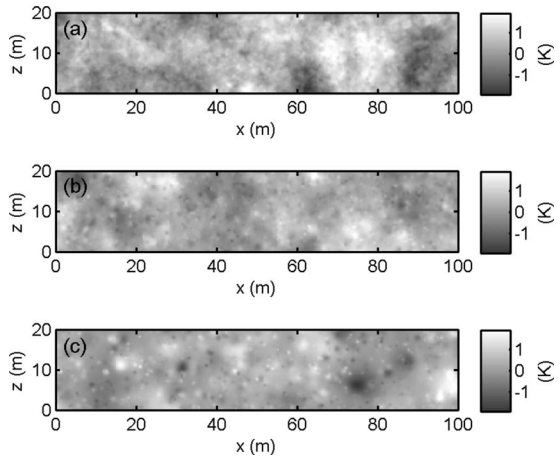


FIG. 2. Examples of synthetic turbulence fields produced by random QW ensembles. (a) Example without intermittency. (b) Example with intrinsic intermittency ($\lambda=0.5$). (c) Same as (b), except that the packing fraction has been multiplied by 1/3, and the amplitudes by $\sqrt{3}$. All scales are in K.

uniformly in a $160\times60\times80$ m^3 volume. A 100×20 m^2 cross section along the xz-plane is shown. To mitigate edge effects, the cross section is centered within the volume such that all displayed points are at least 30 m ($3a_1$) from a side of the volume.

B. Intrinsic intermittency

1. Scaling relationships and cascades

To model temperature fluctuations with intrinsic intermittency, the scaling relationships for a_α, ϕ_α, and ΔT_α must be chosen to capture the main properties of such fluctuations. Regardless of whether intermittency is present, we anticipate that the turbulence is a self-similar cascade process and thus the distribution of size classes continues to be dictated by the first relationship in Eq. (7). However, for turbulence with intrinsic intermittency, small eddies become less and less space filling[12] so that the second relationship in Eq. (7) does not hold. To accommodate this property of intermittent scalar turbulence, we assume (similarly to Ref. 32 where intermittent velocity fluctuations were considered) that the packing fraction ϕ_α decreases as the QW size becomes smaller as follows:

$$\phi_\alpha = \phi(a_\alpha/a_1)^\lambda. \tag{8}$$

Here, ϕ is a constant that is equal to the packing fraction ϕ_1 of the largest QWs and $\lambda\geqslant0$ is a parameter characterizing a degree of intrinsic intermittency. The greater the parameter λ, the more intermittent the turbulence; $\lambda=0$ corresponds to the nonintermittent case.

The cascade mechanism by which the small eddies become less space filling controls the spatial distribution of the QWs. In turbulence literature, the cascade mechanism is often conceived as a process repeated from one "generation" to the next: Parent eddies produce children eddies that are contained within, yet do not entirely fill, the space of the parents. For example, a parent eddy of size a may break down into several eddies of size $a/2$. In three dimensions, 2^3 children eddies would be needed to maintain the packing fraction. Less than 2^3 eddies would decrease the packing fraction. This description coincides to the well-known beta model of Frisch *et al.*,[12] which defines the parameter β as the number of children eddies divided by 2^3. Since $\beta=\phi_{\alpha+1}/\phi_\alpha=(a_{\alpha+1}/a_\alpha)^\lambda$ when $a_{\alpha+1}/a_\alpha=1/2$, it follows that $\beta=2^{-\lambda}$. Frisch *et al.*[12] showed that $\beta=2^{D-3}$, where $D\approx2.5$ is the fractal dimension. Hence, $\lambda=3-D\approx0.5$.

Schmitt *et al.*[41] described a procedure for "scale densification" of the beta model that removes the restriction of the parent eddy size being an integer multiple of the size of the children. Lovejoy and Scherzer generalized the beta model so that regions are not perfectly active or inactive.[42] Conceivably, the beta model or such a generalization thereof could be adapted to place the QWs in space according to a cascade process. However, there are many challenges, such as how to provide a continuous placement of the QWs and how to derive statistical results when the size classes are not independent. Also, the beta model is highly idealized: Actual eddies are swept through space during a cascade process and the cascade process may be at different ages in different regions of space. For present purposes, we do not attempt to formulate a cascade spatial construction with the QWs; the positions of QWs in each size class are independent of each other and of the positions of QWs in the preceding size class. This assumption is actually consistent with the fractal sum of pulses method described by Lovejoy and Mandelbrot.[39]

Intrinsic intermittency of turbulence can also affect the scaling relationship for QW amplitudes ΔT_α. To accommodate this possibility, the scaling relationship for ΔT_α is specified as follows:

$$\Delta T_\alpha = \Delta T_1(a_\alpha/a_1)^{1/3}(\phi/\phi_\alpha)^\eta = \Delta T_1(a_\alpha/a_1)^{1/3-\lambda\eta}, \tag{9}$$

where η is another parameter characterizing the intrinsic intermittency. If $\eta=0$, the scaling relationship described by Eq. (9) coincides with Eq. (7). The value of the parameter η depends on the problem under consideration. For example, the value of η for turbulent temperature fluctuations might differ from that for random heterogeneities within Earth. For turbulence, a value for η can be derived from conservation considerations, as has been previously demonstrated for a QW model of turbulent velocity fluctuations with intrinsic intermittency.[32] Here, we extend this treatment to conservative scalar fluctuations such as temperature.

First, consider the specific (per unit mass) turbulent kinetic energy (TKE). The TKE for size class α is proportional to $\phi_\alpha v_\alpha^2$, where v_α is the characteristic velocity scale. The *rate of transfer* of TKE from eddies with scale a_α to smaller scales should be proportional to the TKE divided by the time scale for the size class, a_α/v_α. Since kinetic energy is neither created nor destroyed within the inertial subrange, the rate of transfer should be invariant for turbulence in equilibrium. Hence, $\phi_\alpha v_\alpha^3/a_\alpha=\phi v_1^3/a_1$ for all α. Substituting with Eq. (8), we then have

$$v_\alpha = v_1(a_\alpha/a_1)^{(1-\lambda)/3}. \tag{10}$$

Similarly, the temperature variance associated with size class α is $\phi_\alpha\Delta T_\alpha^2$. If the rate of variance transfer is to be preserved (as it would be for a conservative scalar in steady turbulence), we must have $\phi_\alpha\Delta T_\alpha^2 v_\alpha/a_\alpha=\phi\Delta T_1^2 v_1/a_1$. Substituting with Eqs. (8) and (9) then yields

$$\Delta T_\alpha = \Delta T_1 (a_\alpha / a_1)^{(1-\lambda)/3} \tag{11}$$

and hence η in Eq. (9) is 1/3. The value $\eta > 0$ implies that the eddies (QWs) must spin relatively faster, and have a stronger temperature amplitude, to compensate for the decreasing packing fraction.

Figures 2(b) and 2(c) are similar to Fig. 2(a), except that intrinsic intermittency with $\lambda = 0.5$ and $\eta = 1/3$ is included. Figure 2(c) furthermore decreases the packing fraction ϕ by 1/3 (to 0.000 76) for all size classes while increasing ΔT_1 by $\sqrt{3}$ (to 0.866 K); according to Eq. (4), these rescalings do not alter the spectrum or correlation. With intrinsic intermittency, there are noticeably fewer small QWs, although they are stronger. Decreasing packing fraction also enhances the appearence of intermittency.

2. Spectra and correlations

With these revised scaling relationships, we can now calculate the spectrum $\Phi(\kappa)$ from Eq. (4). In this equation, ϕ_α and ΔT_α are replaced with their values given by Eqs. (8) and (9), respectively, and a_α is replaced with its value given by the first relationship in Eq. (7). In the resulting formula, assuming that $\mu \ll 1$, the sum over α is replaced with the integral over a using Eq. (8) from Ref. 30. As a result, we obtain a formula for the 3D spectrum of temperature fluctuations with intrinsic intermittency as follows:

$$\Phi(\kappa) = C(\kappa a_1)^{-11/3-\nu} \int_{ka_N}^{ka_1} \xi^{8/3+\nu} F^2(\xi) d\xi. \tag{12}$$

Here, the new parameter ν is a combination of the intrinsic intermittency parameters λ and η,

$$\nu \equiv \lambda(1 - 2\eta) \approx 1/6, \tag{13}$$

and the coefficient C is given by

$$C \equiv 8\pi^3 \frac{\phi a_1^3 \Delta T_1^2}{\mu}. \tag{14}$$

Equation (12) describes a 3D spectrum of temperature fluctuations with intrinsic intermittency for an arbitrary parent function $F(\xi)$. If $\nu = 0$, this spectrum coincides with that for nonintermittent temperature fluctuations [Eq. (9) in Ref. 30]. Similarly, from Eq. (5) one finds

$$B(R) = \frac{\phi \Delta T_1^2}{\mu} \left(\frac{R}{a_1} \right)^{2/3+\nu} \int_{R/a_1}^{R/a_N} \chi^{-5/3-\nu} f_2(\chi) d\chi. \tag{15}$$

For the Gaussian spectral parent function $F(\xi) = \exp(-\xi^2/2)$, the integral on the right-hand side of Eq. (12) can then be calculated as

$$\Phi_G(\kappa) = \frac{C}{2} (\kappa a_1)^{-11/3-\nu} [\gamma(11/6 + \nu/2, \kappa^2 a_1^2) - \gamma(11/6 + \nu/2, \kappa^2 a_N^2)]. \tag{16}$$

Here, $\gamma(a,x)$ is the incomplete gamma function and the subscript G stands for "Gaussian." For the case of nonintermittent turbulence, when $\nu = 0$, Eq. (16) coincides with Eq. (11) from Ref. 30 as it should. For sound propagation in a turbulent atmosphere, the sound wavelength is nearly always greater than the inner scale of turbulence, which is of order a_N.

Although Eq. (16) is specific to the Gaussian parent function, results for other reasonable parent functions have a similar dependence on wave number.[28,30] The scaling laws [Eq. (7) for classical turbulence, or Eqs. (8) and (9) for intermittent turbulence], when combined with the fractal size distribution, lead to inertial subranges with a slope independent of the parent function. The choice of parent function is mainly significant in the energy-containing subrange ($\kappa a_1 \ll 1$) and in the transition between subranges.

As for the correlation function, one can show for the Gaussian parent function that

$$f_2(\chi) = 8\pi^{9/2} e^{-\chi^2/4}. \tag{17}$$

One then finds, from Eq. (15),

$$B_G(R) = \frac{\pi^{3/2} C}{2a_1^3} \left(\frac{R}{2a_1} \right)^{2/3+\nu} [\Gamma(-1/3 - \nu/2, R^2/4a_1^2) - \Gamma(-1/3 - \nu/2, R^2/4a_N^2)], \tag{18}$$

where $\Gamma(a,x)$ is the complimentary incomplete gamma function. Since $-1/3 - \nu/2$ is negative for $\nu > -2/3$, and the complete gamma function $\Gamma(a)$ is undefined for a negative argument, it is helpful to apply the recursion formula $\Gamma(a+1,x) = a\Gamma(a,x) + x^a e^{-x}$ to rewrite Eq. (18) as

$$B_G(R) = \frac{\pi^{3/2} C}{(2/3 + \nu) a_1^3} \left[e^{-R^2/4a_1^2} - \left(\frac{R}{2a_1} \right)^{2/3+\nu} \times \Gamma(2/3 - \nu/2, R^2/4a_1^2) - \left(\frac{a_N}{a_1} \right)^{2/3+\nu} e^{-R^2/4a_N^2} + \left(\frac{R}{2a_1} \right)^{2/3+\nu} \Gamma(2/3 - \nu/2, R^2/4a_N^2) \right]. \tag{19}$$

Temperature fluctuations with scale less than a_N often do not affect the coherence function and other statistical moments for line-of-sight sound propagation.[6] Therefore, in Eqs. (16) and (19), it is often reasonable to set $a_N = 0$. The terms involving a_N thus vanish. In the energy subrange of turbulence, where $\kappa a_1 \ll 1$, the incomplete gamma function in Eq. (16) can be approximated as follows: $\gamma(11/6 + \nu/2, \kappa^2 a_1^2) \approx (\kappa a_1)^{11/3+\nu} / (11/6 + \nu/2)$. In this subrange, the 3D spectrum of temperature fluctuations does not depend on κ and is given by $\Phi_G = C/(11/3 + \nu)$. In the inertial subrange, where $\kappa a_1 \gg 1$, the incomplete gamma function in Eq. (16) can be replaced with its asymptotic value for large values of the argument. As a result (with $a_N = 0$), we have

$$\Phi_G(\kappa) = \frac{C\Gamma(11/6 + \nu/2)}{2} (\kappa a_1)^{-11/3-\nu}. \tag{20}$$

It follows from this formula that increasing intermittency (increasing ν) steepens the decay of $\Phi_G(\kappa)$ in the inertial subrange. It is worthwhile to compare $\Phi_G(\kappa)$ given by Eq. (20) with the Kolmogorov spectrum

$$\Phi(\kappa) = AC_T^2\kappa^{-11/3}, \tag{21}$$

which is valid in the inertial subrange of nonintermittent turbulence. Here, $A \simeq 0.033$ is a numerical constant and C_T^2 is the structure-function parameter for temperature fluctuations. For $\nu=0$, Eq. (20) has the same κ-dependence as the classical (nonintermittent) Kolmogorov spectrum. If we constrain the parameters in the QW representation such that

$$Ca_1^{-11/3} = 8\pi^3\frac{\phi\Delta T_1^2}{\mu a_1^{2/3}} = \frac{2A}{\Gamma(11/6+\nu/2)}C_T^2, \tag{22}$$

Eq. (22) implies $\Delta T_1 \sim a_1^{1/3}$. Equation (20) now becomes

$$\Phi_G(\kappa) = AC_T^2\kappa^{-11/3}(ka_1)^{-\nu}. \tag{23}$$

This result is consistent with Eq. 2.20 of Hentschel and Procaccia.[22] In the inertial subrange, the intermittency effect modifies the $\kappa^{-11/3}$ power law to $\kappa^{-11/3-\nu} \approx \kappa^{-23/6}$.

For $R/a_1 \ll 1$, the complementary incomplete gamma function in Eq. (18) can be replaced with the complete gamma function, and one has for the correlation function (neglecting the effect of a_N)

$$B_G(R) \approx \frac{\pi^{3/2}C}{(2/3+\nu)a_1^3}\left[1 - \left(\frac{R}{2a_1}\right)^{2/3+\nu}\Gamma(2/3-\nu/2)\right]. \tag{24}$$

The structure function for the field is defined as $D(R) = 2[B(0) - B(R)]$. For small separations, the structure function becomes

$$D_G(R) \approx \frac{\pi^{3/2}C\Gamma(2/3-\nu/2)}{(1/3+\nu/2)a_1^3}\left(\frac{R}{2a_1}\right)^{\nu+2/3}. \tag{25}$$

Substituting with Eq. (22), we have

$$D_G(R) \approx C_T^2R^{2/3}\left(\frac{R}{a_1}\right)^{\nu}, \tag{26}$$

where we have made the definition

$$A = \frac{(1/3+\nu/2)\Gamma(11/6+\nu/2)}{2^{1/3-\nu}\pi^{3/2}\Gamma(2/3-\nu/2)}. \tag{27}$$

(With this definition, $A \simeq 0.033$ when $\nu=0$, as expected.) If a_N is not identically zero, we have for $R \ll a_N$

$$B_G(R) = \frac{\pi^{3/2}C}{2a_1^3}\left(\frac{R}{2a_1}\right)^{\nu+2/3}[\Gamma(-1/3-\nu/2, R^2/4a_1^2) - \Gamma(-1/3-\nu/2, R^2/4a_N^2)]. \tag{28}$$

Using Eq. (20), the 3D spectrum $\Phi_G(\kappa)$ [normalized by $C/(11/3)$, the value of $\Phi_G(0)$ for $\nu=0$] is plotted in Fig. 3 versus the normalized wave parameter κa_1 for $\nu=0$, 0.25, and 0.5. The spectrum $\Phi_G(\kappa)$ has two distinct regions. It is almost constant in the energy subrange, where $\kappa a_1 \ll 1$, and has a power law dependence on κ in the inertial subrange, where $\kappa a_1 \gg 1$. This is consistent with asymptotic behavior of the spectrum in the energy and inertial subranges considered above. Furthermore, it follows from Fig. 3 that the greater the parameter ν, the smaller are the values of the spectrum $\Phi_G(\kappa)$. This tendency is more pronounced in the inertial subrange where the spectral slope steepens with increasing value of this parameter.

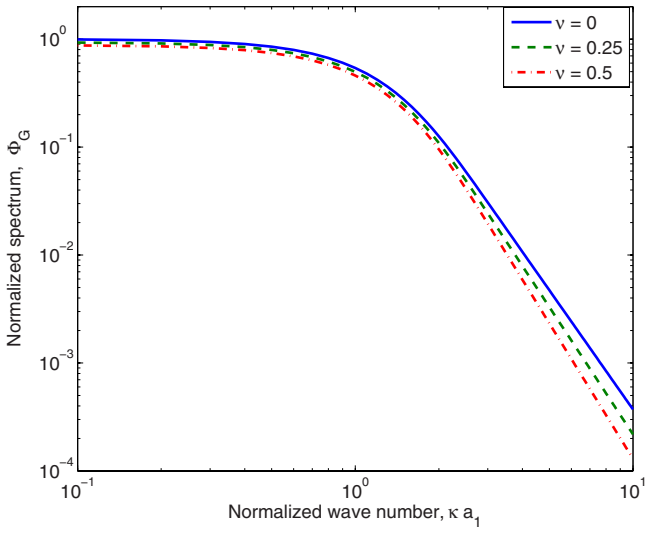


FIG. 3. (Color online) Normalized spectrum Φ_G of intermittent temperature fluctuations vs the normalized wave number κa_1 for different values of the intrinsic intermittency parameter ν.

Figure 4 compares theoretical predictions to the structure function estimated from random realizations of QW fields. The case considered is the same as Fig. 2(b), namely, $a_N=0.2$, $a_1=10$ m, $\phi=0.0023$, and $\Delta T_1=0.5$ K. The estimates were derived from correlation functions of 256 random realizations. The theoretical curves shown are the exact result (with discrete size classes) based on Eq. (5), the continuous size-class approximation based on Eq. (18), and the inertial-subrange approximation, Eq. (25). The simulations and exact result are nearly indistinguishable. For very small separations, the continuous approximation deviates somewhat from the exact result. The inertial-subrange approxima-

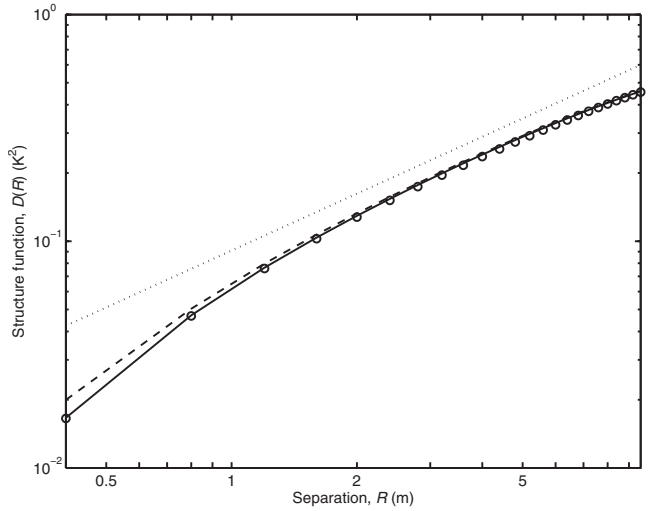


FIG. 4. Structure function for turbulence with intrinsic intermittency. The solid line is an exact curve with discrete spacing between size classes; the dashed line is the corresponding approximation for continuous size classes. The dotted line is an inertial subrange approximation. The circles show the actual structure functions determined from 256 random QW realizations.

tion is seen to be poor for this particular case; apparently, a_N is insufficiently small compared to a_1 for this approximation to be accurate.

3. Variance and kurtosis

The variance of temperature fluctuations σ^2 is needed in many applications. For both intermittent and nonintermittent turbulences, σ^2 can be determined by integrating the 3D spectrum of temperature fluctuations or by taking the limit $R \to 0$ of Eq. (15). Using the latter approach, we have from Eq. (5)

$$\sigma^2 = \sum_{\alpha=1}^{N} \sigma_\alpha^2, \tag{29}$$

where the contribution to the variance from size class α is

$$\sigma_\alpha^2 = \phi_\alpha \Delta T_\alpha^2 f_2(0). \tag{30}$$

To obtain the variance for the intermittent case, ϕ_α and ΔT_α in Eq. (29) are replaced with their values given by Eqs. (8) and (9), respectively. We thus obtain

$$\sigma_\alpha^2 = \phi \Delta T_1^2 (a_\alpha / a_1)^{2/3+\nu}. \tag{31}$$

Setting $a_\alpha = a_1 e^{-\mu(\alpha-1)}$ and recognizing Eq. (29) now as a geometric series, we explicitly determine the sum as

$$\sigma^2 = \phi \Delta T_1^2 f_2(0) \frac{1 - e^{-\mu(2/3+\nu)N}}{1 - e^{-\mu(2/3+\nu)}} \approx \frac{\phi \Delta T_1^2 f_2(0)}{\mu(2/3+\nu)}. \tag{32}$$

The second approximate form is valid when there are many closely spaced size classes ($\mu N \gg 1$ and $\mu \ll 1$). Applying Eqs. (14) and (17) yields the variance for Gaussian QWs as follows:

$$\sigma_G^2 = \frac{\pi^{3/2}}{(2/3+\nu)} \frac{C}{a_1^3}. \tag{33}$$

We see that the variance decreases with increasing intermittency. When $\nu = 1/6$, σ^2 is $4/5$ times its value without intermittency.

The kurtosis K is defined as the normalized fourth moment of a random field, namely, $K = \langle \tilde{T}^4 \rangle / \sigma^4$. Since intermittency generally leads to a kurtosis larger than the value for a Gaussian random variable, namely, $K=3$, determination of the kurtosis is of much interest (e.g., Refs. 43 and 44). A general formula for $\langle \tilde{T}^4 \rangle$ in the QW model is[45]

$$\langle \tilde{T}^4 \rangle = K_t \sum_{\alpha=1}^{N} \phi_\alpha \Delta T_\alpha^4 \Omega + 3 \left(\sum_{\alpha=1}^{N} \sigma_\alpha^2 \right)^2, \tag{34}$$

where $\Omega = \int d^3\xi f^4(\xi)$ and K_t is the fourth moment of the $\tau^{\alpha n}$ (which as described earlier, have a unit variance). For a QW model in which the $\tau^{\alpha n}$ are ± 1 with equal probability, $K_t = 1$. For a model in which the $\tau^{\alpha n}$ are normally distributed, $K_t = 3$. Equation (34) is valid for both intermittent and nonintermittent turbulence. To obtain the value of K for the intermittent case, a_α, ϕ_α, and ΔT_α are replaced with their values given by Eqs. (7)–(9), respectively. We thus obtain

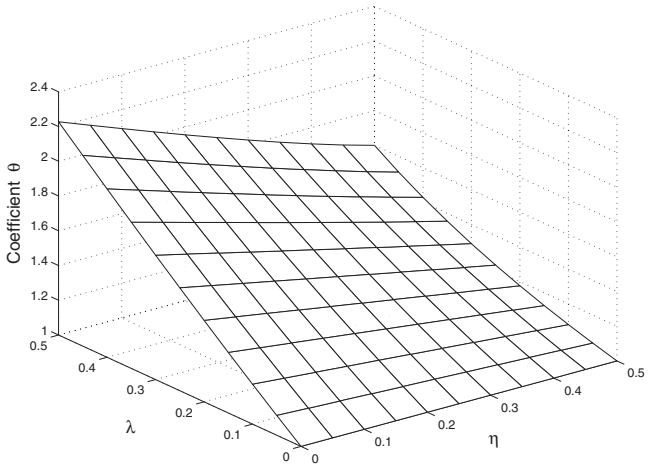


FIG. 5. Coefficient θ, which controls the kurtosis of the QW field, vs the intrinsic intermittency parameters λ and η.

$$K = 3 + \frac{\phi \Delta T_1^4 \Omega K_t}{\sigma^4} \sum_{\alpha=1}^{N} e^{-(\alpha-1)q}, \tag{35}$$

where $q \equiv \lambda(1-4\eta)$. Unlike the variance, the fourth moment [Eq. (34)] is *not* a simple sum of contributions from the size classes. However, Eq. (35) shows that the deviation from Gaussian statistics, $K-3$, can be considered as the sum of contributions from the size classes if the variance is fixed. By calculating the sum in Eq. (35), a formula for the kurtosis of temperature fluctuations with intrinsic intermittency results

$$K = 3 + \frac{\phi \Delta T_1^4 \Omega K_t}{\sigma^4} \frac{1 - e^{-\mu(4/3+q)N}}{1 - e^{-\mu(4/3+q)}} \approx 3 + \frac{\mu(2/3+\nu)^2 \Omega K_t}{\phi(4/3+q)[f_2(0)]^2}. \tag{36}$$

The second, approximate form applies when $\mu N \gg 1$ and $\mu \ll 1$.

The preceding formulas are valid for any parent function. For the Gaussian parent function,[30] $\Omega = (\pi/2)^{3/2}(2\pi)$.[6] Combining this result with Eq. (17), we obtain the kurtosis for the case of the Gaussian parent function

$$K_G = 3 + \frac{\mu \theta K_t}{3(2\pi)^{3/2}\phi}. \tag{37}$$

Here, the coefficient θ is a combination of parameters ν and q (and thus λ and η) as follows:

$$\theta = \frac{(1+3\nu/2)^2}{1+3q/4}. \tag{38}$$

For nonintermittent turbulence, when $\lambda = 0$ and $\theta = 1$, Eq. (37) coincides with Eq. (44) from Ref. 30. The kurtosis can be enhanced either by a low density of QWs (small ϕ/μ) or intrinsic intermittency ($\theta > 1$). Illustrations of the former mechanism can be found in Ref. 30.

In Fig. 5, the coefficient θ is plotted versus the parameters λ and η. It follows from this figure that θ varies in the range from 1 to 2.23 for $0 \leqslant \lambda \leqslant 0.5$ and $0 \leqslant \eta \leqslant 0.5$. Figure 5 can be used in choosing parameters of a QW model of temperature fluctuations leading to a particular value of K_G.

To test the preceding expressions, we compare them to variance and kurtosis from simulations. The cases considered are the same as shown in Fig. 2; results from 256 such random realizations were averaged. For the case without intrinsic intermittency and $\phi=0.0023$, as illustrated in Fig. 2(a), the simulation results (with theoretical values in parentheses) were $\sigma_G^2=0.420$ (0.420) and $K_G=3.41$ (3.34). For the case with intrinsic intermittency and $\phi=0.0023$, as illustrated in Fig. 2(b), the simulation results were $\sigma_G^2=0.343$ (0.350) and $K_G=3.66$ (3.56). For the case with intrinsic intermittency and $\phi=0.000\,76$, as illustrated in Fig. 2(c), the simulation results were $\sigma_G^2=0.350$ (0.350) and $K_G=4.82$ (4.68).

C. Global intermittency

As discussed earlier, global intermittency involves externally imposed variations in turbulence intensity over regions larger than the outer scale of the turbulence. In this section, we consider an approach to incorporating global intermittency into a QW model.

The volume $\mathcal{V}$ where the turbulence is modeled is conceptually subdivided into subvolumes V_i with characteristic scales larger than a_1. Within each of these subvolumes, QWs occur as described in Sec. II B 1, although the parameters may vary from one subvolume to another. In principle, any or all of the parameters ΔT_1, a_1, a_N, ϕ, μ, λ, and η could vary. Variations in the parameters ΔT_1 and a_1 are of particular interest since they depend on the external mechanism creating the turbulence. The inner scale a_N may vary but does not strongly affect sound waves. The ratio ϕ/μ and the parameter λ are expected to have a physical meaning (the packing of QWs as a function of their size) but be fixed for a particular cascade process, such as turbulence. (The individual parameters ϕ and μ are constructive in the QW model but should not vary independently for a particular cascade process.) According to Eq. (14), variations in ΔT_1 and a_1 result in random values of the coefficient C. Then, it follows from Eqs. (22) and (33) that the values of C_T^2 and σ_G^2 are also random and proportional to each other.

To proceed, we need statistical models for the variations in ΔT_1 and/or a_1, or for quantities dependent on them, such as C. Many and varied statistical models for global intermittency have been considered in literature. Mahrt,[11] Petenko and Shurygin,[27] and other authors used dichotomous regions of strong and weak activity. Antonia *et al.*[46] considered a linear ramp model. Tatarskii and Zavorotnyi[21] considered a gamma probability density function (pdf), and Frehlich[24] a log-normal pdf, for C_n^2. This diversity of approaches to modeling global intermittency reflects the different mechanisms producing it as well as the challenges of describing the underlying physics with tractable models.

In the following, we consider a basic model for global intermittency that can be readily related to a QW construction. Depending on how the parameters are adjusted, the model can be made similar to most previous treatments of global intermittency. It involves allowing the quantity C to vary with the coordinate x along a linear path; in our case, this path is the propagation path. The random field $C(x)$ is written as

$$C(x) = \langle C\rangle + C'(x) = \langle C\rangle[1 + \zeta(x)], \tag{39}$$

where $\langle C\rangle$ is the mean value and $C'(x)$ is its fluctuating part, and $\zeta(x)=C'(x)/\langle C\rangle$.

Initially, we consider a two-state Markov process for $C(x)$. (The reader may refer to a text such as Wilks[47] for an introduction to two-state Markov processes.) Within each subvolume V_i along the propagation path, the turbulence is either active (present) or inactive (absent). The inactive state is characterized by $\Delta T_1=0$ and $C=0$; the value for a_1 is thus immaterial. The values of ΔT_1, C, and a_1 are the same for all active subvolumes. Let us designate C_a as the value of C when in the active state, and the actual value of C within V_i as $C_i=C_aX_i$, where X_i is a random variable equal to 0 when V_i is an inactive state and 1 when it is active. The probability of transitioning from an inactive state in the volume V_i to an active state in V_{i+1}, while moving along the propagation path from one subvolume to the next, is designated as p_{01}. The probability of remaining in an active state is designated p_{11}, and so forth. Since there are only two possible outcomes for a transition from a given initial state, we must have $p_{00}+p_{01}=1$ and $p_{10}+p_{11}=1$. This means that only two of the transition probabilities are independent. By convention, these are usually taken to be p_{01} and p_{11}. If the Markov process has positive memory, the probability of transitioning to a particular state is greater if the process is already in that state ($p_{00}>p_{10}$, $p_{11}>p_{01}$). Also of interest are the unconditional probabilities of occurrence for the inactive and active states, designated π_0 and $\pi_1=1-\pi_0$. Since π_1 equals the sum of $p_{01}\pi_0$ and $p_{11}\pi_1$, we have

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}}. \tag{40}$$

The following relationships can also be proven:

$$\langle X_i\rangle = \langle X_i^2\rangle = \pi_1, \tag{41}$$

$$\langle X_i'^2\rangle = \langle (X_i - \langle X_i\rangle)^2\rangle = \pi_0\pi_1, \tag{42}$$

$$\langle X'_{i+j}X'_i\rangle = \pi_0\pi_1(p_{11} - p_{01})^j. \tag{43}$$

Statistics involving C_i or C_i' follow after appropriate scaling by C_a, e.g., $\langle C_i\rangle=C_a\pi_1$ and $\langle C_i'^2\rangle=C_a^2\pi_0\pi_1$. Statistics for ζ_i follow from those for X_i' after scaling by π_1^{-1}. Designating the distance between the subvolumes as ℓ, we can write Eq. (43) as

$$\langle X'(x+\Delta x)X'(x)\rangle = \pi_0\pi_1(p_{11} - p_{01})^{|\Delta x|/\ell} = \pi_0\pi_1 e^{-|\Delta x|/L}, \tag{44}$$

where $L=\ell/\ln[1/(p_{11}-p_{01})]$. The transition probabilities can be determined from the mean activity level π_1 and the correlation length by L as follows:

$$p_{01} = \pi_1(1 - e^{-\ell/L}), \quad p_{11} = \pi_1 + \pi_0 e^{-\ell/L}. \tag{45}$$

Next let us suppose $C(x)$ is the average of M independent, identical two-state Markov processes, namely,

$$C(x) = \frac{C_a}{M} \sum_{m=1}^{M} X_i^{(m)}, \tag{46}$$

where each of the $X_i^{(m)}$ follow Eqs. (40)–(43). We then find

$$\langle C(x) \rangle = C_a \pi_1, \tag{47}$$

$$B_C(\Delta x) = \langle C'(x+\Delta x) C'(x) \rangle = \frac{C_a^2 \pi_0 \pi_1}{M} e^{-|\Delta x|/L}, \tag{48}$$

$$B_\zeta(\Delta x) = \langle \zeta(x+\Delta x) \zeta(x) \rangle = \frac{\pi_0}{M \pi_1} e^{-|\Delta x|/L}. \tag{49}$$

Hence, the mean is unaffected, but the variance decreases with increasing M. Since it counts the number of "successful outcomes" of M independent trials with probability π_1, $C(x)$ has a binomial distribution. For large M and π_1 near 0.5, the binomial distribution approaches a Gaussian one. Thus, an advantage of this description for global intermittency is that it encompasses a wide range of behaviors, from two-state (purely active or inactive) to Gaussian.

The local spectrum follows from Eq. (16) but with the random $C(x)$, namely,

$$\Phi_G(x;\kappa) = \frac{C(x)}{2} (\kappa a_1)^{-11/3-\nu} [\gamma(11/6+\nu/2, \kappa^2 a_1^2) - \gamma(11/6+\nu/2, \kappa^2 a_N^2)]. \tag{50}$$

When the contribution from the term involving a_N^2 is negligible, $\kappa a_1 \gg 1$, and intrinsic intermittency is absent ($\nu=0$), our approach reduces to earlier works[21,24] on inertial-subrange intermittency where C_n^2 was allowed to vary with position. Within the volume $\mathcal{V}$, we can calculate the mean spectrum Φ_G by averaging the right-hand side of Eq. (50) over an ensemble of realizations of $C(x)$, which according to Eq. (47) amounts to replacing $C(x)$ with $C_a \pi_1$.

Figure 6 shows two realizations of QW fields with global intermittency. The positions and amplitudes of the QWs are actually the same as Fig. 2(b) (a case with intrinsic intermittency) but are then modulated by a random $C(x)$ process with mean $\langle C(x) \rangle$ matching the original field; hence, the variance of the fields is unchanged. The value of $C(x)$ at the *center position* $\mathbf{b}^{\alpha n}$ is used to tailor its amplitude $\Delta T^{\alpha n}$. Figure 6(a) is a realization of $C(x)/\langle C(x) \rangle$ with $M=1$, $\pi_1 = 0.25$, and $L=10$ m. The corresponding QW field is shown in Fig. 6(b). This model results in dramatic regions of weak and strong activity. Figure 6(c) is a realization of $C(x)/\langle C(x) \rangle$ with $M=4$, $\pi_1=0.5$, and $L=10$ m. The corresponding QW field is shown in Fig. 6(d). Although the global intermittency is much weaker than Fig. 6(b), there are still pronounced variations in turbulent activity.

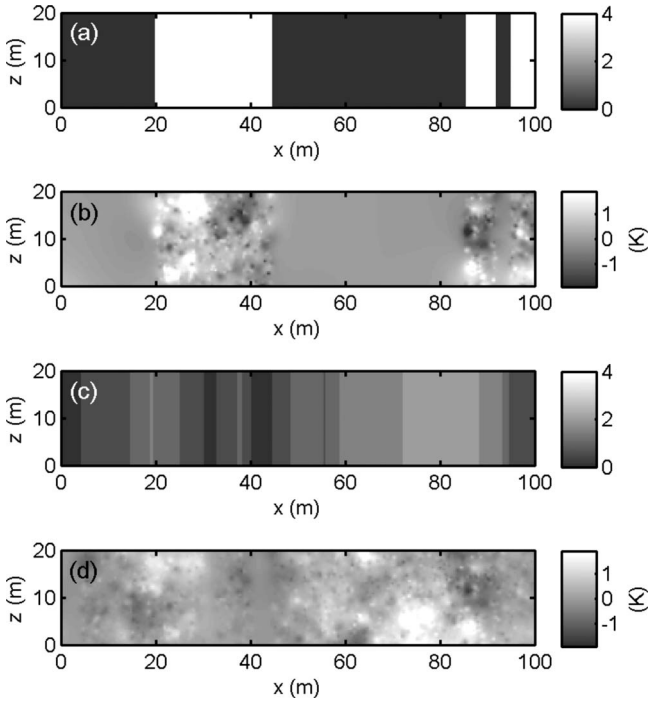


FIG. 6. Examples of synthetic turbulence fields produced by random QW ensembles with global intermittency. (a) Random global intermittency for $C(x)/\langle C \rangle$ produced by a Markov process with $\pi_1=0.25$ and $L=10$ m. (b) Same as the synthetic field in Fig. 2(b), except that the field is modulated by the process in (a). (c) Random global intermittency for $C(x)/\langle C \rangle$ produced by the average of four Markov processes with $\pi_1=0.5$ and $L=10$ m. (d) Same as the synthetic field in Fig. 2(b), except that the field is modulated by the process in (c). Scales for (a) and (c) are dimensionless; scales for (b) and (d) are in K.

III. COHERENCE FUNCTION

In this section, we derive a theory for the coherence of propagating planar sound waves based on the intermittent QW model derived in the previous two sections. The coherence function is defined here as

$$\Gamma(x,r) = \langle p(x,\mathbf{r}_0) p^*(x,\mathbf{r}_0+\mathbf{r}) \rangle. \tag{51}$$

Here, $p(x,\mathbf{r})$ is the acoustic pressure, x is the distance from the source plane to the observation plane, and $\mathbf{r}_0$ and $\mathbf{r}_0+\mathbf{r}$ are points in the observation plane. Statistical homogeneity in the observation plane is assumed, so that the coherence is independent of $\mathbf{r}_0$. As mentioned in the Introduction, theories for coherence have already been the subject of much development. Typically, several important assumptions are made: validity of the parabolic approximation, a long propagation path compared to the size of the inhomogeneities, and Gaussian statistics. Intermittency may weaken the validity of the latter two approximations. For example, atmospheric boundary-layer thermals and cloud-driven variations in surface heating, which are sources of intermittency, may extend over several kilometers. In this section, we first consider conceptually how intermittency affects the coherence, and then examine simulation results.

A. Local propagation paths

Initially, let us consider the coherence when the propagation path length x is short compared to the scale L of global intermittency. Paths 3–5 in Fig. 1 illustrate this situation. Intrinsic intermittency, as well as parameter variations induced by global intermittency, can impact on the signal propagation. For example, with the two-state Markov model for global intermittency considered in Sec. II C, the propagation path would be subject to either fully active or inactive

turbulence. This is important because we can consider the statistics of the turbulence to be local and stable over the propagation path.

Previous studies of intermittency have often examined the statistics of point samples and line (or volume) averages of the field. The coherence function actually involves a nonlinear combination of these types of statistics. The interpretation can be made clear for geometric acoustics, in which the sound follows ray paths directly from the source to the receiver. However, interpretation is less clear when diffraction becomes important. In the geometric-acoustics approximation, $p(x,r)=p_0\exp[i\phi(x,r)]$, where $p_0=p\ (x=0,r)$ for all r, and $\phi(x)$ is a random phase factor given by

$$\phi(x,r)=k\int_0^x n'(x',r)dx'. \tag{52}$$

Here, $n'(x,r)=(c_0/c)-1\approx -c'/c_0$ is a random variation in the index of refraction. (The sound speed is c, c' is its fluctuation, and c_0 its mean.) For temperature fluctuations, which in air are proportional to the squared sound speed, $n'\approx -T'/2T_0$ for temperature fluctuations. Hence, the coherence function is

$$\Gamma(x,r)=p_0^2\left\langle \exp\left[ik\int_0^x [n'(x',0)-n'(x',r)]dx' \right]\right\rangle. \tag{53}$$

This shows that the coherence function for geometric acoustics involves the difference between two line averages separated by a distance r. A nonlinear function (the exponential) of this quantity is then averaged to obtain the coherence. Usually, QWs with size comparable to r will most strongly affect coherence. QWs large compared to r are strong but tend to affect both paths in the same manner, whereas those smaller than r are relatively weaker.

If the phase factors $\phi(x,r)$ are random variables with a Gaussian distribution, the formula $\langle\exp(\varepsilon)\rangle=\exp(\langle\varepsilon^2\rangle/2)$ (where ε is Gaussian with zero mean) leads to

$$\Gamma(x,r)=p_0^2\exp[-D_\phi(x,r)/2], \tag{54}$$

where $D_\phi(x,r)$ is the geometric-acoustics phase structure function, defined as

$$D_\phi(x,r)=\langle[\phi(x,0)-\phi(x,r)]^2\rangle = \frac{k^2}{2T_0^2}\int_0^x\int_0^x [B(x'-x'',0)-B(x'-x'',r)]dx'dx''. \tag{55}$$

When the propagation path is much longer than the size of the inhomogeneities ($x\gg a_1$),

$$D_\phi(x,r)\approx\frac{k^2x}{2T_0^2}\int_{-\infty}^{\infty}[B(x',0)-B(x',r)]dx'. \tag{56}$$

Using the spectrum, the preceding result can alternatively be written as

$$D_\phi(x,r)\approx\frac{2\pi^2k^2x}{T_0^2}\int_0^\infty \kappa[1-J_0(\kappa r)]\Phi(\kappa)d\kappa. \tag{57}$$

Here, J_0 is the Bessel function of zero order. Hence, we have the following hierarchy: Equation (53) is valid for geometric acoustics, the additional assumption of Gaussian statistics leads to Eqs. (53) and (55), and the further additional assumption of a long propagation path leads to Eqs. (56) and (57). One of the remarkable results of the conventional theory of wave propagation in random media, based on Gaussian statistics and the Markov approximation, is that Eq. (54) with Eq. (57) remains valid in the parabolic approximation even when geometric-acoustics approximations do not. Put another way, when the coherence is calculated with geometric acoustics, the correct general result is obtained, even though the calculation method does not correctly capture the underlying physics involving diffraction or strong scattering. Thus, we may identify limitations to the validity of the conventional theory, such as those stemming from non-Gaussian phase statistics induced by intermittency, by calculating the coherence based on the geometric acoustics. It is plausible that the geometric-acoustics based calculations of the coherence remain correct even when there is intermittency.

Introducing now the complication that C may be a random function (but constant along a particular propagation path), we generalize Eq. (54) as

$$\Gamma_C(x,r)=p_0^2\exp[-D_{\phi,C}(x,r)/2], \tag{58}$$

where $D_{\phi,C}$ and $\Gamma_C(x,r)$ are the structure function and coherence associated with a particular value of C. For the global intermittency model described in Sec. II C, C takes on discrete values C_m following a binomial distribution, so the average coherence function is

$$\bar{\Gamma}(x,r)=\sum_{m=1}^{M}P_m\Gamma_{C_m}(x,r) = p_0^2\sum_{m=1}^{M}P_m\exp[-D_{\phi,C_m}(x,r)/2], \tag{59}$$

where P_m is the probability associated with C_m. If the $D_{\phi,C_m}(x,r)$ are small (i.e., the coherences are high for values of C_m), the approximation $\exp[-D_{\phi,C_m}(x,r)/2]\approx 1-D_{\phi,C_m}(x,r)/2$ leads to the following result:

$$\bar{\Gamma}(x,r)\approx p_0^2\exp[-\bar{D}_\phi(x,r)/2], \tag{60}$$

where

$$\bar{D}_\phi(x,r)=P_mD_{\phi,C_m}(x,r) \tag{61}$$

is the average phase structure function. Since $D_{\phi,C_m}\propto C_m$, $\bar{D}_\phi\propto\langle C\rangle$. However, if the coherence is low and the phase fluctuations are non-Gaussian, we should not expect Eq. (60) to hold.

B. Global propagation paths

We next consider propagation paths that are long compared to the scale L of global intermittency, as illustrated by

paths 1 and 2 in Fig. 1. Such paths involve a varying turbulence spectrum. A treatment for this situation follows from the equation

$$\Gamma(x,r) = p_0^2 \exp\left\{ -\frac{\pi^2 k^2}{T_0^2} \int_0^x dx' \times \int_0^\infty \kappa[1 - J_0(\kappa r)]\Phi(x';\kappa)d\kappa \right\}, \quad (62)$$

which was derived in Ref. 7 for inhomogeneous turbulence. Equation (62) assumes validity of the parabolic equation and the Markov approximation for the random medium, and that the turbulent fluctuations are Gaussian. When the strength of the turbulence varies along the propagation path as described by Eq. (50), we have the following formula for the coherence:

$$\Gamma(x,r) = p_0^2 \exp\left\{ -\frac{k^2}{a_1^2 T_0^2}[W(r/a_1,1) - W(r/a_1,a_N/a_1)] \int_0^x C(x')dx' \right\}, \quad (63)$$

where W is a deterministic function given by

$$W(\rho,s) = \frac{\pi^2}{2} \int_0^\infty \xi^{-8/3-\nu}[1 - J_0(\xi\rho)]\gamma(11/6 + \nu/2, \xi^2 s^2)d\xi. \quad (64)$$

The solution for this integral is given in the Appendix. Note that the right-hand side of Eq. (63) contains a function (the integral) of the random variable $C(x)$. Therefore, to obtain the desired expression for the mean coherence function, both sides of Eq. (63) are averaged over an ensemble of realizations of this integral. (The physical meaning of such averaging is discussed in detail in Ref. 21.) By applying Eq. (39), we can recast Eq. (63) in the following form:

$$\Gamma(x,r) = p_0^2 \exp[-\bar{D}_\phi(x,r)/2] \times \exp\left[-\frac{\bar{D}_\phi(x,r)}{2x} \int_0^x \zeta(x')dx' \right], \quad (65)$$

where

$$\bar{D}_\phi(x,r) = \frac{2\langle C\rangle k^2 x}{a_1^2 T_0^2}[W(r/a_1,1) - W(r/a_1,a_N/a_1)]. \quad (66)$$

The average coherence function is thus

$$\bar{\Gamma}(x,r) = p_0^2 \exp[-\bar{D}_\phi(x,r)/2] \times \left\langle \exp\left[-\frac{\bar{D}_\phi(x,r)}{2x} \int_0^x \zeta(x')dx' \right] \right\rangle. \quad (67)$$

The final part of this expression, with angle brackets, represents the global intermittency effect. It is bounded as follows:

$$1 \leqslant \left\langle \exp\left[-\frac{\bar{D}_\phi(x,r)}{2x} \int_0^x \zeta(x')dx' \right] \right\rangle < \exp[\bar{D}_\phi(x,r)/2]. \quad (68)$$

The first inequality follows from Jensen's inequality, as previously noted in Refs. 21 and 26. The second inequality follows from the argument that $(1/x)\int\zeta(x')dx'$ cannot be smaller than -1, which would correspond to a state of complete inactivity. Thus,

$$1 \geqslant \bar{\Gamma}(x,r)/p_0^2 \geqslant \exp[-\bar{D}_\phi(x,r)/2]. \quad (69)$$

Hence, for a fixed value of $\langle C\rangle$, intermittency always *increases* the average coherence.

According to the global intermittency model in Sec. II C, in some situations $\zeta(x)$ approaches a Gaussian distribution. Actually, since the integral of $\zeta(x)$ is effectively the sum of many samples of $\zeta(x)$ when $x \gg L$, it is even less restrictive to consider the integral as a Gaussian random variable. Assuming this is the case, we obtain the average coherence function

$$\bar{\Gamma}(x,r) = p_0^2 \exp[-\bar{D}_\phi(x,r)/2 + \bar{D}_\phi^2(x,r) I_\zeta/8x^2], \quad (70)$$

where I_ζ is the following integral:

$$I_\zeta = \int_0^x \int_0^x B_\zeta(x_1 - x_2)dx_1 dx_2 = 2\sigma_\zeta^2 Lx\left[1 - \frac{L}{x}(1 - e^{-x/L}) \right]. \quad (71)$$

The final result for I_ζ corresponds to the correlation function B_ζ given by Eq. (49), with $\sigma_\zeta^2 = B_\zeta(0) = \pi_0/M\pi_1$. Equation (70) provides a formula for the coherence function of a plane sound wave propagating in a turbulent atmosphere with intrinsic and global intermittency. It can actually be applied to local propagation paths such that $x \ll L$. Then, $e^{-x/L} \approx 1 - (x/L) + (x/L)^2/2$, and $I_\zeta \approx \sigma_\zeta^2 x^2$; that is, L ceases to affect the result. The main limitation of Eq. (70) is that integrals of the function ζ along the propagation path are assumed to have a Gaussian distribution.

Figure 7 compares coherence calculations from Eq. (70) with various combinations of intrinsic and global intermittency conditions. The turbulence parameters, based on the QW model, $a_N = 0.2$ m, $a_1 = 10$ m, $\phi/\mu = 0.0721$, and $\Delta T_1 = 0.866$ K. The path length is $x = 1$ km, the frequency 680 Hz (wavelength 0.5 m), and $L = 500$ m. The case with no intermittency ($\nu = 0$ and $\sigma_\zeta^2 = 0$) has the lowest coherence for all separations r. Intrinsic intermittency alone ($\nu = 1/6$ and $\sigma_\zeta^2 = 0$) increases the coherence for small separations ($r \lesssim a_1$), whereas global intermittency alone ($\nu = 0$ and $\sigma_\zeta^2 = 1/4$) increases the coherence for large separations. The combined effect ($\nu = 1/6$ and $\sigma_\zeta^2 = 1/4$) is an overall increase in coherence; it is nearly doubled for large separations.

Figure 8 shows the effect of changing the length scales a_1 and L. For each of the curves, $\nu = 1/6$ and $\sigma_\zeta^2 = 1/4$. One of the curves is for $a_1 = 10$ m and $L = 500$ m, as in Fig. 7. When a_1 is increased to 20 m, coherence increases for small separations but decreases for large ones. Increasing L to 2000 m, on the other hand, increases the coherence for all separations.

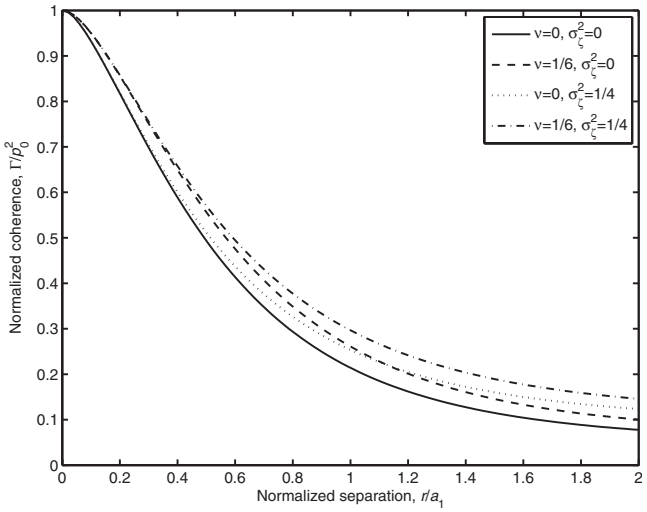


FIG. 7. Normalized coherence function vs the normalized sensor separation r/a_1 for different values of intrinsic intermittency (ν) and global intermittency σ_ξ^2. The correlation length of global intermittency $L=1$ km and the propagation path length is $x=5$ km.

C. Simulation results

In this subsection, the effects of intrinsic and global intermittency on the theoretical coherence function $\bar{\Gamma}(x,r)$ are studied with simulated random QW fields. The acoustic phases are calculated by numerically integrating Eq. (52) with the trapezoidal method. The use of geometric acoustics greatly simplifies the simulations and is justifiable for the reasons discussed in Sec. II B 2. (Simulations of random scattering by QWs, based on a finite-difference method, are described in Ref. 37. These do not involve the geometric approximation but are far more computationally intensive.) The simulations are for essentially the same situation in Figs. 7 and 8, although with intrinsic intermittency ($\nu=1/6$) always present. There are 362 QW size classes, and buffer regions were used around the edges as described in Sec. II A. Varying global intermittency conditions are considered:

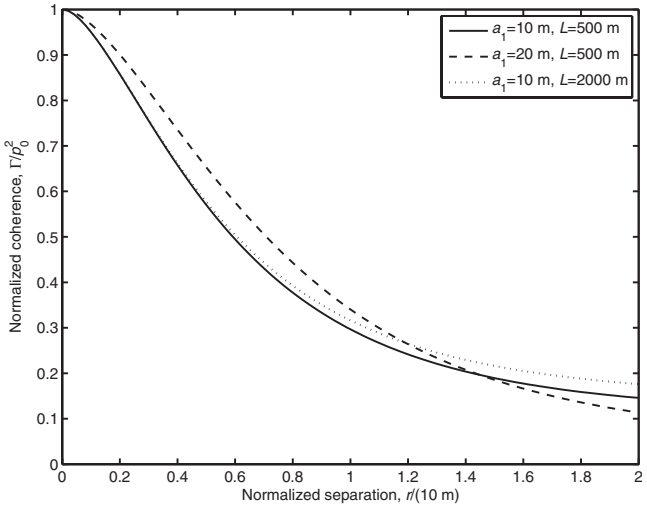


FIG. 8. Normalized coherence function vs sensor separation for different values of the outer scale a_1 and the correlation length of global intermittency L. All calculations include both intrinsic and global intermittency.

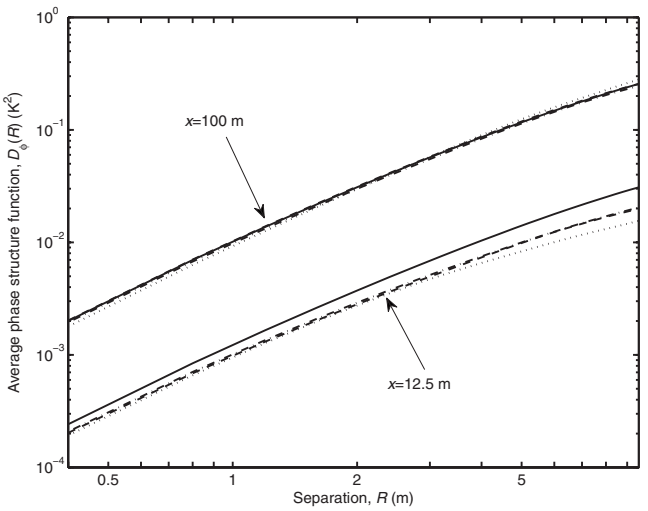


FIG. 9. Average phase structure function $\bar{D}_\phi(x,r)$ at distances $x=12.5$ m and $x=100$ m from the source, as a function of transverse separation between the observation points. The outer scale is 10 m and the correlation length of global intermittency is 500 m. The solid line is a theoretical prediction. The other lines are simulation data: Dashed is without global intermittency, dash dotted is moderate global intermittency, and dotted line is strong global intermittency.

none; $M=4$, $\pi_1=0.5$, and $L=500$ m; and $M=1$, $\pi_1=0.25$, and $L=500$ m. For the second of these cases, $\sigma_\xi^2=1/4$. We designate this case *moderate* global intermittency. The third case, for which $\sigma_\xi^2=3$, is designated *strong* global intermittency. All statistics were determined from 1024 random realizations.

Figures 9 and 10 show results for the average phase structure function, $\bar{D}_\phi(x,r)$, and average coherence, $\bar{\Gamma}(x,r)$, for propagation distances $x=12.5$ m and $x=100$ m. At these distances, the discussion in Sec. III A regarding local propagation paths ($x \ll L$) applies. The simulated curves for $\bar{D}_\phi(x,r)$ (Fig. 9) are nearly independent of the global intermittency condition, as they should be, since $\langle C \rangle$ is the same for each case. Theoretical predictions based on Eq. (66) agree very well with the simulations at $x=100$ m, but are too high at $x=12.5$ m. The reason for this overprediction is that

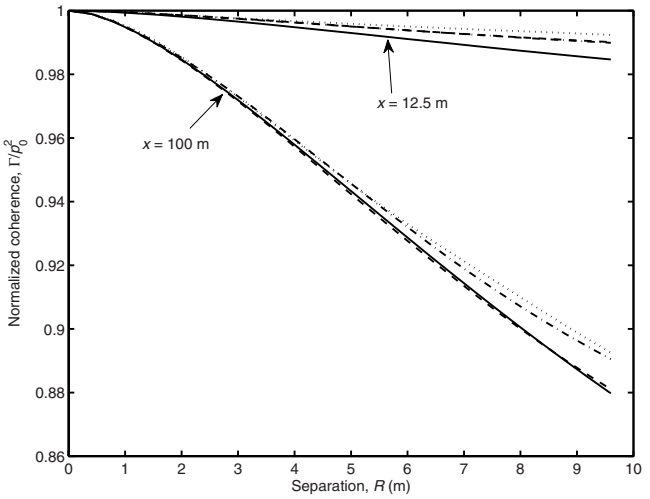


FIG. 10. (Color online) Normalized coherence function as a function of transverse separation. The curves correspond to the same cases in Fig. 9.

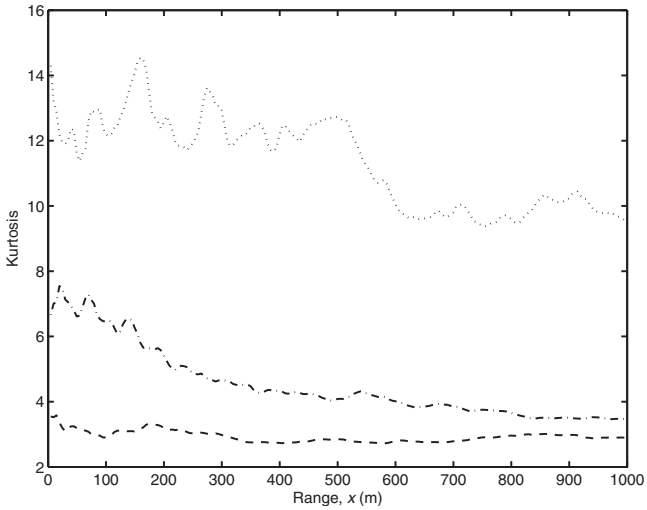


FIG. 11. Kurtosis of the phase difference fluctuations as a function of range x from the source. The separation between the observation points is 20 m. All curves are from an analysis of simulation data. The dashed line is without global intermittency, the dash-dotted line is moderate global intermittency, and the dotted line is strong global intermittency.

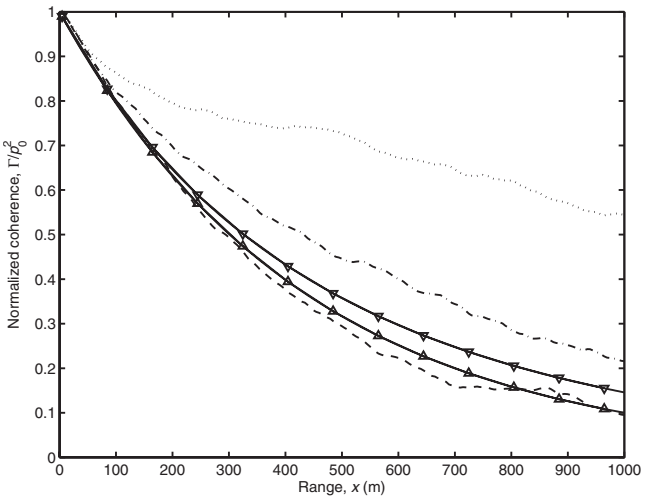


FIG. 12. Average coherence as a function of range x from the source. The separation between the observation points is 20 m. The solid lines are theoretical prediction without global intermittency (upward pointing triangles) and with moderate global intermittency (downward pointing triangles). The other lines are simulation data with the same interpretations as in Fig. 11.

$x \sim a_1$ at this distance, which means that the approximation $x \gg a_1$ inherent to Eqs. (57) and (62) is inapplicable. As might be expected, disagreement between the simulated average coherence at $x=12.5$ m and predictions based on the Gaussian approximation, Eq. (70) (Fig. 10), also results. Only one prediction from Eq. (70), corresponding to no global intermittency, is actually shown in Fig. 10 for each distance. This is because the Gaussian predictions are nearly independent of the global intermittency characteristics at these short distances. A particularly significant feature of Fig. 10 is the disagreement between the prediction and simulations at $x=100$ m when global intermittency is present, but not when global intermittency is absent. Since the average phase structure functions *are* correctly predicted (Fig. 9) in all cases, the assumption of Gaussian statistics upon which Eq. (70) is based must be at issue.

The significance of the non-Gaussian nature of the global intermittency is made clearer by Fig. 11, which shows the kurtosis of the phase differences, $[\phi(x,0)-\phi(x,r)]$ for $r=20$ m, as a function of range. The kurtosis for the case lacking global intermittency is nearly 3 at all ranges, as expected. For moderate intermittency, the kurtosis decays from 7 near the source to about 3.5 at $x=1$ km. For the strong intermittency, the kurtosis starts near 13 and remains quite high, around 10, at $x=1$ km. One would expect such strong non-Gaussianity to affect the average coherence and, indeed, as shown in Fig. 12, this is the case. The theoretical prediction and simulation without global intermittency are in good agreement. For moderate global intermittency, the simulated average coherence becomes somewhat higher than the prediction at distances $x>100$ m. No prediction is shown on the figure for strong global intermittency, because the prediction diverged in this case. The simulated coherence is dramatically raised by the strong global intermittency: It is about 0.55 at 1 km versus 0.1 without global intermittency. When there is such strong intermittency, the coherence is dominated by the many cases in which there is little turbulence along the propagation path.

IV. CONCLUSION

In this paper, a QW model of temperature fluctuations with both intrinsic and global intermittency was developed. Intrinsic intermittency was modeled by allowing the packing fraction to decrease with eddy size. Formulas for the 3D spectrum, variance, and kurtosis of temperature fluctuations were subsequently derived. It was shown that increasing the parameter ν, which characterizes the degree of intrinsic intermittency, results in decreasing variance, but increasing kurtosis, of temperature fluctuations. The increase in ν also leads to steepening of the inertial-subrange spectrum beyond the familiar $-11/3$ power law.

To introduce global intermittency into the QW model, the overall turbulence volume was partitioned into subvolumes in which intrinsic intermittency of the temperature fluctuations occurs as described above. Then, global intermittency was modeled by allowing the temperature amplitudes ΔT_1 of the largest QWs (eddies) to change randomly from one subvolume to another according to one or more superimposed Markov processes.

The developed statistical framework for describing intrinsic and global intermittency was then used as the basis for a theory of the coherence of a plane sound wave propagating through intermittent temperature fluctuations. Calculations based on this theory show that intrinsic and global intermittency both increase the average coherence. Simulations suggest that when global intermittency is very strong, such as might occur in stably stratified, night-time conditions, signal phase fluctuations become highly non-Gaussian and the coherence is dominated by episodes with little turbulence. This finding is important, e.g., for the performance of modern acoustic beam forming arrays, and should be examined experimentally. It would also be valuable to find analytical predictions for coherence in non-Gaussian conditions.

The QW model of temperature fluctuations with intrinsic and global intermittency is applicable with appropriate parameter adjustments to other intermittent scalar inhomogeneities, such as contaminants in the atmosphere and density fluctuations of geological heterogeneities within Earth, and can also be applied to other types of wave propagation, such as seismic and electromagnetic.

ACKNOWLEDGMENTS

This work was supported by U.S. Army In-House Laboratory Research Initiative (ILIR) and by U.S. Army Research Office Grant No. W911NF-06-1-0007. We thank J. C. Wyngaard and J. G. Brasseur (Pennsylvania State University) for many insightful discussions. We dedicate this manuscript to the memory of Dr. Steven Clifford and his many outstanding contributions to wave propagation in random media.

APPENDIX: SOLUTION FOR THE INTEGRAL *W*

A solution for the integral equation (64) is derived in this appendix. We begin by defining

$$I(y,s,a) = \int_0^\infty \xi^{-2a+1}\gamma(a,\xi^2 s^2)J_0(\xi y)d\xi. \tag{A1}$$

Then, Eq. (64) becomes

$$W(y,s) = (\pi^2/2)[I(0,s,a) - I(y,s,a)] \tag{A2}$$

with $a=11/6+\nu/2$.

We now determine the integral $I(y,s,a)$. First, we use Eq. (6.5.12) from Ref. 48 to write the incomplete gamma function as a confluent hypergeometric function. This gives

$$I(y,s,a) = a^{-1}s^{2a}\int_0^\infty \xi {}_1F_1(a;1+a;-\xi^2 s^2)J_0(\xi y)d\xi.$$

This integral can be solved with Eq. (7.663.6) from Ref. 49, which reads (with $\nu=0$)

$$\int_0^\infty x\,{}_1F_1(a;b;-\lambda x^2)J_0(xy)dx = \frac{2^{1-a}\Gamma(b)}{\Gamma(a)\lambda^{a/2}}y^{a-2}e^{-y^2/8\lambda}W_{a/2-b+1,a/2-1/2}\left(\frac{y^2}{4\lambda}\right). \tag{A3}$$

Here, $W_{k,\mu}$ is the Whittaker function. Setting $b=1+a$ and $\lambda=s^2$, we find

$$I(y,s,a) = 2^{1-a}s^a y^{a-2}e^{-y^2/8s^2}W_{-a/2,a/2-1/2}\left(\frac{y^2}{4s^2}\right). \tag{A4}$$

We can use Eq. (13.1.33) in Ref. 48 to rewrite this as

$$I(y,s,a) = \frac{1}{2}\left(\frac{y^2}{4}\right)^{a-1}e^{-y^2/4s^2}U\left(a,a,\frac{y^2}{4s^2}\right), \tag{A5}$$

where U is Kummer's confluent hypergeometric function. Finally, we use Eq. (13.6.28) in Ref. 48 to recast this result with an incomplete gamma function as follows:

$$I(y,s,a) = \frac{1}{2}\left(\frac{y^2}{4}\right)^{a-1}\Gamma\left(1-a,\frac{y^2}{4s^2}\right). \tag{A6}$$

A practical problem with this result for $I(y,s,a)$ is that routines for incomplete gamma functions in many numerical libraries do not allow negative arguments. In our case, without intermittency ($\nu=0$), we have $1-a=-5/6$. However, $\Gamma(a',x)$ [unlike $\gamma(a',x)$] is still a convergent function even when $a'<0$. This problem can be avoided by using the recursion relationship

$$\Gamma(a'+1,x) = a'\Gamma(a',x) + x^{a'}e^{-x}. \tag{A7}$$

Applying this recursion formula twice, we find

$$I(y,s,a) = \frac{s^{2(a-1)}}{2(a-1)}\left[\left(1-\frac{1}{a-2}\frac{y^2}{4s^2}\right)e^{-y^2/4s^2} + \frac{1}{a-2}\left(\frac{y^2}{4s^2}\right)^{a-1}\Gamma\left(3-a,\frac{y^2}{4s^2}\right)\right]. \tag{A8}$$

The argument to the incomplete gamma function is now positive. We also see that in the limit $y/s\to 0$,

$$I(0,s,a) = \frac{s^{2(a-1)}}{2(a-1)}. \tag{A9}$$

Finally, we have

$$W(y,s) = \frac{\pi^2 s^{2(a-1)}}{4(a-1)}\left[1-\left(1-\frac{1}{a-2}\frac{y^2}{4s^2}\right)e^{-y^2/4s^2} - \frac{1}{a-2}\left(\frac{y^2}{4s^2}\right)^{a-1}\Gamma\left(3-a,\frac{y^2}{4s^2}\right)\right]. \tag{A10}$$

For small arguments, $\Gamma(a,z)=\Gamma(a)-\gamma(a,z)\simeq\Gamma(a)-z^a/a$. Hence, for small y/s,

$$W(y,s) \simeq \frac{\pi^2 s^{2(a-1)}}{4(a-1)}\left\{1-\left(1-\frac{1}{a-2}\frac{y^2}{4s^2}\right)e^{-y^2/4s^2} - \frac{1}{a-2}\left(\frac{y^2}{4s^2}\right)^{a-1}\left[\Gamma(3-a)-\frac{1}{3-a}\left(\frac{y^2}{4s^2}\right)^{3-a}\right]\right\}. \tag{A11}$$

For $a=11/6$, the term proportional to $\Gamma(3-a)$ is the leading order term, and we have

$$W(y,s) \simeq \frac{\pi^2\Gamma(1/6)}{20}\left(\frac{y}{2}\right)^{5/3}. \tag{A12}$$

[1] V. Mellert and B. Schwarz-Rohr, "Correlation and coherence measurements of a spherical wave traveling in the atmospheric boundary layer," in Proceedings of the Seventh International Symposium on Long Range Sound Propagation, Lyon, France (1996), pp. 391–405.
[2] S. L. Collier, V. E. Ostashev, and D. K. Wilson, "Maximum likelihood estimation of the angle of arrival for an acoustic wave propagating in atmospheric turbulence," in Proceedings of the 2006 Meeting of the Military Sensing Symposia (MSS), Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors, Laurel, MD (2006).
[3] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[4] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4, Wave Propagation Through Random Media* (Springer, Berlin, 1989).
[5] L. A. Chernov, *Waves in Randomly-Inhomogeneous Media* (Nauka, Mos-

cow, 1975) (in Russian).
[6]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN SPON, London, 1997).
[7]V. E. Ostashev and D. K. Wilson, "Coherence function and mean field of plane and spherical sound waves propagating through inhomogeneous anisotropic turbulence," J. Acoust. Soc. Am. **115**, 497–506 (2004).
[8]V. E. Ostashev, I. P. Chunchuzov, and D. K. Wilson, "Sound propagation through and scattering by internal gravity waves in a stably stratified atmosphere," J. Acoust. Soc. Am. **118**, 3420–3429 (2005).
[9]A. N. Kolmogorov, "A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number," J. Fluid Mech. **13**, 82–85 (1962).
[10]A. M. Oboukhov, "Some specific features of atmospheric turbulence," J. Fluid Mech. **13**, 77–81 (1962).
[11]L. Mahrt, "Intermittency of atmospheric turbulence," J. Atmos. Sci. **46**, 79–95 (1989).
[12]U. Frisch, P.-L. Sulem, and M. Nelkin, "A simple dynamical model of intermittent fully developed turbulence," J. Fluid Mech. **87**, 719–736 (1978).
[13]A. Muschinski, R. G. Frehlich, and B. B. Balsley, "Small-scale and large-scale intermittency in the nocturnal boundary layer and the residual layer," J. Fluid Mech. **515**, 319–351 (2004).
[14]A. S. Gurvich and V. P. Kukharets, "The influence of intermittence of atmospheric turbulence on scattering of radio waves," Sov. J. Commun. Technol. Electron. **30**, 52–58 (1986).
[15]A. Muschinski, "Local and global statistics of clear-air Doppler radar signals," Radio Sci. **39**, RS1008 (2004).
[16]D. K. Wilson, J. C. Wyngard, and D. I. Havelock, "The effect of turbulent intermittency on scattering into an acoustic shadow zone," J. Acoust. Soc. Am. **99**, 3393–3400 (1996).
[17]D. K. Wilson, "Scattering of acoustic waves by intermittent temperature and velocity fluctuations," J. Acoust. Soc. Am. **101**, 2980–2982 (1997).
[18]D. E. Norris, D. K. Wilson, and D. W. Thomson, "Atmospheric scattering for varying degrees of saturation and turbulent intermittency," J. Acoust. Soc. Am. **109**, 1871–1880 (2001).
[19]V. E. Ostashev and D. K. Wilson, "Line-of-sight sound propagation through intermittent atmospheric turbulence," in Proceedings of the Ninth International Congress on Sound and Vibration, Orlando, FL (2002).
[20]A. S. Gurvich and A. M. Yaglom, "Breakdown of eddies and probability distributions for small-scale turbulence," Phys. Fluids **10**, 59–65 (1965).
[21]V. I. Tatarskii and V. U. Zavorotnyi, "Wave propagation in random media with fluctuating turbulent parameters," J. Opt. Soc. Am. A **2**, 2069–2076 (1985).
[22]H. G. E. Hentschel and I. Procaccia, "Passive scalar fluctuatations in intermittent turbulence with applications to wave propagation," Phys. Rev. A **28**, 417–426 (1983).
[23]R. Frehlich, "Laser scintillation measurements of the temperature spectrum in the atmospheric surface layer," J. Atmos. Sci. **49**, 1494–1509 (1992).
[24]R. Frehlich, "Effects of global intermittency on laser propagation in the atmosphere," Appl. Opt. **33**, 5764–5769 (1994).
[25]J. Gozani, "Effect of the intermittent atmosphere on laser scintillations," Opt. Lett. **24**, 436–438 (1999).
[26]J. Gozani, "Clarifying the concepts of wave propagation through intermittent media," Opt. Lett. **24**, 108–110 (1999).
[27]I. V. Petenko and E. A. Shurygin, "A two-regime model for the probability density function of the temperature structure parameter in the convective boundary layer," Boundary-Layer Meteorol. **93**, 381–394 (1999).
[28]G. H. Goedecke, V. E. Ostashev, D. K. Wilson, and H. J. Auvermann, "Quasi-wavelet model of von Kármán spectrum of turbulent velocity fluctuations," Boundary-Layer Meteorol. **112**, 33–56 (2004).
[29]G. H. Goedecke and H. J. Auvermann, "Acoustic scattering by atmospheric turbules," J. Acoust. Soc. Am. **102**, 759–771 (1997).
[30]G. H. Goedecke, D. K. Wilson, and V. E. Ostashev, "Quasi-wavelet models of turbulent temperature fluctuations," Boundary-Layer Meteorol. **120**, 1–23 (2006).
[31]G. H. Goedecke, V. E. Ostashev, and D. K. Wilson, "Quasi-wavelet models of turbulent temperature and shear-driven velocity fluctuations," in Proceedings of the 11th International Symposium on Long Range Sound Propagation, Fairlee, VT (2004), pp. 225–239.
[32]D. K. Wilson, G. H. Goedecke, and V. E. Ostashev, "Quasi-wavelet formulations of turbulence with intermittency and correlated field properties," in Proceedings of AMS-BLT Conference (2006).
[33]D. A. De Wolf, "A random motion model of fluctuations in a nearly transparent medium," Radio Sci. **18**, 138–142 (1983).
[34]W. E. McBride, H. E. Bass, R. Raspet, and K. E. Gilbert, "Scattering of sound by atmospheric turbulence: Predictions in a refractive shadow zone," J. Acoust. Soc. Am. **91**, 1336–1340 (1992).
[35]G. H. Goedecke, R. C. Wood, H. J. Auvermann, V. E. Ostashev, D. Havelock, and C. Ting, "Spectral broadening of sound scattered by advecting atmospheric turbulence," J. Acoust. Soc. Am. **109**, 1923–1934 (2001).
[36]D. K. Wilson, V. E. Ostashev, G. H. Goedecke, and H. J. Auvermann, "Quasi-wavelet calculations of sound scattering behind barriers," Appl. Acoust. **65**, 605–627 (2004).
[37]N. P. Symons, D. F. Aldridge, D. H. Marlin, D. K. Wilson, E. G. Patton, P. P. Sullivan, S. L. Collier, V. E. Ostashev, and D. P. Drob, "3D staggered-grid finite-difference simulation of sound refraction and scattering in moving media," in Proceedings of the 11th International Symposium on Long Range Sound Propagation, Fairlee, VT (2004), pp. 481–499.
[38]H. Sato and M. C. Fehler, *Seismic Wave Propagation and Scattering in the Heterogeneous Earth* (Springer, New York, 1998).
[39]S. Lovejoy and B. B. Mandelbrot, "Fractal properties of rain, and a fractal model," Tellus, Ser. A **37A**, 209–232 (1985).
[40]The simplest distribution for the τ^{an} is to set them to ± 1 with equal probability. Alternatively, they could be specified with a unit-variance Gaussian distribution. The distinction becomes important when discussing modeling field statistics higher than second order.
[41]F. Schmitt, S. Vannitsem, and A. Barbosa, "Modeling of rainfall time series using two-state renewal processes and multifractals," J. Geophys. Res. **103**, 23181–23193 (1998).
[42]S. Lovejoy and D. Scherzer, "Scale invariance, symmetries, fractals, and stochastic simulations of atmospheric phenomena," Bull. Am. Meteorol. Soc. **67**, 21–32 (1986).
[43]C. R. Chu, M. B. Parlange, G. G. Katul, and J. D. Albertson, "Probability density functions of turbulent velocity and temperature in the atmospheric surface layer," Water Resour. Res. **32**, 1681–1688 (1996).
[44]M. A. Jiménez and J. Cuxart, "Study of the probability density functions from a large-eddy simulation for a stably stratified boundary layer," Boundary-Layer Meteorol. **118**, 401–420 (2006).
[45]This formula is based on Eq. (41) in Ref. 30 [see Eq. (41) from that reference]. The primary difference is the appearance of the factor K_l, which allows various distributions to be used for the τ^{an}. The presence of this factor can be readily understood from the discussion leading up to Eq. (38) in Ref. 30 [see Eq. (41) from that reference]. Multiplication of the first term on the right side of Eq. (38) extends that equation to any distribution for the τ^{an} having a unit variance. When this extension is carried through the subsequent derivation, the modification to Eq. (41) results.
[46]R. A. Antonia, A. J. Chambers, and E. F. Bradley, "Relationships between structure functions and temperature ramps in the atmospheric surface layer," Boundary-Layer Meteorol. **23**, 395–403 (1982).
[47]D. S. Wilks, *Statistical Methods in the Atmospheric Sciences* (Academic, Oxford, 2006).
[48]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, San Francisco, CA, 1965).
[49]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, San Diego, CA, 1994).

Sound transmission at ground level in a short-grass prairie habitat and its implications for long-range communication in the swift fox *Vulpes velox*

Safi K. Darden and Simon B. Pedersen
Animal Behaviour Group, Department of Population Biology, Institute of Biology, University of Copenhagen, Tagensvej 16, DK-2000 Copenhagen N, Denmark

Ole N. Larsen
Institute of Biology, University of Southern Denmark, DK-5230 Odense M, Denmark

Torben Dabelsteen
Animal Behaviour Group, Department of Population Biology, Institute of Biology, University of Copenhagen, Tagensvej 16, DK-2000 Copenhagen N, Denmark

(Received 5 July 2007; revised 21 May 2008; accepted 28 May 2008)

The acoustic environment of swift foxes *Vulpes velox* vocalizing close to the ground and the effect of propagation on individual identity information in vocalizations were quantified in a transmission experiment in prairie habitat. Sounds were propagated (0.45 m above the ground) at distances up to 400 m. Effects of transmission were measured on three sound types: synthesized sweeps with 1.3 kHz bandwidths spanning in the range of 0.3–8.0 kHz; single elements of swift fox barking sequences (frequency range of 0.3–4.0 kHz) and complete barking sequences. Synthesized sweeps spanning 0.3–1.6 and 1.2–2.5 kHz propagated the furthest and the latter sweeps exhibited the best transmission properties for long-range propagation. Swift fox barking sequence elements are centered toward the lower end of this frequency range. Nevertheless, measurable individual spectral characteristics of the barking sequence seem to persist to at least 400 m. Individual temporal features were very consistent to at least 400 m. The communication range of the barking sequences is likely to be farther than 400 m and it should be considered a long-ranging vocalization. However, relative to the large home ranges of swift foxes (up to 16 km^2 in the experimental area) the barking sequence probably functions at intermediate distances.
 [DOI: 10.1121/1.2946704]

PACS number(s): 43.28.Gq [MCH] Pages: 758–766

I. INTRODUCTION

As sound travels through a medium changes occur in both the time and the frequency domain. This is known as sound degradation. In addition to spherical spreading, there are effects induced by atmospheric and habitat features that through medium absorption, defraction, reflection, and refraction of sound waves cause frequency dependent attenuation and amplification, reverberation, and irregular amplitude fluctuations (see Embleton, 1996). Animals that use acoustic signals are subjected to the constraints imposed by the degradation properties of their signaling environments, which can vary drastically with, for example, season (Blumenrath and Dabelsteen, 2004), time of day (Rasmussen, 1986), and signaler and receiver heights above the ground (Mathevon *et al.*, 2005). Thus, by characterizing an animal's acoustic environment as it appears during natural signaling interactions, we can test hypotheses regarding signal design and function. For example, we can estimate a signal's actual range and the range of information coded in a given set of a signal's acoustic parameters, which in turn will give us a better understanding of who the intended receivers are when an acoustic signaling event occurs.

Several studies have analyzed the coding of information in acoustic signals across taxa, in particular, information on individual identity (e.g., Frommolt *et al.*, 1997; Mathevon, 1997; Reby *et al.*, 1998). In these studies, the signals analyzed are typically long-ranging vocalizations and we would expect the perception of information coded in such signals to be affected by degradation during sound transmission from sender to receiver. Previous studies have looked at the overall effect of degradation on sound transmission and found that there seem to be varying degrees of adaptation in terms of signal design for long-range propagation (e.g., Daniel and Blumstein, 1998; Nemeth *et al.*, 2001; Perla and Slobodchikoff, 2002; Nicholls and Goldizen, 2006). Although less commonly studied, it is equally important to investigate the transmission of specific types of information coded in vocalizations (e.g., McComb *et al.*, 2003; Blumenrath *et al.*, 2004; Blumstein and Munos, 2005; Mitchell *et al.*, 2006). In this study, we aim to quantify aspects of the acoustic environment close to the ground in natural swift fox *Vulpes velox* habitat and to evaluate the degree to which individual identity information seemingly coded in a particular vocalization, the barking sequence (see Darden *et al.*, 2003a), remains intact over long distances.

Swift foxes are native to the short and mixed grass prairies of North America and occupy relatively large home ranges (range of 8–43 km^2 across the species' distribution;

Moehrenschlager *et al.*, 2004). During the mating season, December to March depending on latitude (Moehrenschlager *et al.*, 2004), the foxes regularly use long-ranging barking sequences that have highly individual characteristics (Darden *et al.*, 2003a). These characteristics probably code information on individual identity as do a number of vocalizations with individual characteristics in other species (e.g., Ceugniet and Izumi, 2004; Searby *et al.*, 2004). However, with the distances that barking sequences are likely to propagate before reaching a receiver, we are unsure of the types of information the foxes actually are making available during a natural signaling event. Barking sequences are used in a territorial context (Darden and Dabelsteen, 2008) and we expect that it is important for foxes to transmit information on individual identity during territorial conflicts and announcements. Since there is a strong effect of the ground during sound propagation close to the surface causing differential attenuation and amplification of a signal's spectral components (see Rasmussen, 1981), we predict that frequency characteristics of the vocalization cannot be used reliably by a receiver for identification, but temporal characteristics will be quite stable due to what is likely to be a very low level of reverberation in the open landscape (e.g., Wiley and Richards, 1989) of native swift fox habitat.

TABLE I. Spectral features of the model sounds in the transmission sequence (each sound represented once in the sequence) used in quantifying sound propagation in natural swift fox habitat [USWL=synthesized long upsweep, DSWL=synthesized long down sweep, USW=synthesized short upsweep, DSW=synthesized short down sweep, FB=single elements (barks) of swift fox barking sequences].

Model sound	Center frequency (kHz)	Frequency range (kHz)	Duration (ms)
USWL, DSWL	3.9	0.5–7.4	1361
USW1, DSW1	0.9	0.3–1.6	27
USW2, DSW2	1.8	1.2–2.5	27
USW3, DSW3	2.8	2.1–3.4	27
USW4, DSW4	3.7	3.0–4.3	27
USW5, DSW5	4.6	4.0–5.2	27
USW6, DSW6	5.5	4.9–6.2	27
USW7, DSW7	6.4	5.8–7.1	27
USW8, DSW8	7.4	6.7–8.0	27
FB1	1.0	0.5–2.6	142
FB2	1.0	0.4–3.1	126
FB3	1.0	0.3–2.0	113
FB4	1.2	0.5–3.1	122
FB5	1.3	0.7–3.9	161
FB6	1.3	0.6–3.9	122
FB7	1.4	0.4–4.0	152
FB8	1.4	0.6–4.0	118

II. METHODS

A. Sound transmission

We conducted a transmission experiment in native swift fox habitat on the Pawnee National Grassland (40°49′N, 104°46′W; elevation of 1650 m) in Weld County, CO, in February 2004, which is the time of year when the foxes exhibit the highest calling activity. Dominant vegetation included short grasses (key: *Bouteloua gracilis* and *Buchloe dactyloides*) and succulents (key: *Opuntia polyacatha*), with some forbs (key: *Sphaeralcea coccinea*) and half-shrubs (key: *Eriogonum effusum, Chrysothamnus nauseosus*, and *Gutierrezia sarothrae*). Average ground characteristics of the sites used include a soil density of 1.5 g/cm^3 and a vegetation to bare ground ratio of 2.2. Our transmission sequence consisted of a series of synthetically generated sounds and high quality recordings of swift fox vocalizations, each spaced by 1 s silent intervals. The synthesized sounds were generated in SIGPRO (S. B. Pedersen, DK) and consisted of a series of short upsweeps, long upsweeps, short down sweeps, long down sweeps, and pure tones (not analyzed here) in the frequency range of 0.3–8.0 kHz (see Table I for details of the synthesized sounds used in the analysis). The swift fox vocalizations used in the transmission sequence were recorded from a maximum distance of 10 m from foxes in a captive swift fox population (Cochrane Ecological Institute, Alberta, Canada) with an Audio-Technica directional microphone (AT815b, Audio-Technica, Ltd., Leeds, UK) and a SONY DAT-recorder (TDC-D8, SONY Corporation, Tokyo, Japan). We included entire barking sequences (Fig. 1) and single elements from each barking sequence (second to the last element in a sequence) in the experimental transmission sequence.

The transmission sequence was played back using a SONY Vaio laptop computer connected to a Denon DCA-600 power-amplifier (Denon Electronics, LLC) output to a JBL Control 5 speaker, which has a flat frequency response (±3 dB) from 75 Hz to 20 kHz (JBL Professional, Northridge, CA). Transmission trials were initiated on two nights with low wind forecasts (<2 m/s) and carried out between 1800 and 2000 h at two stations with flat terrain and uniform vegetation. Temperatures during trials ranged from 2 to −6 °C. Swift foxes stand about 30–32 cm at the shoulder (Scott-Brown *et al.*, 1987) and we estimated a natural sound source height of 45 cm above the ground for the center of the speaker's low frequency membrane (50–3000 Hz). Propagated sounds were recorded with a Brüel and Kjær 2236 sound pressure level (SPL) meter (Brüel and Kjær Inc., Nærum, DK) output to an HHB PDR 1000 PORTADAT recorder (HHB Communications Ltd., London, UK). The microphone was placed on a tripod at a membrane height of

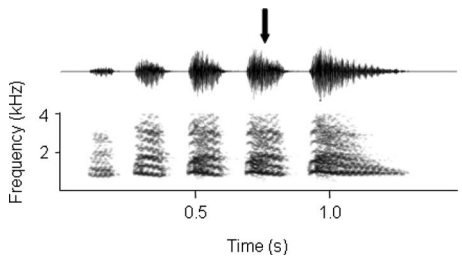


FIG. 1. Example of a swift fox barking sequence (spectrogram and waveform) with five barks (elements). The arrow indicates the placement of elements that were included singly in the transmission sequence in addition to the entire barking sequence (Avisoft SASLAB PRO 3.2, Hamming window, 512 point FFT, 100% frame size, 87.5% overlap; 43 Hz and 2.9 ms resolution).

45 cm above the ground (estimated natural listening height of a swift fox). We had nine transmission distances (distance from speaker to microphone): 10, 20, 40, 80, 120, 200, 300, and 400 m where sounds were recorded consecutively. We refer to the recordings of these propagated sounds as our *observation sounds*. The microphone gain was adjusted as needed at each distance to get acceptable signal strength. This adjustment was taken into account in all calculations resulting from these recordings. The transmission sequence was also recorded with the speaker placed on its back on a padded, nonreflective surface and the microphone suspended in the air, centered above the low frequency membrane at a distance of 1.25 m. This recording of the sequence is referred to as our *model sound* and is used as a reference, i.e., the sound that comes out of the speaker and gets transmitted through the environment. In the analysis portion, we compared all observation sounds to their respective model sounds to avoid measuring effects of the sound emitting system [computer sound card, amplifier, speaker, and associated cables that together are likely to cause a discrepancy between the input and output sounds (frequency and amplitude distortion)].

B. Sound analysis

Our observation and model sounds were digitally transferred to a PC at a 44.1 kHz sampling rate for sound analysis. We used the programs SIGPRO, MATHCAD 2001 (Mathsoft Applications, Inc.), and AVISOFT SASLAB PRO. V. 3.2 (R. Specht, DE) for different aspects of the sound analysis. Each recorded sound was bandpass filtered in SIGPRO using filter settings corresponding to the frequency range of the untransmitted sound (filter transition band, 150 Hz; filter attenuation, 30 dB) (Table I).

1. Sound propagation

To quantify the propagation properties of observation sounds, we used two of the measures described in detail by Dabelsteen *et al.* (1993) and Holland *et al.* (1998): signal-to-noise ratio (SNR) and excess attenuation (EA: attenuation in excess of spherical spreading), and an additional measure, tail-to-signal ratio (TSR: ratio of energy in the tail of reverberations to energy in the observation, see Holland *et al.* 2001). In Dabelsteen *et al.* (1993), the results were based on envelope functions of the blackbird signals whereas the present investigations are performed directly on the signals as a meaningful envelope function of the swift fox signal cannot be computed due to the time-frequency structure of the signal.

We used the signal match function in SIGPRO on the filtered signal waveforms of the short upsweeps and downsweeps from two trials in the sequence of synthetic sounds and the single elements of the barking sequences (Table I) to make the measurements on our observation sounds. In short, the signal match function aligns the signal waveform of an observation sound and the corresponding model sound in time using cross correlation. The result of the signal match is the mean total attenuation k for the aligned observation sound relative to the model sound. With d the distance (in meters) from microphone to loudspeaker and g the combined microphone gains (in dB) for the model and the observation sounds, the EA can be calculated as

$$\mathrm{EA} = \mathrm{g} - 20\log(d/10) - 20\log(k).$$

In the signal match function of SIGPRO provisions are made to compute the root mean square (rms) value of approximately 500 ms of the background noise in the filtered observation recording prior to the arrival of the observation sound. Using this measure and the rms value of the observation within the aligned position, the program calculates

$$\mathrm{SNR} = 20\log(\mathrm{rms(observation)/rms(noise)}).$$

Finally, the sound analyst can set a cursor marking the observed end of echoes ("the tail") that extend beyond the end of the aligned observation sound and the program computes

$$\mathrm{TSR} = 10\log(\mathrm{energy(tail)/energy(aligned\ observation)}).$$

To get a general overview of frequency dependent attenuation in this habitat, we used the long upsweeps and long downsweeps (synthesized sounds, Table I). An autocorrelation function algorithm in MATHCAD 2001 was used to calculate the energy density spectra of model and observation sounds (1024 FFT, Hann window, 22.5 ms window width; see Darden *et al.*, 2003b for details on the algorithm). The autocorrelation algorithm performs computation of the energy density spectrum of a signal and smoothes the result by averaging of adjacent frequency components to minimize the effect of stochastic noise. To quantify the attenuation, we then subtracted the energy density spectrum of the model from the energy density spectrum of each observation and adjusted the spectrum for differences in microphone gain and spherical spreading.

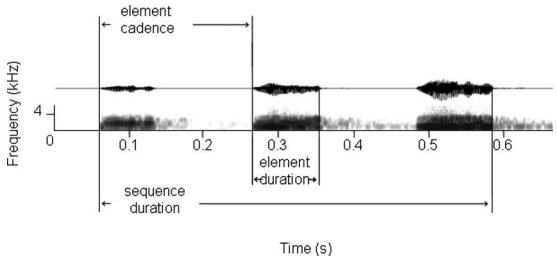


FIG. 2. Temporal measurements made on high temporal resolution spectrograms with the signal waveform pictured simultaneously (Avisoft SASLAB PRO 3.2, Hamming window, 128 point FFT, 25% frame size, 87.5% overlap; 172 Hz and 0.73 ms resolution).

2. Stability of individual vocal characteristics

We subjected our propagated swift fox vocalizations to an analysis similar to that described in Darden *et al.* (2003a). Avisoft SASLAB PRO was used for temporal measurements from spectrograms with high temporal resolution displayed simultaneously with the signal waveform (Hamming window, 128 point FFT, 25% frame size, 87.5% overlap; 172 Hz and 0.73 ms resolution) (Fig. 2): total duration of the barking sequence (DUR); mean cadence of barking sequence elements (MCAD); variance of element cadence measurements within a sequence (CADV), and the on/off duty cycle of a

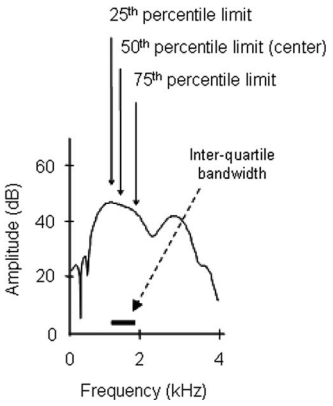


FIG. 3. Spectral measurements made from single barking sequence elements using a windowed autocorrelation function (see text). Arrows indicate the measurement results for the energy distribution in this particular example.

sequence (sum of element durations divided by total duration of the sequence) (DCYCL). We again used the autocorrelation algorithm in MATHCAD 2001 for two spectral measurements on the second to the last barking sequence elements that were transmitted singly: center frequency (CEN) and interquartile bandwidth (25%–75% of the energy density spectrum) (QBAN) (Fig. 3).

C. Statistical analysis

We tested for differences in propagation over the nine experimental distances and the model sound with repeated measures general linear models (PROC GLM with REPEATED statement) in SAS/STAT 9.1 (SAS Institute, Inc., Cary, NC,) and *post hoc* Tukey multiple comparison tests. We used a separate GLM model for each analysis to test for the effect of distance on (1) the propagation properties of the short upsweeps, (2) the propagation properties of single barking sequence elements, and (3) the individual characteristics of barking sequences. For GLM models 1 and 2, we also included the center frequency of the model sound for the corresponding observation sound and its interaction with distance as factors. To evaluate the consistency of the measured barking sequence variables during transmission, we correlated (Pearson product) the measures from the model sound, with the corresponding measures of our observation sounds within a given distance.

III. RESULTS

A. Sound propagation

1. Synthesized sounds

While it was possible to discriminate all sweeps from the background noise [20–25 dB(A) at each recording location] at transmission distances less than 120 m, we could only measure SNR and EA at all distances for short sweeps with 0.9 and 1.8 kHz center frequencies (Fig. 4). We therefore ran two separate GLM analyses, one with all the sweeps up to 120 m and another with only the sweeps centered at 0.9 and 1.8 kHz up to 400 m. Both SNR and EA were significantly affected by transmission distance ($F_{4,45}$=510.58 and $F_{4,45}$=91.82, respectively, $p<0.0001$) and the center frequency of the transmitted short sweep ($F_{7,45}$=41.01 and $F_{7,45}$=34.34, respectively, $p<0.0001$) at distances less than 120 m. We can see from the *post hoc* comparison of means that the average SNR [Fig. 5(a)] of transmitted sweeps starts to decrease significantly already at a distance of 40 m and the average EA of transmitted sweeps increases significantly at a distance of 120 m [Fig. 5(b)]. Sweeps with frequencies centered at 2.8 and 3.7 kHz had the highest SNR and those centered at 0.9 and 7.4 kHz had the lowest SNR for sounds transmitted up 120 m [Fig. 5(c)]. The pattern for EA mirrors that of the SNR [Fig. 5(d)]. The higher frequencies have the greatest variability in measurement, which is likely to be due to small fluctuations in the local climate during the transmission trial that affect these frequencies the most (e.g., Rasmussen, 1986).

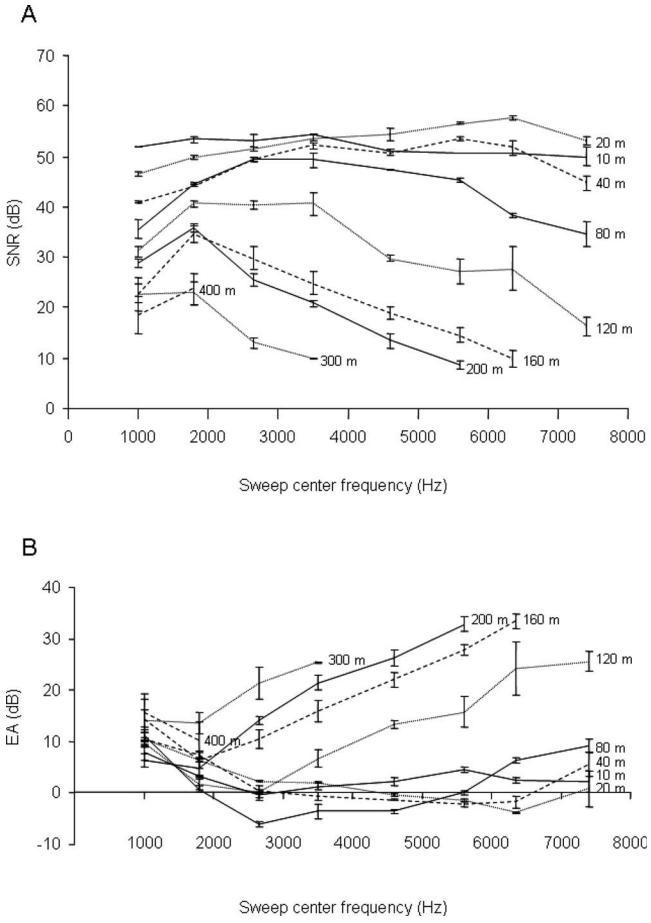


FIG. 4. Degradation measures for 16 synthetic sounds (short upsweeps and downsweeps, see Table I) transmitted up to a distance of 400 m; (A) average SNR and (B) average EA. There are missing values at some frequencies at distances where they could not be distinguished from the background noise. Bars indicate standard error.

For distances up to 400 m, there was a significant effect of distance on both SNR and EA ($F_{8,40}$=32.49, $p<0.0001$ and $F_{8,40}$=3.05, p=0.0090, respectively) (Fig. 6). The only significant difference in EA was between 200 m, where the average EA was lowest, and 300 m, where the average EA was greatest [Fig. 6(b)]. There was also a significant effect of the center frequency of the model sound such that the average SNR was higher [Fig. 6(a)] and EA lower [Fig. 6(b)] for the sweeps centered at 1.8 than at 0.9 kHz ($F_{1,40}$=18.69, p

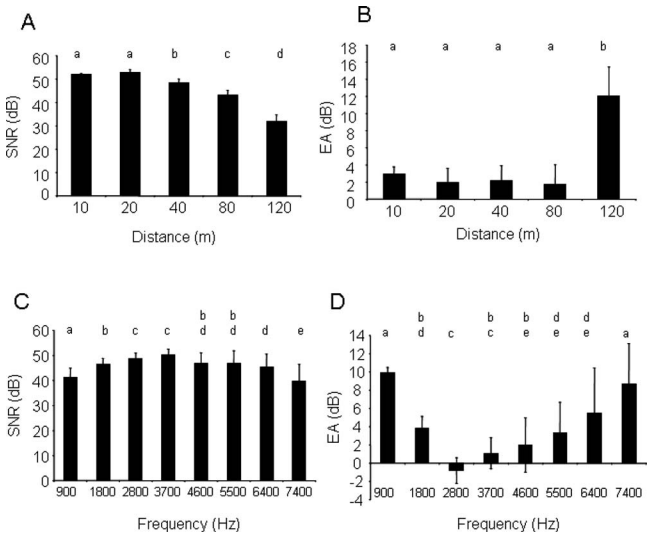


FIG. 5. Degradation measures for 16 synthesized sounds (short upsweeps and downsweeps, see Table I) for transmission distances up to 120 m; (A) average SNR at all frequencies, (B) average EA at all frequencies, (C) average SNR of sweeps transmitted 10, 20, 40, 80, and 120 m, and (D) average EA of sweeps transmitted 10, 20, 40,80, and 120 m. Columns with the same letter did not differ significantly in *post hoc* multiple comparison tests (Tukey). Bars indicate standard error.

<0.0001 and $F_{1,40}=11.89$, $p=0.0013$, respectively). In this respect, when we looked more closely at the frequency distribution over distance with the windowed autocorrelation function, we could see that on average, there was differential attenuation of frequencies below about 1.2 kHz (and above about 3.0 kHz) (Fig. 7).

2. Swift fox vocalizations

The TSR of transmitted barking sequence elements was very low (max. −24.2 dB, average -34.3 ± 5.5 dB, $n=56$),

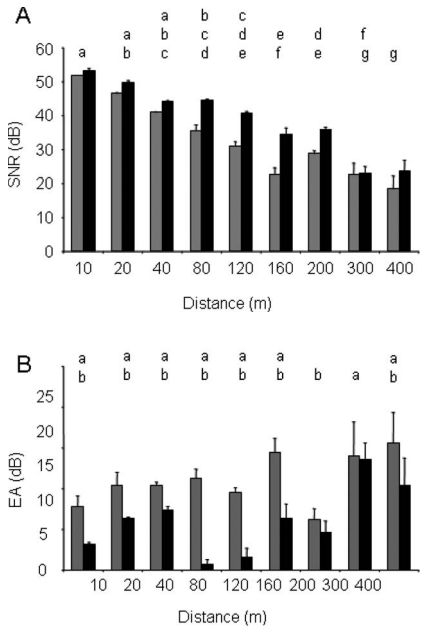


FIG. 6. Degradation measures for four synthesized sounds (short upsweeps and downsweeps) centered at 0.9 (grey) and 1.8 (black) kHz (see Table I) for transmission distances up to 400 m; (A) average SNR and (B) average EA. Columns with the same letter did not differ significantly in *post hoc* multiple comparison tests with values for all four sounds pooled (Tukey). Bars indicate standard error.

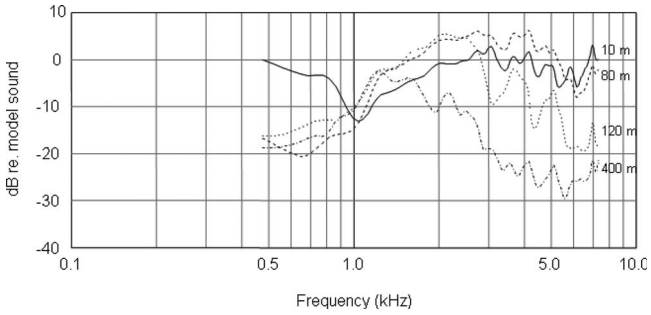


FIG. 7. Spectral representation of the relative sound pressure level between experimentally transmitted observation sounds and reference model sounds (see text) for two synthetic sounds (long upsweep and downsweep). Curves show average spectra for the two sounds at four distances.

probably due to lack of reflecting vegetation in the habitat (see also Balsby *et al.*, 2003), and we did not conduct any further analysis on this variable. The SNR of barking sequence elements was significantly affected by transmission distance ($F_{8,63}=55.87$, $p<0.0001$) and the *post hoc* Tukey analysis showed that there were four groupings of distance between which the SNR decreased significantly with increasing distance: between 20 and 40 m, 120 and 160 m, and 200 and 300 m [Fig. 8(a)]. As the exception, the SNR at 80 m was not significantly different from that at 160 and 200 m. There was also a significant effect of distance on EA of single barking sequence elements, but the pattern was much more disjunctive [Fig. 8(b)]. The lowest EA for swift fox vocalizations occurred at 120 m. Figure 9 illustrates the ef-

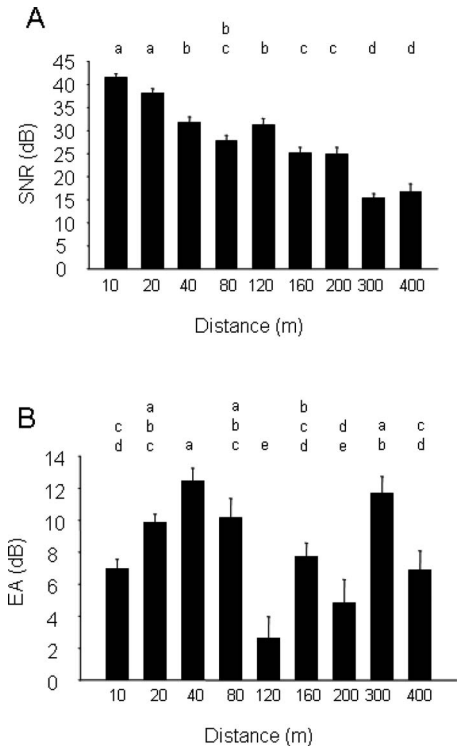


FIG. 8. Degradation measures for eight single swift fox barking sequence elements (see Table I) transmitted at distances up to 400 m; (A) average SNR and (B) average EA. Columns with the same letter did not differ significantly in *post hoc* multiple comparison tests (Tukey, $\alpha=0.05$). Bars indicate standard error.

fect of distance on single swift fox barking sequence elements.

B. Stability of individual characteristics of swift fox barking sequences

1. Effect of distance on measured variables

As expected due to the low level of reverberation in this habitat, there was only a significant effect of distance on variables describing the spectral properties of a barking sequence element and the temporal variables had the highest level of consistency (Table II). As expected from the results for the synthesized sounds, center frequencies of swift fox barks were consistently higher at each transmitted distance [Fig. 10(a)] compared to the model sounds because of the attenuation of lower frequencies. Quartile bandwidths were broader than those of the corresponding model sounds for most distances, but narrower for others [Fig. 10(b)]. From examination of the results of the Tukey comparisons, we could see that although the measured center frequency at all distances except 10 and 200 m differed significantly from the model, there were very few other comparisons that differed significantly (80 m versus 10 m and 80 m versus 200 m). This was not the case for the quartile bandwidth, as reflected in the consistency scores summarized in Table II. The range of natural variation that we find in the barking sequences produced by an individual is on average a 200 ± 22 Hz difference in the average CENF of an individual's barking sequences and the maximum and minimum center frequencies of barking sequences that the individual produces (i.e., the maximum measured CENF of barking sequences produced by an individual minus the average CENF of barking sequences produced by that individual and the average CENF of barking sequences produced by an individual minus the minimum measured CENF of barking sequences produced by that individual). The identical measure for QBAN gives an average difference of 188 ± 32 Hz. From these numbers, we can see that, with the exception of CENF at 40 and 80 m and QBAN at 300 m, the variation caused by transmission was within the natural range of variation exhibited by individual foxes [Figs. 10(a) and 10(b)].

In our original analysis of the individuality of swift fox barking sequences, there was a 98.5% correct classification of barking sequences to individual with the variables DUR, MCAD, CADV, DCYCL, CENF, and QBAN (see Darden *et al.*, 2003a). If we remove the QBAN variable from this analysis, we still get a rather high correct classification, 92.8%. If we remove both spectral measures, i.e., also CENF, the correct classification drops to 80.4%. However, these percentages are still significantly greater than that expected due to chance if all the variables were included (64.1%, $p = 0.0001$).

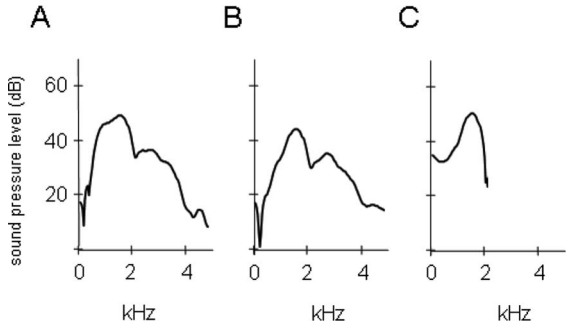


FIG. 9. Example of the spectral composition of the same swift fox barking sequence element (A) prior to transmission, (B) transmitted 80 m, and (C) transmitted 400 m. Spectra have been smoothed with a windowed autocorrelation function (see text) and the SPL has been adjusted for loss due to spherical spreading.

TABLE II. Summary results for the analysis of the effect of transmission distance on variables measured from swift fox barking sequences (see text). R=Pearson product correlation coefficient for model sounds plotted against observation sounds at each transmission distance (see text).

Variable	$F_{9,139}$	P	Mean R	SE R
DUR	0.24	0.9883	0.990	0.002
MCAD	0.04	1.0000	0.987	0.005
CADV	0.62	0.7758	0.876	0.060
DCYCL	1.12	0.3521	0.914	0.020
CENF	5.26	<0.0001	0.807	0.045
QBAN	11.29	<0.0001	0.406	0.136

IV. DISCUSSION

The natural swift fox habitat characterized in this study exhibits a range of qualities that makes barking sequences

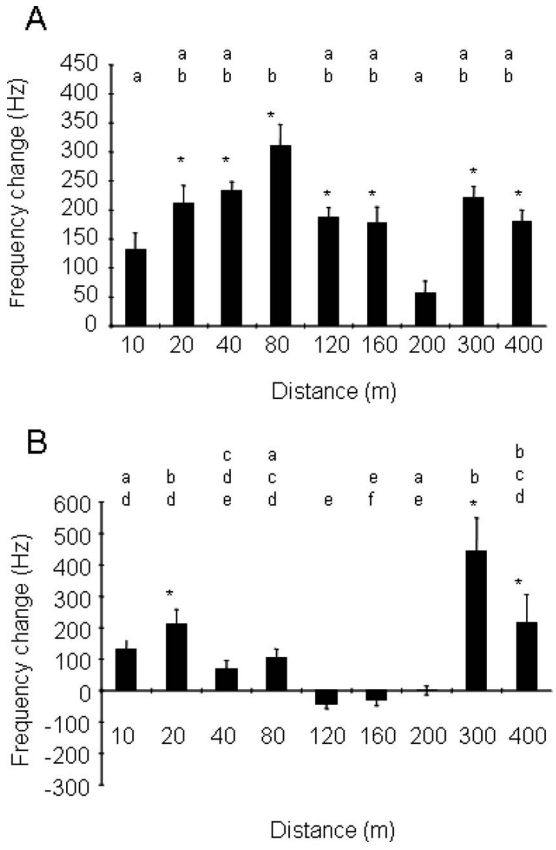


FIG. 10. Mean differences in (A) CENF and (B) (QBAN) between a swift fox barking sequence element transmitted at distances up to 400 m and the same element prior to transmission (transmitted minus untransmitted). Columns with the same letter did not differ significantly in *post hoc* multiple comparison tests (Tukey, $\alpha = 0.05$). Bars indicate standard error [$^{*}p < 0.05$ in comparison with model (untransmitted) sound].

theoretically detectible at distances over 1 km and makes it possible for foxes to communicate individually distinctive features of barking sequences over a range of at least 400 m. These ranges are of course based on propagation over flat terrain, while in most instances, the landscape is likely to experience a number of rises and falls that will influence the sound's propagation distance (Embleton, 1996). At distances less than 120 m, most frequencies were propagated at a relatively high SNR, although the lower frequencies (those centered around 0.9 kHz) exhibited higher levels of EA. At distances of 120 m and greater, frequencies above about 3 kHz were greatly attenuated and we could only measure SNR and EA for the sweeps that were in the frequency range of 0.3–1.6 kHz and 1.2–2.5 kHz at distances all the way up to 400 m. At these distances, the sweeps in the range of 1.2–2.5 kHz experienced the least amount of EA and had the highest SNR. These two patterns (low and high frequency attenuation) are consistent with the predicted transmission properties close to the ground in an open habitat where we would expect low and high frequency attenuations due to the interference from ground and reflected waves for those wavelengths, and differential attenuation of higher frequencies due to their absorbance at a faster rate. Bradbury and Vehrencamp (1998, pp. 113–139) provide a good description of these effects, but to summarize, the ground represents a change in transmission medium (air to ground) and works to reflect and absorb incident sound waves. Reflection will be frequency dependent and at the receiving end will be manifested as attenuation or amplification of specific components of the emitted sound dependent on distance traveled, ground impedance, and sender and receiver heights.

The tails of echoes added to barking sequence elements during propagation were short and of very low energy due to the paucity of reflective surfaces in this habitat. This means that forward masking (Holland *et al.*, 2001) of successive barks in a sequence is essentially nonexistent and temporal features of the barking sequences were thus very consistent. These were also the features of highest importance in our discriminant function analysis for classifying barking sequences to the correct individual. We can see that the frequencies of barking sequences, which on average have 95% of their energy in the range of 0.7–2.1 kHz (SKD unpublished data) and are on average centered around 1.2 kHz (Darden *et al.*, 2003a), are not necessarily designed for what appears to be the most optimal range for long distance propagation of all frequency components. However, from a closer look at how our measures changed with transmission distance, we could see that for the most part, shifts in the center frequency and quartile bandwidth of a barking sequence element were within the natural range of variation exhibited by individual foxes. As such, it may still be possible for receivers to extract individual identity information from the spectral components of the barking sequence at various distances. In fact, they may be quite important for increasing the reliability of individual identity cues. Without the inclusion of the two spectral measures, our percentage correct classification dropped by nearly 20%. A study by Mitchell *et al.* (2006) found that individual identity cues, mostly frequency dependent, measured from coyote *Canis latrans* howls remained intact for at distance of at least 1 km when transmitted from a height of 50 cm over flat terrain in an open landscape. Coyote howls are harmonic and have a dominant, lower frequency around 1.0 kHz from which the measurements were made. Swift fox barking sequences incorporate frequency components that are lower than this and they are produced at a lower intensity than coyote howls [91 dB SPL at 1 m (SKD unpublished data) versus 105 dB SPL at 1 m (Mitchell *et al.*, 2006), respectively], both of which would explain a lot of the difference in reliable propagation of relative spectral information in these calls. Of course, differences in habitat structure will also have an effect.

So what do we postulate the communication range of this vocalization to be and why are the foxes using calls with frequency components that are below those that are most optimally propagated over long distances? Using measurements of SPL at the source, Dabelsteen *et al.* (1993) estimated the detection range by blackbirds *Turdus merula* of blackbird song by calculating EA per doubling of distance and adding this to attenuation due to spherical spreading. Similarly, Van Staaden and Römer (1997) used regression equations to model transmission distances of male and female bladder grasshopper *Bullacris membracioides* sounds to estimate the communication ranges of these sounds. However, such calculations are not possible in this study for swift fox barks since EA is not directly related to transmission distance [Fig. 8(b)], most likely because of the strong ground effect. We have only rarely observed swift fox vocal interactions in the wild, but we have used barking sequences to conduct a number of playback trials simulating the presence of a strange male fox in the home ranges of our study animals (see Darden and Dabelsteen, 2008). From these trials, we can see that foxes respond to the playback with vocalizations (barking sequences) when they have been anywhere from 90 to 400 m of the playback speaker. If we assume that foxes are most likely to respond to a vocalizing fox in their home range after they have acquired some information about that fox, they are able to get that information at a distance of at least 400 m. Of course, foxes may also upon detecting a barking sequence call back in order to initiate an interaction that will allow them to localize the individual and approach or move into a better listening position to gather more information about that individual from its barking sequences.

In any case, the estimates we have made in this study regarding the transmission distance of individual identity information and signal range under optimal transmission conditions (flat terrain, very low background noise, very low wind) are probably conservative since we have not taken into account the hearing ability of foxes. The outer ear has a significant effect on sound reception (e.g., Carlile, 1990; Aitkin *et al.*, 1994; Young *et al.*, 1996) and sounds are most likely amplified by fox ears. Even so, the spectral structure of barking sequences would remain more intact, and the vocalization propagate farther, if foxes shifted their energy to higher frequency components. Also, foxes do not always call under optimal conditions so our estimates may represent a realistic average when considering effects incurred by time of day, wind conditions, background noise, and varying terrain.

There could be several factors contributing to the design as it is. First, although foxes have vocalizations in their repertoire that are centered at frequencies that average up to 3.4 kHz (Darden and Dabelsteen, 2006), they may simply be constrained by physiological factors in their ability to produce vocalizations at the amplitude of the barking sequence in the more optimal range. In fact, their other high amplitude vocalizations, meows, barks, and screams, are on average centered at 1.2, 0.9, and 1.4 kHz, respectively. Second, a major cost of a large detection distance may be the increased chance of attracting a predator. In this respect, the observed design may reflect a trade-off between having a sound propagate only a short distance to make it less likely to attract a predator and travel a long distance to increase the chance of the signal reaching intended receivers without the need for a high repetition rate (which may also increase predation risk). Of course, another explanation may be that foxes want to limit the distance at which individual identity and other information are transmitted to avoid, for example, the costs associated with having unintended receivers (see Dabelsteen, 2005). Foxes in our study area occupy home ranges of up to 15.8 km^2 (Darden and Dabelsteen, 2008), but they can be at least twice as large in some parts of the species' range (Moehrenschlager *et al.*, 2004). At this order of magnitude, barking sequences may function at what we would term intermediate distances to defend particular resources or areas of a home range when intruders approach or there is a high likelihood that an intruder is in the area (see Darden and Dabelsteen, 2008), but not necessarily as a broadcast signal to what might be defined as an individual's social network (mate, neighbors, etc.). As such, the possibility of foxes being part of a communication network (McGregor and Dabelsteen, 1996) on the basis of this vocalization may be quite low, since the signal would have to reliably range at least a couple of kilometers during signaling events to get well into a neighboring home range.

ACKNOWLEDGMENTS

We thank K. K. Jensen for generously composing our transmission sequence. The study was funded by the SGS Long-Term Ecological Research project, the Danish National Science Foundation, a Ph.D. Fellowship from the University of Copenhagen to S.K.D and two Framework Grants from the Danish National Research Council (No. 21-04-0403 to T.D. and No. 23155-4 to O.N.L.).

Aitkin, J. E., Nelson, J. E., and Shepherd, R. K. (**1994**). "Hearing, vocalization and the external ear of a marsupial, the Northern Quoll, *Dasyurus hallacatus*," J. Comp. Neurol. **349**, 377–388.

Balsby, T. J. S., Dabelsteen, T., and Pedersen, S. B. (**2003**). "Degradation of whitethroat vocalizations: Implications for song flight and communication network activities," Behaviour **140**, 695–719.

Blumenrath, S. H., and Dabelsteen, T. (**2004**). "Degradation of great tit (*Parus major*) song before and after foliation: Implications for vocal communication in a deciduous forest," Behaviour **141**, 935–958.

Blumenrath, S. H., Dabelsteen, T., and Pedersen, S. B. (**2004**). "Being inside nest boxes: Does it complicate the receiving conditions for great tit *Parus major* females?," Bioacoustics **14**, 209–223.

Blumstein, D. T., and Munos, O. (**2005**). "Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls," Anim. Behav. **69**, 353–361.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication*. (Sinauer Associates, Inc., Sunderland, MA).

Carlile, S. (**1990**). "The auditory periphery of the ferret II: The spectral transformations of the external ear and their implications for sound localization," J. Acoust. Soc. Am. **88**, 2196–2204.

Ceugniet, M., and Izumi, A. (**2004**). "Vocal individual discrimination in Japanese monkeys," Primates Med. **45**, 119–128.

Dabelsteen, T. (**2005**). "Public, private or anonymous? Facilitating and countering eavesdropping," *Animal communication networks*, edited by P. McGregor (Cambridge University Press, Cambridge), pp. 38–62.

Dabelsteen, T., Larsen, O. N., and Pedersen, S. B. (**1993**). "Habitat-induced degradation of sound signals: Quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song," J. Acoust. Soc. Am. **93**, 2206–2220.

Daniel, J. C., and Blumstein, D. T. (**1998**). "A test of the acoustic adaptation hypothesis in four species of marmots," Anim. Behav. **56**, 1517–1528.

Darden, S. K., and Dabelsteen, T. (**2008**). "Acoustic territorial signaling in a small socially monogamous canid," Anim. Behav. **75**, 905–912.

Darden, S. K. and Dabelsteen, T. (**2006**). "Ontogeny of swift fox *Vulpes velox* vocalizations: production, usage and response," Behaviour **143**, 659–681.

Darden, S. K., Dabelsteen, T., and Pedersen, S. B. (**2003a**). "A potential tool for swift fox (*Vulpes velox*) conservation: Individuality of long-range barking sequences," J. Mammal. **84**, 1417–1427.

Darden, S. K., Pedersen, S. B., and Dabelsteen, T. (**2003b**). "Methods of frequency analysis of complex mammalian vocalizations," Bioacoustics **13**, 247–263.

Embleton, T. F. W. (**1996**). "Tutorial on sound propagation outdoors," J. Acoust. Soc. Am. **100**, 31–48.

Frommolt, K. H., Kruchenkova, E. P., and Russig, H. (**1997**). "Individuality of territorial barking in Arctic foxes, *Alopex lagopus* (L., 1758)," Z. Säugetier. **62**, 66–70.

Holland, J., Dabelsteen, T., and Pedersen, S. B. (**1998**). "Degradation of wren *Troglodytes troglodytes* song: Implications for information transfer and ranging," J. Acoust. Soc. Am. **103**, 2154–2166.

Holland, J., Dabelsteen, T., Pedersen, S. B., and Paris, A. L. (**2001**). "Potential ranging cues contained within the energetic pauses of transmitted wren song," Bioacoustics **12**, 3–20.

Mathevon, N. (**1997**). "Individuality of contact calls in the greater flamingo *Phoenicopterus ruber* and the problem of background noise in a colony," Ibis **139**, 513–517.

Mathevon, N., Dabelsteen, T., and Blumenrath, S. H. (**2005**). "Are high perches in the blackcap Sylvia atricapilla song or listening posts? A sound transmission study," J. Acoust. Soc. Am. **117**, 442–449.

McComb, K., Reby, D., Baker, L., Moss, C., and Sayialel, S. (**2003**). "Long-distance cues to social identity in African elephants," Anim. Behav. **65**, 317–329.

McGregor, P. K., and Dabelsteen, T. (**1996**). "Communication networks," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 409–425.

Mitchell, B. R., Makagon, M. M., Jaeger, M. M., and Barrett, R. H. (**2006**) "Information content of coyote barks and howls," Bioacoustics **15**, 289–314.

Moehrenschlager, A., Cypher, B. L., Ralls, K., List, R., and Sovada, M. A. (**2004**). "Swift and kit foxes: Comparative ecology and conservation priorities of swift and kit foxes," in *Biology and Conservation of Wild Canids*, edited by D. W. Macdonald and C. Sillero-Zubiri (Oxford University Press, Oxford), pp. 185–198.

Nemeth, E., Winkler, H., and Dabelsteen, T. (**2001**). "Differential degradation of antbird songs in a Neotropical rainforest: Adaptation to perch height?," J. Acoust. Soc. Am. **110**, 3263–3274.

Nicholls, J. A., and Goldzien, A. W. (**2006**). "Habitat type and density influence vocal signal design in satin bowerbirds," J. Anim. Ecol. **75**, 549–558.

Perla, B. S., and Slobodchikoff, C. N. (**2002**). "Habitat structure and alarm call dialects in Gunnison's prairie dog (*Cynomys gunnisoni*)," Behav. Ecol. Sociobiol. **13**, 644–850.

Rasmussen, K. B. (**1981**). "Sound propagation over grass covered ground," J. Sound Vib. **78**, 247–255.

Rasmussen, K. B. (**1986**). "Outdoor sound propagation under the influence of wind and temperature gradients," J. Sound Vib. **104**, 321–335.

Reby, D., Joachim, J., Lauga, J., Lek, S., and Aulagnier, S. (**1998**). "Individuality in the groans of fallow deer (*Dama dama*) bucks," J. Zool. **245**,

79–84.
Scott-Brown, J. M., Herrero, S., and Reynolds, J. (**1987**). "Swift Fox," in *Wild Furbearer Management and Conservation in North America*, edited by M. Novak, J. A. Baker, M. E. Obbard, and B. Malloch (Ministry of Natural Resources, Ontario), pp. 433–441.
Searby, A., Jouventin, P., and Aubin, T. (**2004**). "Acoustic recognintion in macaroni penguins: An original signature system," Anim. Behav. **67**, 615–625.
Van Staaden, M. J., and Römer, H. (**1997**). "Sexual signaling in bladder grasshoppers: Tactical design for maximizing calling range," J. Exp. Biol. **200**, 2597–2608.
Wiley, R. H., and Richards, D. G. (**1982**). "Adaptations for acoustic communication in birds: Sound transmission and signal detection," in *Acoustic Communication in Birds*, edited by E. H. Kroodsma and D. E. Miller (Academic, New York), Vol. **I**, pp. 131–181.
Young, E. D., Rice, J. J., and Tong, S. C. (**1996**). "Effects of pinna position on head-related transfer functions in the cat," J. Acoust. Soc. Am. **99**, 3064–3076.

Directionality and maneuvering effects on a surface ship underwater acoustic signature

Mark V. Trevorrow[a] and Boris Vasiliev
Defence R&D Canada-Atlantic, Dartmouth, Nova Scotia, Canada B2Y 3Z7

Svein Vagle
Fisheries and Oceans Canada, Institute of Ocean Sciences, Sidney, British Columbia, Canada V8L 4B2

(Received 13 February 2008; revised 29 April 2008; accepted 8 May 2008)

This work examines underwater source spectra of a small (560 tons, 40 m length), single-screw oceanographic vessel, focusing on directionality and effects of maneuvers. The measurements utilized a set of four, self-contained buoys with GPS positioning, each recording two calibrated hydrophones with effective acoustic bandwidth from 150 Hz to 5 kHz. In straight, constant-speed runs at speeds up to 6.2 m s^{-1}, the ship source spectra showed spectral levels in reasonable agreement with reference spectra. The broadband source level was observed to increase as approximately speed to the fourth power over the range of 2.6–6.1 m s^{-1}, partially biased at low speeds by nonpropulsion machinery signals. Source directionality patterns were extracted from variations in source spectra while the ship transited past the buoy field. The observed spectral source levels exhibited a broadside maximum, with bow and stern aspect reduced by approximately 12–9 dB, respectively, independent of frequency. An empirical model is proposed assuming that spectral source levels exhibit simultaneous variations in aspect angle, speed, and turn rate. After correction for source directionality and speed during turning maneuvers, an excess of up to 18 dB in one-third octave source levels was observed.
 [DOI: 10.1121/1.2939128]

PACS number(s): 43.30.Nb, 43.30.Jx [RCG] Pages: 767–778

I. INTRODUCTION

The low-frequency underwater acoustic output generated by surface ships is a significant contributor to background ambient noise in the sea. This is especially true in littoral regions, near harbors and shipping lanes. With minimal acoustic attenuation at low frequencies (LFs), ship acoustic footprints can extend significant distances, potentially of the order tens to hundreds of kilometers for LF (<100 Hz) signals. In the civilian domain, concerns regarding acoustic noise pollution arise with the effects of shipping on marine mammals and fisheries. In the naval area, underwater radiated signatures represent a vulnerability for both surface vessels and submarines, and thus these vessels are specifically designed to minimize their acoustic signature over a broad range of frequencies. Additionally, naval vessels are often required to perform aggressive maneuvers, so it is important to understand changes in their acoustic signature during such maneuvers. Although measurement of the broadband acoustic signature from ships is straightforward in principle, there are several subtle features of ship signatures which deserve attention. Specifically, the directionality of vessel radiated signatures and the effects of turning maneuvers are examined in this work.

In this study, the underwater acoustic output from a ship will be called *signature*. As a result of a number of studies since the second World War,[1–4] the general features of a ship's signature are reasonably well known. There are two major components: narrowband propeller and machinery lines, and broadband cavitation, turbulence, and bubble mediated output. Fundamental propeller shaft and blade-rate signals lie in the 1–20 Hz range, dependent on ship and propeller design and speed, with multiple harmonics extending upwards to 100 Hz[3]. Gray and Greeley[5] presented a model for propeller blade-rate signatures, identifying propeller cavitation as the source mechanism. Other narrowband shipboard machinery (e.g., generator and compressor) signals may extend as high as 3 kHz. Broadband propeller cavitation, turbulent flow, and bubble generation sounds extend from roughly 100 Hz upwards, with a spectral level varying approximately as inverse frequency squared.[1] In some literature,[2,4] the ship acoustic signatures have been assumed horizontally omnidirectional. Other recent studies[3] present directivity measurements showing reduced levels at bow and stern aspects.

While a few deep-water acoustic ranges exist (e.g., Atlantic Underwater Test and Evaluation Center—AUTEC, utilized by Arveson and Vendittis[3]), the authors are aware of a number of fixed, shallow-water sound ranges. These shallow-water facilities can be useful for routine stationary or low-speed measurements, particularly for relative measurements, e.g., to assess changes due to differing machinery states. However, such facilities are generally too shallow and close to shore to allow ships to maneuver freely at greater speeds. Air-dropped sonobuoys have been used in some deep-water studies;[4] however, precise positioning of the ship and buoys and precise acoustic calibration remain problem-

[a]Author to whom correspondence should be addressed. Electronic mail: mark.trevorrow@drdc-rddc.gc.ca.

atic. Additionally, measurement of ship signatures at longer ranges often requires a difficult compensation for acoustic propagation effects.[2] In this work, a set of four broadband underwater recording buoys (BURBs) was developed to overcome some of these limitations. In addition to allowing the vessel to maneuver freely well away from shore, operating in deeper waters minimizes or removes seabed interactions, simplifying estimation of the low-frequency source levels.

The BURB system was developed in 2004/2005 through a collaboration between Defence R&D Canada Atlantic and the Institute of Ocean Sciences. This prototype system was intended to demonstrate the concept of a portable acoustic ranging system for the Canadian Navy. The BURBs were designed to provide continuous recording of two calibrated hydrophones, each with 20 kHz bandwidth. Relative positioning of the individual BURBs and the target ship was accomplished through commercially available differential global positioning systems (DGPSs). This work describes the first major sea trial of the BURBs against a small oceanographic vessel (CCGS VECTOR, 39.7 m LOA). While these particular sea trials uncovered some practical limitations in these prototype buoys, an examination of the data uncovered several interesting features of ship signatures. In particular, the ship's acoustic directionality and the increased source levels generated by turning maneuvers were investigated and described herein. A goal of the analysis was to devise a simple, empirical model that could account for the major dependencies in the signature during maneuvers. In particular, signature variations due to frequency, aspect angle, speed, and turn rate were investigated. Section II describes the BURBs and the experimental approach. An important discussion of surface-reflected acoustic multipath effects is presented in Sec. II E. Section III will present an overview of the acoustic results from the three days of sea trials in mid-April 2005. Finally, Sec. IV will present summary discussions and recommendations.

II. EXPERIMENT

A. Broadband underwater recording buoys

The BURBs (see photo in Fig. 1) were developed to provide broadband, short time-scale measurement of vessel underwater radiated signatures. A detailed description of the BURB system is given by Trevorrow *et al.*[6] The BURB system was intended to remedy problems of positioning, calibration, data dynamic range, lifetime, and background noise encountered with the use of conventional sonobuoys. There were four identical BURBs, each completely self-contained with an operational duration in excess of 24 h. Each buoy continuously recorded two hydrophone channels onto an internal hard drive. Each hydrophone channel was sampled at 40 000 samples/s with 16 bit (92 dB) dynamic range.

Additionally, an automatic gain control (AGC) adjusted the recording levels by up to 58 dB. The buoys recorded their own position through an onboard DGPS receiver at 1 s intervals. Each buoy was approximately 110 cm high by 60 cm maximum diameter, with a weight of 45 kg. In these trials, the two hydrophones on each BURB were suspended

FIG. 1. Photograph of a BURB during deployment from the CCGS Vector, April 2005. Inset at lower right shows the hydrophone, cable, strength member, and lead weight.

at depths of 5 and 15 m below the surface package. The hydrophones were inexpensive commercial units approximately 10 cm in length by 2 cm diameter with integral preamplifier. Note that the use of two different hydrophone depths allowed a consistency check on the propagation loss corrections, to be explained in Sec. II E. For simplicity, the hydrophones were deployed without special suspension gear (e.g., damping plates and elastic sections). The hydrophones and cables were attached to a 1 cm diameter braided rope with a 2.5 kg weight at the bottom (see inset of Fig. 1). The hydrophones were held away from the rope and cables with a u-shaped wire stiffener. It was assumed that this would be acceptable for the relatively calm conditions expected in these trials.

Detailed acoustic calibrations of the eight BURB hydrophones were conducted after the trials. The calibration technique involved broadcasting a series of cw tones of known source level at frequencies from 500 to 20 000 Hz in steps of 100 Hz, and comparing results to a reference hydrophone. The measured low-frequency hydrophone responses were near −185 dB re 1 V/μPa. Unfortunately, the hydrophones exhibited a strong resonance in the 6–10 kHz region, with multiple smaller resonances above 10 kHz. These resonances greatly complicated attempts at spectral compensation above 6 kHz. In all that follows, the acoustic data were numerically low-pass filtered with a corner frequency of 5500 Hz. While unfortunate, this reduced bandwidth covered the dominant portion of the ship's signature and was sufficient to explore the effects of directionality and maneuvers on ship source spectra.

Instrumental and background noise levels imposed some constraints on the ship signature measurements. Internal electronic noise levels had an equivalent sound pressure level (SPL) near 50 dB (re 1 μPa^2/Hz), which was more than 40 dB below the typical ship signatures encountered during the trials. More importantly, measured background

acoustic noise SPL were between 60 and 70 dB, decreasing with frequency. An approximate fit to the background noise during the sea trials was found to be $\mathrm{SPL}=118-16\log_{10}(f)$ (dB re 1 $\mu\mathrm{Pa}^2/\mathrm{Hz}$), or approximately equivalent to open-ocean wind-wave noise at Beaufort scale five (wind speeds of 8.5–11 m s^{-1}, as cited in Wenz[7]). Even taking into consideration the shallow-water conditions, which are known to exhibit higher ambient noise at a given sea state, this noise level was in excess of that expected from the local winds (5–6 m s^{-1}). This suggested the presence of the buoys and/or inadequate isolation of the hydrophones from buoy motions as the source of this background noise. Overall, this posed some limitations for higher frequency (>2 kHz) measurements at ship-buoy ranges greater than a few hundred meters. Because of this background noise, the acoustic data were only used for ship-buoy ranges <300 m and where the measured SPL exceeded the background noise curve by 6 dB. This range limit corresponded to approximately 1.5 water depths. Additionally, although the BURB data bandwidth extended as low as 10 Hz, during these trials it was found that cable vibrations and wind-wave impact against the buoy hull generated excess background noise up to approximately 100 Hz. As a result, signals below 100 Hz were ignored.

B. Source ship

These field measurements utilized a small coastal oceanographic vessel, the CCGS VECTOR. This ship is 39.7 m overall length, with maximum beam of 9.5 m, draft of 3.5 m, and displacement of 560 tons. VECTOR has a maximum speed near 6.1 m s^{-1} (11.8 knots). It has a single, three-bladed, variable pitch, 1.8 m diameter propeller driven by an 825 hp (600 kW) diesel engine. The propeller is located 3 m forward of the transom with its axis at 2.3 m depth. A 1.2 m wide by 2.5 m tall rudder is mounted immediately astern of the propeller. In terms of underwater profile, the ship has a straight keel from the bow to approximately 10 m forward of the stern, changing to an upsloping hull form reaching the water line at the transom. A solid fin skeg extends the keel line aft to the rudder post, partially enclosing the propeller. The hull shape in the middle third of the ship blocks near-horizontal forward paths from the propeller within ±28° of the bow.

For these trials, VECTOR was outfitted with two DGPS recorders, one at the bow and one on the aft deck. The aft DGPS was located 4.5 m ahead of the propeller location. The ship's position, speed, and heading over ground were recorded on each system at 1 s intervals. The use of two separate DGPS receivers allowed estimation of the instantaneous ship heading through incoherent differencing of the antenna locations. This technique was used to avoid the complication of tapping into the ship's gyrocompass.

C. Experimental location

These field trials were conducted in Saanich Inlet, near Victoria, BC (Canada). This inlet is approximately 30 km long by 5–8 km at its widest, with depths in the central inlet reaching up to 220 m, overlying a soft sandy mud bottom. For these short-range (<300 m) measurements, seabed reflections were near normal incidence, thus suffering an estimated seabed reflection loss of at least 10 dB at frequencies under consideration here (100 Hz–5 kHz). This reflection loss, in combination with the greater path length (400–500 m), suggests that seabed reflections would be roughly 15–30 dB lower than the direct paths, depending on range. Thus, seabed reflections can be ignored. The inlet is also completely sheltered against ocean swell and has only modest tidal currents. This location is only a few kilometers from the Institute of Ocean Sciences, which provided shore facilities and access to small boats. In April 2005, the surface waters of the inlet were relatively unstratified, with only a mild (0.03 s^{-1}) upward-refracting sound-speed gradient in the upper 90 m. Acoustic propagation modeling suggested that the dominant propagation effects were due to the spherical spreading and surface reflections, and that refraction effects would only be relevant at ranges beyond 1 km. During the trials, the local winds (monitored via a nearby meteorological buoy) were typically 5–6 m s^{-1}, with local wind-wave heights <0.5 m.

D. Signal processing

The fundamental measurement provided by the buoys was the SPL, converted to a frequency spectral density (dB re 1 $\mu\mathrm{Pa}^2/\mathrm{Hz}$) using the individual hydrophone calibrations and the AGC. The goal of this analysis was to determine the vessel spectral source level (SSL, dB) corrected to a standard distance of 1 m. Variations in this SSL with aspect angle, ship speed, and turning rate were investigated, thus requiring estimates at relatively short time intervals. Generally, the measured SPL was strongly dependent on the distance between the source vessel and the buoy. At these short distances, the transmission loss can be approximated by a spherical spreading law and a frequency-dependent absorption term, i.e.,

$$\mathrm{SPL}(f) = \mathrm{SSL}(f) - 20\log_{10}[r] - \alpha(f)r \ (\mathrm{dB\ re\ 1}\ \mu\mathrm{Pa}^2/\mathrm{Hz}), \quad (1)$$

where r is the slant range (m) between vessel and hydrophone and f is the frequency (Hz). The absorption term, $\alpha(f)$, is dependent on water properties (e.g., temperature and salinity)[8] and is generally small ($<3\times10^{-4}$ dB m^{-1}) at frequencies below 5 kHz. It should be noted that reflections from the ocean surface have the potential to complicate the transmission loss calculation (to be discussed in the next section).

The fundamental analysis technique was the calculation of frequency spectra at 1 s intervals. A decision was made to sacrifice frequency resolution in favor of stable, short-duration spectral estimates. The precise measurement of narrowband machinery signals was not a goal in this work. Each BURB channel was processed using 4096-pt fast Fourier transforms, demeaned and tapered using a Hanning window, yielding a 9.8 Hz frequency resolution. For each 1 s block of data (40 000 samples), a total of 17 50% overlapped raw spectra were averaged. In addition to this basic spectral analysis, the data were further averaged within *one-third oc-*

tave bands. A set of industry-standard center frequencies one-tenth decade apart were used, defined by

$$f_c = 10^{n/10}, \quad n = 1,2,3, \ldots \ . \tag{2}$$

For this analysis only six representative bands were used, with corresponding center frequencies of 160, 250, 500, 1000, 2000, and 4000 Hz. Each band was one-third octave wide, with upper and lower frequency limits bounded by multipliers of $2\hat{}(-1/6)$ and $2\hat{}(1/6)$, or 0.8909 and 1.1225 times the center frequency. Averaging into one-third octave bands had the benefits of reducing spectral variability and removing effects of narrowband lines. Spectral averaging was always done using the linear (not decibel) values.

The experimental geometry was well resolved, with estimates of ship-to-buoy horizontal range and ship speed, heading, and turn rate produced at 1 s intervals. The slant range to each hydrophone depth was then used to convert each 1 s averaged SPL to an equivalent SSL using Eq. (1). The typical measurement ranges were between 20 and 300 m. A detailed investigation of the SPL time series found that the maximum acoustic output was coincident with the closest point of approach (CPA) of the propeller. A small geometric correction was applied to the aft-DGPS position to translate the effective center of the ship to the propeller location. The ship turn rate was calculated by differentiation (with smoothing) of the ship heading versus time. An additional calculation was the source azimuthal aspect angle to each buoy. This angle was defined as zero at the bow, 90° at broadside (abeam), and 180° at stern aspect, with an assumption of port-starboard symmetry.

E. Lloyd's mirror effects

It is well known that acoustic interference effects can be created by the combination of direct and surface-reflected acoustic paths. This interference pattern, seen as a range-dependent pattern of alternating peaks and nulls in frequency-time spectrograms, is known as the *Lloyd's mirror* (LM) effect. Important features of this effect are that reflections from the ocean surface can incur only small losses and that the reflection incurs a 180° phase shift. Owing to this reflection, phase shift destructive interference is particularly important at longer ranges and LF, where the difference between direct and surface-reflected path lengths becomes small. In the limit of low-frequency or long range, the source with its surface reflection behaves as a vertically oriented dipole (see Ref. 1, Chap. 4). A LM effect is also possible for seabed reflections, modified by the fact that seabed reflections do not have a 180° phase shift.

A relatively simple theory can be used to quantify the LM effect.[9] The basic theory assumes a point source and receiver, at depths d_0 and d_1, respectively, and specular reflection from a nominally flat ocean surface. The surface-reflected path is assumed to emanate from a virtual source located a distance d_0 above the boundary. The theory also allows use of a reflection coefficient, η, which is less than 1.0 if the boundary reflection is imperfect (this is also useful in the case of seabed reflections). Using these parameters, Lloyd's mirror interference pattern (LMIP) relative to spherical spreading can be given by

$$\mathrm{LMIP} = 10 \log_{10}[1 + \eta^2 - 2\eta \cos(4\pi d_0 d_1 f/(cr))] \quad (\text{in dB}), \tag{3}$$

where c is the (average) sound speed (m s^{-1}) and r is slant range (m). This relation will exhibit a series of peaks and nulls in frequency and/or range, with amplitude controlled by η. Note that this reflection coefficient can differ from unity due to the rough surface scattering and bubble scattering and absorption losses, and thus should generally be regarded as frequency, sea-state, and grazing-angle dependent. For illustration purposes in this discussion, η will be assumed to be 0.8. This theory[9] can also account for refraction effects, which are ignored in this work. This relation predicts frequencies of the first two LM peaks at

$$f_{p1} = cr/(4d_0 d_1) \quad \text{and} \quad f_{p2} = 3f_{p1}. \tag{4}$$

For example, at a slant range of 30 m with $d_0 = 2$ m and $d_1 = 15$ m, the first two peaks are at 375 and 1175 Hz.

Note that the determination of the effective source depth for a ship is not straightforward, as in reality a ship is a complicated, distributed acoustic source. It is generally accepted that the propeller is the dominant acoustic source at lower frequencies.[5] This was confirmed for the VECTOR with the finding that maximum acoustic output coincided with the propeller CPA. However, unless special mountings are used, additional machinery noise can be coupled into the hull and radiate from regions not colocated with the propeller. Additionally, broadband acoustic contributions from breaking bow, quarter, and stern waves and turbulent hull flow will be distributed over the length and depth of the ship hull. Finally, blockage by the hull and absorption by wake bubbles will modify the far-field acoustic output near the bow and stern aspects. Overall, the question can be reduced to consideration of an *effective* source location, spatial distribution, and beam pattern which includes all these effects. Gray and Greeley[5] suggested that an effective source depth for a propeller could be taken as the ship draft minus 85% of the propeller diameter. This is based on the assumption that a given propeller would only cavitate over the top portion of its travel. For the case of VECTOR, this would correspond to a source depth of 2.0 m, which is considerably shallower than ships studied in the references. A modification was proposed by Wales and Heitmeyer,[4] where a vertically distributed acoustic source following a Gaussian distribution in depth was assumed. The corresponding LMIP was then

$$\mathrm{LMIP}_G = 10 \log_{10}\left[\int_0^D (1 + \eta^2 - 2\eta \cos(4\pi d d_1 f/(cr)))W(d)\,\delta d\right], \tag{5}$$

where

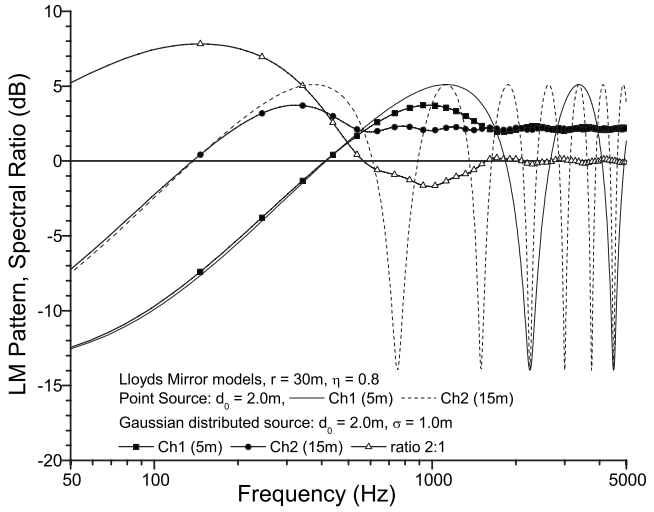


FIG. 2. Predicted LIMP. Example is calculated using point source (d_0 =2.0 m) and Gaussian distributed source (d_0=2.0 m, σ=1.0 m) models with r=30 m and η=0.8. The solid and dashed lines are for point source model with d_1=5 and 15 m, respectively. The solid square and circles denote Gaussian distributed model, with spectral ratio (2:1) denoted by open triangles.

$$W(d) = (\sigma\sqrt{2\pi})^{-1} \exp[-(d-d_0)^2/(2\sigma)]. \quad (6)$$

σ is the Gaussian source standard deviation (m), and D is the ship draft (3.5 m). A normalization constant, which is the integral of Eq. (6) from 0 to D, must be applied. Wales and Heitmeyer[4] found that this provided a better match to the observed LM pattern than using a single (point) source depth. For these trials with the VECTOR, where the ship was operated near top speed during the majority of runs, a significant fraction of the 1.8 m diameter propeller was assumed to be a cavitating acoustic source. Thus, as a starting point, a vertically distributed source with mean depth of 2.0 m and a standard deviation of 1.0 m was assumed in Eqs. (5) and (6). In principle, the source could also be distributed longitudinally, for example, by cavitation bubbles shed by the propeller or rudder. A consequence of such longitudinal source extension would be an increasing directionality at higher frequencies.

A comparison of the predicted LMIP using the two source models is plotted in Fig. 2 as a function of frequency for the two hydrophone depths (d_1=5 and 15 m). A slant range of 30 m (a typical CPA in these trials) and a constant sea-surface reflection coefficient of 0.8 were assumed for this example. The point source model shows the expected alternating peaks and nulls continuing with constant amplitude to frequencies of 5 kHz and beyond. The prediction for the deeper hydrophone shows the pattern shifted to lower frequencies. The Gaussian source case shows similar behavior at frequencies up to the first peak, followed by an increasing attenuation of the higher frequency peaks and nulls. For the Gaussian source case, the frequencies of the first peaks (908 and 303 Hz at 5 and 15 m, respectively) are slightly lower than predicted by Eq. (4), and the peak LMIP was approximately 75% of the point source model peak levels. At frequencies well above the first peak, the distributed source model trends toward a constant LMIP of 2.1 dB for both hydrophone depths. This in agreement with the theoretical asymptote of $10\log_{10}[1+\eta^2]$. At frequencies below the first peak (both point and distributed source models), the LMIP decreases monotonically at approximately 5 dB/octave. In this low-frequency limit, the effective transmission loss approaches twice that due to spherical spreading. It can be shown that for well-behaved source distributions (i.e., symmetric about d_0 and diminishing to zero for large $|d-d_0|$), the first peak in LMIP will always be present, at a frequency very similar to the point source case. The behavior between the first peak and high-frequency limit depends on the specific source distribution. Finally, note that this pattern should scale directly proportional to ship-to-buoy range, such that at ranges >200 m the peaks should be pushed to frequencies greater than a few kilohertz, making the low-frequency roll highly important.

An advantage of measuring the ship signature at two different depths is that the spectral ratio in SSL (uncorrected for LM effect) between the two hydrophones can be used to isolate the frequency dependence of the LM effect independent of the shape of the ship source spectrum, i.e., that $\mathrm{SSL}_{15\ \mathrm{m}}-\mathrm{SSL}_{5\ \mathrm{m}}=\mathrm{LMIP}_{15\ \mathrm{m}}-\mathrm{LMIP}_{5\ \mathrm{m}}$. With a single hydrophone, the LM effects are not easily separable from structures in the ship SSL. Specifically, the distributed source depth, standard deviation, and the frequency dependence in η are free parameters which can estimated through fitting Eqs. (5) and (6) to this ratio. The predicted ratio in LMIP between the two channels is plotted in Fig. 2. The distributed source model prediction is that the ratio will be up to +7 dB below 400 Hz, dipping to a −2 dB null near 1000 Hz, and finally trending to +0 dB at frequencies above approximately 1500 Hz. For the same source depth, the point source model would yield a recurring series of strong peaks and nulls extending to high frequencies. This sensitivity to source depth distribution illustrates the potential use of this spectral ratio technique in estimating the effective source distribution. This will be examined using sea-trial data in the next section.

III. ACOUSTIC MEASUREMENTS

A goal of this analysis is to quantitatively determine SSL dependencies on frequency, speed, aspect angle, and turn rate. The intended outcome is a simple empirical model that can be used to predict ship signature during maneuvers. In this work, it was assumed that the variation in SSL can be decomposed into independent variations with ship speed (U, m s^{-1}), aspect angle (θ, deg), and turn rate (TR, deg/s), i.e., as

$$\mathrm{SSL}(f,\theta,U,\mathrm{TR}) = \mathrm{SSL}_B(f) + S_U(U) + S_{\mathrm{AA}}(\theta) + S_{\mathrm{TR}}(\mathrm{TR}) \quad (\text{in dB}), \quad (7)$$

where SSL_B is the broadside SSL at a reference speed, and S_U, S_{AA}, and S_{TR} are the normalized variations due to speed, aspect angle, and turn rate, respectively. The broadside SSL and variations due to speed and aspect angle will be extracted from the straight, constant-speed runs. Analysis of the turning maneuvers will attempt to isolate the variation due to turn rate. Note that Eq. (7) might also include contributions from ship acceleration. However, no specific tests were con-

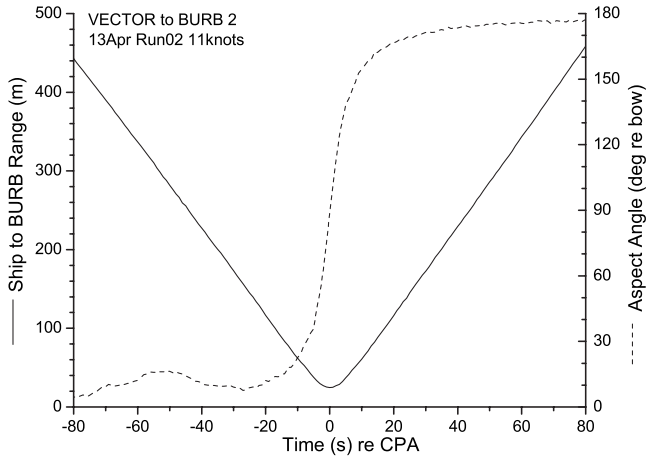


FIG. 3. Variation in ship-to-BURB range and aspect angle during a straight-line run at 5.7 m s^{-1} (11.0 knots). CPA was 24.8 m.

ducted in this trial to measure this effect, so it will be ignored herein.

A. Examination of Lloyd's mirror effects

Prior to examining specific results, it was important to understand and possibly compensate for the LM effects. An example typical of the straight, constant-speed runs will be examined in detail. In these runs, the four buoys were placed in a line and the ship directed to steam at a constant speed along a parallel course approximately 30–50 m away from the line of buoys. Figure 3 shows an example of the variation in ship-buoy geometry as the VECTOR conducted a straight, constant-speed run at 11.0 knots (5.7 m s^{-1}). In this example, the ship passed through a CPA of 28.4 m and was within 300 m of the buoy for a period of about 106 s. As it transited past the buoy, the ship exhibited a hyperbolic range-time trajectory, i.e.,

$$r^2 = r_0^2 + (U(t - t_0))^2, \tag{8}$$

where r_0 and t_0 are the range (m) and time (s) of CPA. The aspect angle varied smoothly from approximately 5° to 175° during the transit, with a rapid transition through broadside aspect within ±10 s of CPA. In this geometry, the sea-surface grazing angles were relatively small, increasing from approximately 1.3° at 300 m range (shallow hydrophone) to a maximum of 30° (deeper hydrophone) at the 30 m CPA.

Figure 4 shows the SSL spectrograms from the two channels on one of the BURBs during this straight, constant-speed transit. These spectra were corrected as in Eq. (1) but not corrected for the LM effect described in Sec. II E. Lines indicating the expected first LM peak frequencies [Eq. (3)] are overplotted. Both spectrograms show a relatively smooth decrease in radiated signature with increasing frequency, with several clusters of narrow-band lines (300–800, 1100–1600, and near 2320 Hz) due to ship's machinery. Although there was a clear SSL maximum at CPA, suggesting a broadside directional peak, a clear LM pattern was not observed. If LM effects were present, the SSL spectra in both channels should exhibit a hyperbolic shaped peak feature (similar to the dashed lines), centered at CPA, with a strong decrease in SSL below the first peak. This low-frequency roll-off should be particularly obvious at times >±10 s from CPA, corresponding to ranges >100 m. Also, the shallower hydrophone (channel 1) should show a stronger low-frequency decrease than the deeper hydrophone. These features were not observed. This lack of LM structure is believed to be due to the

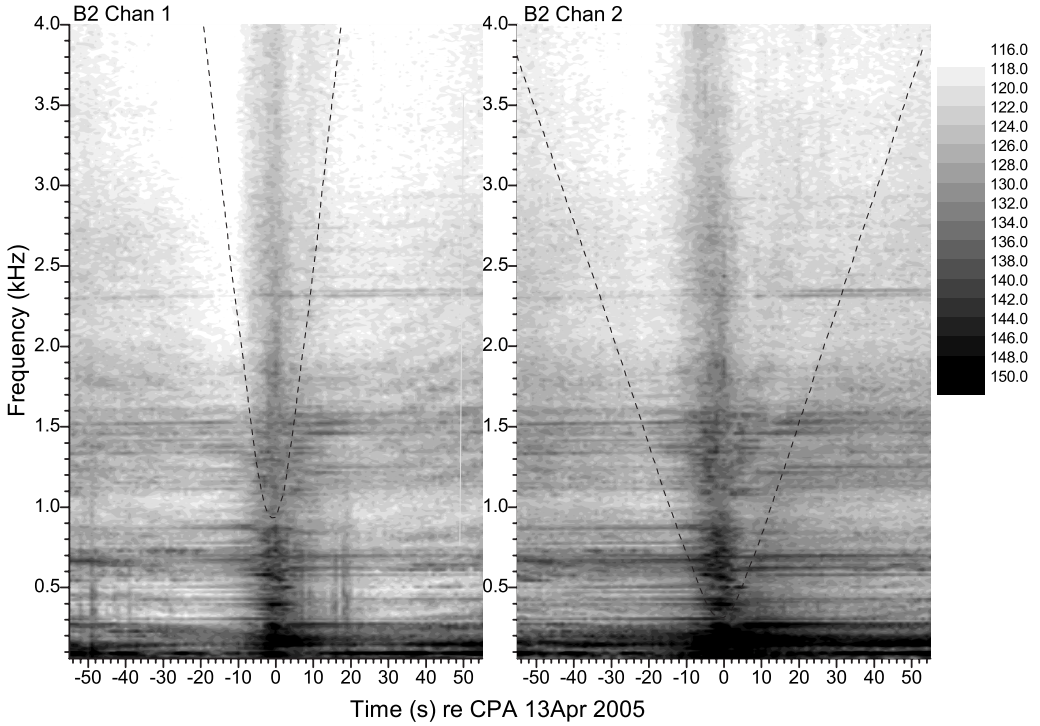


FIG. 4. Example frequency-time SSL spectrograms for the two channels (left 5 m depth, right 15 m depth) of a BURB during a straight, constant-speed run at 5.7 m s^{-1} (11 knots). Intensity in dB re 1 $\mu Pa^2/Hz$ at 1 m. Spectra have not been corrected for LM propagation effects. Frequencies of the first LM peak [Eq. (4)] in each channel are overplotted.

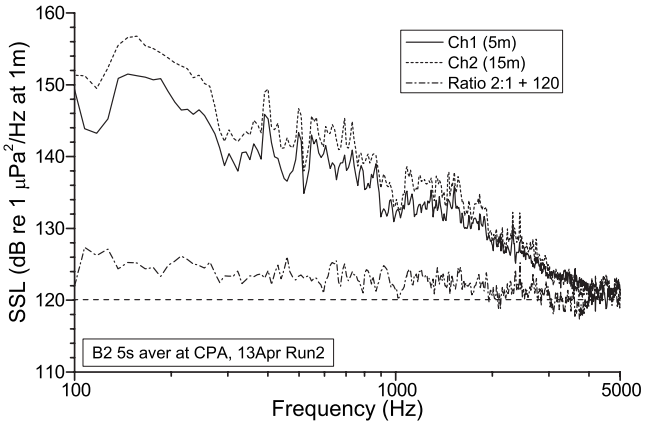


FIG. 5. Example SSL spectra at two depths near CPA from one BURB during a straight, constant-speed run at 5.7 m s^{-1} (11 knots). Spectra are averaged over 5 s, but uncorrected for LM effects. Spectral ratio (Chan 2:1) +120 dB is plotted as dash-dot line.

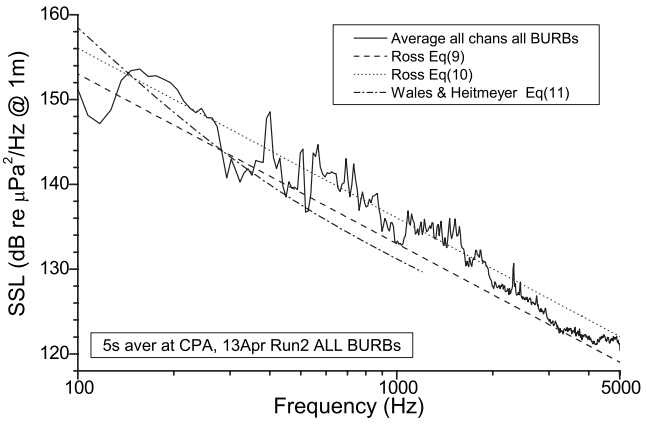


FIG. 6. Comparison of averaged SSL spectrum at broadside incidence from a straight, constant-speed run at 5.7 m s^{-1} with regressions due to Ross (Ref. 1) Eqs. (9) and (10) and Wales and Heitmeyer (Ref. 4) Eq. (11). Measured spectrum was averaged over 5 s at CPA, then averaged over the two channels from all four BURBs.

shallow source depth (~2.0 m) combined with blocking of the surface reflection from the ship's hull at forward aspect and acoustic scattering and absorption by the wake near stern aspect.

A further check for the presence of the LM effect can be made by detailed comparison of the ship's SSL at the two hydrophone depths near CPA (Fig. 5), corresponding to broadside incidence. At CPA, the LM effects should be pronounced due to larger grazing angles and minimal hull blockage. The figure shows a slightly stronger SSL in the deeper hydrophone at frequencies below 2 kHz, with a maximum spectral ratio near +6 dB at 100 Hz, in rough agreement with the prediction. The predicted asymptotic decrease in spectral ratio to near unity above 2 kHz (shown in Fig. 2) is also present; however, the prominent null near 1 kHz is absent. Even if the actual source depth distribution differed from the assumed values, there should still be a frequency-varying spectral ratio and a prominent null in this ratio in the vicinity of 1 kHz. Furthermore, below roughly 500 Hz (particularly in channel 1) the observed uncorrected SSL should be approximately flat, with the 5 dB/octave decrease due to the LMIP canceling out the roughly f^{-2} behavior (6 dB/octave increase) in the true SSL (see Sec. III B) at LF. This was not observed. It is possible that measurements below 200 Hz were contaminated by background noise, particularly in channel 1 which has a lower SSL. In the absence of a clear LM pattern, and without any confirmation of the exact form of the low-frequency LMIP, it would seem imprudent to apply any corrections in this case. Averaging the SSL estimates between hydrophone channels at two depths, and between buoys at different locations, should minimize these LM effects, with the caveat that resulting estimates below 200 Hz may be underestimated by up to approximately 6 dB.

B. Source level estimates and directionality

A series of straight, constant-speed runs was performed to establish a base line SSL for the VECTOR as a function of speed. Additionally, the variation in estimated SSL as the ship transited past the BURB field was used to infer the source directionality. The measured SSL near broadside incidence were in reasonable agreement with empirical reference spectra, as shown in Fig. 6. In this figure, the estimated SSL spectrum was averaged over 5 s at CPA, then over the two hydrophone channels in each buoy, and then over similar CPA at each of the four buoys. The variability in this estimate between the buoys decreased from ±2 dB at 100 Hz to ±1 dB at 500 Hz and above. The reference spectrum due to Ross[1] is

$$\mathrm{SSL}(f) = 190 + 53\log_{10}(U/5.15) - 20\log_{10}(f) \quad (\mathrm{dB\ re\ 1\ }\mu\mathrm{Pa}^2/\mathrm{Hz}), \tag{9}$$

where U is the ship speed (m s^{-1}). Ross also proposed an alternate reference curve based on propeller characteristics,

$$\mathrm{SSL}(f) = 195 + 60\log_{10}(U_t/25) + 10\log_{10}(B/4) - 20\log_{10}(f), \tag{10}$$

where U_t is the propeller tip speed (m s^{-1}) and B is the number of blades. For the VECTOR during this run, the propeller was operated at 290 rpm, equivalent to U_t=27.3 m s^{-1}. A more recent study by Wales and Heitmeyer[4] generated a different relation, without a speed dependence, fit to an ensemble of measurements on roughly 50 merchant ships over a 30–1200 Hz band, i.e.,

$$\mathrm{SSL}(f) = 230 - 35.9\log_{10}(f) + 9.17\log_{10}(1 + (f/340)^2). \tag{11}$$

The typical ship speeds in the Wales and Heitmeyer study were between 10 and 15 knots. All of these empirical references represent averages among an ensemble of different merchant and naval ships, with typical variability of ±6 dB about the mean, thus the signature of particular ship may not be an exact match. Overall, there is a reasonable agreement in level and frequency dependence between the measured and reference spectra at frequencies above 150 Hz. This suggests that VECTOR can be considered a *typical* ship from an acoustic signature perspective, in spite of its relatively small size. The VECTOR signature contained a number of narrowband machinery lines approximately 2–5 dB higher than the

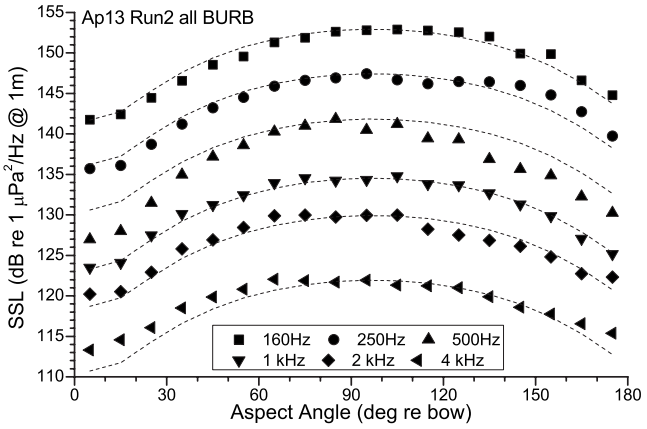


FIG. 7. Variation in one-third octave averaged SSL at six frequencies vs aspect angle from a straight, constant-speed run at 5.7 m s^{-1} (11 knots). SSL was averaged within 10° bins over the two channels from all four BURBs. The dashed lines are Eq. (12) scaled to match broadside aspect values at each frequency.

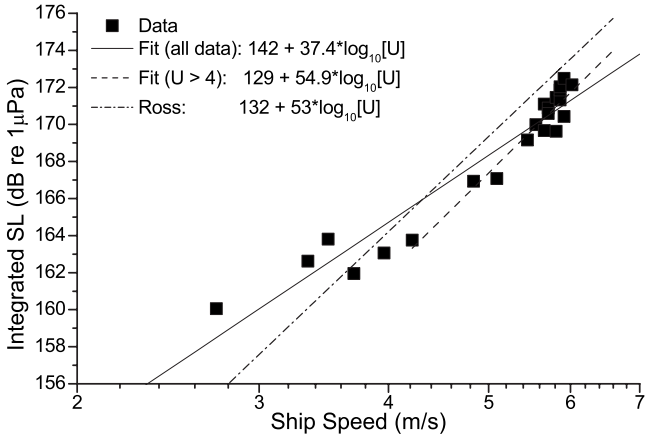


FIG. 8. Variation in BSL (dB re 1 μPa) vs speed (m s^{-1}) from all straight, constant-speed runs conducted on April 12 and 13, 2005. Source levels were calculated from 5 s averaged SSL taken at CPA (broadside aspect) to each of three buoys. The solid line is the least-squares linear regression to the entire data set. The dashed line is fit to data where speed >4 m s^{-1}. The dash-dot line is the equivalent speed dependence from Ross (Ref. 1).

references. There were particularly prominent lines at 400, 500, 550–870, 1060–1570, and 2320 Hz. Either this measured SSL, or a fit of the data to $\log_{10}(f)$, can be utilized as an estimate of $\mathrm{SSL}_B(f)$ in Eq. (7).

By averaging the SSL spectra over the two hydrophones on each BURB and then into one-third octave frequency bins, stable estimates of the source directionality were produced (Fig. 7). This plot was generated by averaging within 10° angular bins all available SSL versus aspect angle estimates, taken at 1 s intervals, from all four BURBs. Averaging within 10° angular bins compensated for the large number of estimates near bow and stern aspects, but only a few near broadside aspect. The figure shows that VECTOR exhibited a clear broadside peak approximately 8–12 dB higher than at bow incidence. This directionality appeared to be marginally stronger at lower frequencies. There was also a clear bow-stern asymmetry, particularly at lower frequencies, with the stern aspect approximately 2–4 dB higher than at bow aspect. There was only a small variation in SSL within 20° of the bow across all frequencies. This bow-stern asymmetry was presumed to be due to acoustic blocking by the hull in the forward direction.

A simple, frequency-independent model for this aspect angle variation was created. This was performed by normalizing the variation at each frequency and then averaging the resultant directionality patterns across frequencies. The resulting normalized variation with aspect angle, denoted S_{AA}, was found to be in close agreement with

$$S_{\mathrm{AA}} = 10\log_{10}[\cos^{1.95}(\theta - \theta_0) + 0.08] \quad (\mathrm{dB}), \tag{12}$$

where $\theta_0 = 97°$ is the peak direction. The small constant offset was added to handle small values near bow aspect. This curve, suitably denormalized to match the broadside aspect SSL, is plotted in Fig. 7, and will be used to correct for aspect angle effects in subsequent sections. The data-model comparison highlights the slightly greater directivity at the three lower frequencies. This angular variation is close to the expected angular variation for a horizontally oriented dipole (i.e., cosine squared), with the main lobe axis oriented 7° aft of beam (and assumed symmetric port-starboard). The data do not show any significant increase in directivity with frequency, as might be expected from a horizontally distributed source.

C. Speed dependence

A total of 12 straight runs at nominal speeds from 2.6 to 6.2 m s^{-1} were conducted during the first two days of operations, allowing examination of the speed dependence in SSL. Here, we were interested only in the variation in broadband source level (BSL) with speed, integrated from 120 Hz to 5 kHz. This was computed from 5 s averaged SSL at CPA from each buoy. In practice, there were always small variations in speed, roughly ± 0.25 m s^{-1}, during each run as the ship attempted to maintain the requested speed while lining up the buoys for a 30–50 m CPA. Using the measured speed at CPA to each buoy, a number of independent BSL versus speed estimates were created from each run (see Fig. 8). Although there was some scatter in the measurements, there is a clear trend of increasing BSL versus speed. A linear regression to these data yielded the relation

$$\mathrm{BSL} = 142.0 + 37.4\log_{10}(U). \tag{13}$$

The regression generated a correlation coefficient of 0.97 from a total of 22 data points. Note that this results suggests a speed dependence significantly weaker than the $U^{5.3}$ specified by Ross [in Eq. (9)]. A possible explanation for this difference is that these measurements spanned a lower range of speeds compared to Ross' regressions, which were based on speeds from 8 to 24 knots (4.1–12 m s^{-1}). At the lower speeds, the acoustic signature of the VECTOR likely included increasing contributions from nonpropulsion related machinery (e.g., generators and ventilation). If we restrict attention to BSL estimates at speeds greater than 4.0 m s^{-1}, the best fit line has slope of 54.9 (see Fig. 8), which is in agreement with Ross' result. When compensating for speed variations between 4 and 6 m s^{-1} during turning maneuvers (described below), this latter speed dependence seems appropriate.

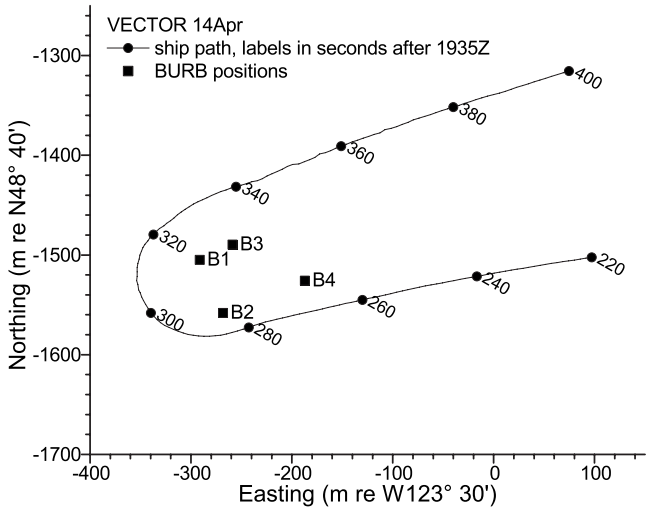


FIG. 9. Plan view of trajectory of CCGS Vector relative to the four BURBs during a 180° turning maneuver on April 14, 2005. Inbound ship speed was 5.8 m s^{-1} (11.2 knots). Labels are time (s) after start of run at 1935Z.

D. Turning maneuvers

During the turning maneuvers, the VECTOR exhibited significant increases in acoustic signature relative to the straight runs. This was true of both 90° and 180° turns. An important feature of these maneuvers was a simultaneous variation in ship speed and turning rate through the evolution of the turn. This reduction in speed was a consequence of the turning maneuver and not a result of a deliberate reduction in engine power or rpm. An additional complication was that, because the buoys occupied slightly different locations, they each were presented with different aspect angles during the turn. Figure 9 shows a plot of the ship track relative to the buoys through a 180° turning maneuver, and Fig. 10 shows the corresponding speed and turning rate. In this maneuver, the inbound speed was 5.8 m s^{-1} (11.2 knots), executing a 180° turn to starboard of radius 68 m in roughly 55 s. In this example, the turn was initiated at 275 s. After a brief transition, there was a period of relatively constant turn rate, near 4 deg/s, between 295 and 315 s. Note that there were two significant events in this maneuver: (1) the turn initiation extending over the first 15 s, and (2) a point of minimum

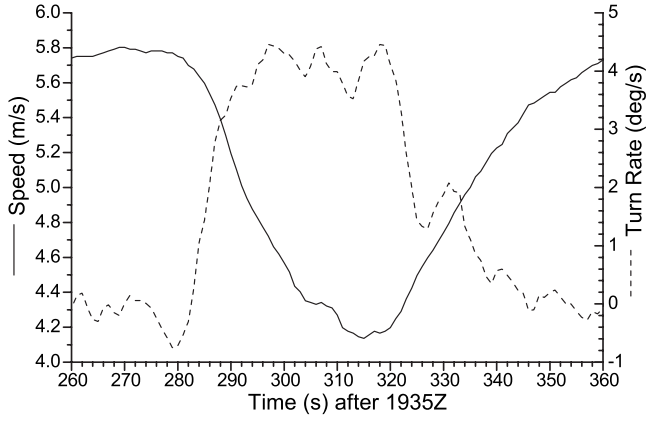


FIG. 10. Variation in ship speed (m s^{-1}) and turn rate (deg/s) during a 180° turning maneuver on April 14, 2005 (as shown in Fig. 9).

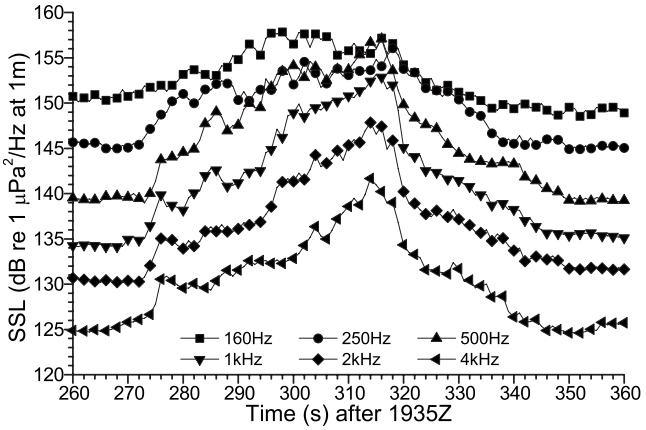


FIG. 11. Time series of corrected one-third octave averaged SSL at six frequencies from the same 180° turning maneuver, as shown in Figs. 9 and 10. SSL data have been corrected for aspect angle and speed dependence as described in the text, then averaged among three buoys.

speed, approximately 50 s into the turn. These events have acoustic significance (discussed below). Through the course of the turn, the ship presented nearly the full range of aspect angles to the individual BURBs.

The first step in this analysis was to correct for the simple variations due to speed and aspect angle [Eqs. (12) and (13)], applied to the one-third octave averaged SSL values. Since the speed was always above 4 m s^{-1}, the speed correction was taken as $S_U = 54.9 \log_{10}[U/U_0]$, where U_0 is the inbound speed. After the aspect angle and speed correction, the time series of one-third octave SSL estimated from the individual BURB data were quite similar, and so were averaged together. Figure 11 shows the corrected one-third octave averaged time series, equivalent to $SSL_B + S_{TR}$ at a speed of 5.8 m s^{-1}, through this turning maneuver. At the initiation of the turn (275 s), there was a relatively rapid rise in SSL of between 4 and 10 dB, particularly at the higher frequencies, as the ship sets its rudder and began to heel into the turn. At this point, the ship was still near its in-run speed. After the turn initiation, the behavior at lower (160 and 250 Hz) and higher frequencies began to diverge. At LF, the corrected SSL exhibited a broad plateau up to 8 dB above the in-run condition, dropping away only after the minimum speed point (315 s). The higher frequencies exhibited a slow ramp-up during the turn ending in a pronounced peak coincident with the minimum speed point. For example, at 4 kHz the SSL rose from a base line near 125 dB to a peak of 141 dB. At the minimum speed point, the ship speed was 4.2 m s^{-1} (8.1 knots), or 72% of its inbound speed. At this point, the commanded propeller rpm (still set for 5.8 s^{-1}) was furthest from constant-speed equilibrium. This can be seen through changes in the propeller advance ratio,

$$J = U/(nD_p), \tag{14}$$

where n is the propeller rotational speed (rev/s^{-1}) and D_p is the propeller diameter (1.8 m). At the in-run speed of 5.8 m s^{-1}, the propeller was at 285 rpm, thus operating at $J = 0.674$. At the minimum speed point, the advance ratio was reduced to 0.488. In general, as J decreases, the propeller thrust and torque increase and the propulsive efficiency de-

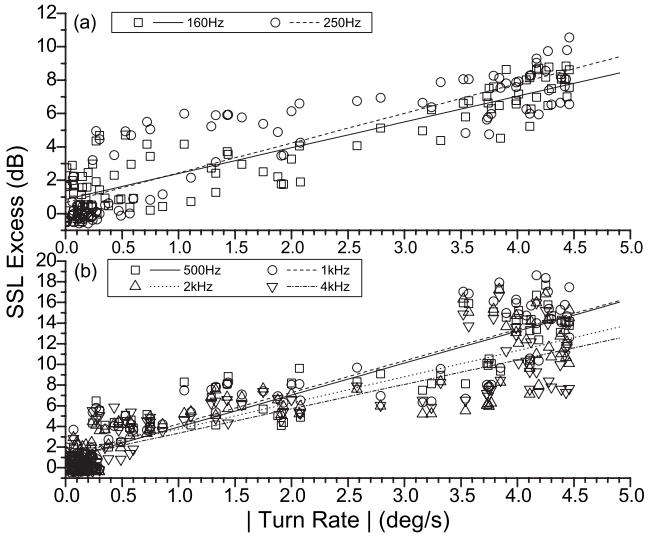


FIG. 12. Variation of SSLE with ship turn rate, from the same 180° turn maneuver shown in Figs. 9–11. (a) One-third octave variations at 160 and 250 Hz. (b) One-third octave variations at 0.5, 1.0, 2.0, and 4.0 kHz. Lines are least-squares linear regressions to data.

creases (see Ref. 1 Sec. 8.3). Increasing the propeller thrust and torque presumably created the observed increase in radiated signature levels. It is also possible that strong application of the rudder contributed to the initial signature peak (near 276 s).

The dependence of SSL excess (SSLE=S_{TR}) on turn rate was complicated. Figure 12 shows the variation of SSLE with turn rate explicitly, calculated from data shown in Fig. 11 through normalizing by SSL_B calculated from the in-run. The figure separates the dependency at the two lowest frequencies (160 and 250 Hz) from the higher frequency behavior. Figure 12(a) shows an approximately linear dependence of SSLE on turn rate, with slopes near 1.65. The maximum low-frequency SSLE was near 10 dB. In particular, for the relatively constant turn rate portion (TR$>$3 deg/s), the variance was small. However, at 500 Hz and above [Fig. 12(b)], the dependence on turn rate was more complicated. The best fit lines have higher slopes (3.1, 3.1, 2.5, and 2.4 at 500 Hz, 1 kHz, 2 kHz, and 4 kHz, respectively), with significantly greater variance at higher turn rates. In particular, there appeared to be a region of higher SSLE, up to 18 dB at turn rates above 3.5 deg/s, which was not present at the lower frequencies. These higher SSLE values occurred near the minimum speed point, which is suggestive that there may be an additional dependence on the propeller advance ratio. There were also some anomalously low SSLE values at TR $>$3 deg/s in the 4 kHz data, which occurred early in the turn.

IV. SUMMARY AND RECOMMENDATIONS

A. Summary discussions

This work describes an investigation of surface ship underwater acoustic signatures with particular examination of directionality and the effects of maneuvers. The prototype instrumentation, while suffering from some practical limitations, successfully delivered broadband acoustic measurement with a well-resolved experimental geometry. Overall, this supported the establishment of empirical models for the aspect angle, speed, and turning rate variations. The use of self-contained buoys was successful from a ship-handling perspective, allowing the ship to freely maneuver close to the buoys in relatively deep water. The target ship could quickly deploy and recover these buoys, making this approach convenient for self-ranging.

Broadband measurements from approximately 150 Hz to 5 kHz were demonstrated herein. The lower bound was set by LF noise contamination due to wave impacts against the buoy hull and inadequate hydrophone suspension. The upper frequency limit was imposed by hydrophone response problems. Both of these limitations can be overcome. Improvements in LF performance can be made by deploying the hydrophones at greater depths below the surface buoys (e.g., $>$20 m) and adding a compliant, damped suspension similar to that utilized by sonobuoys. The greater receiver depth also pushes the LM patterns to lower frequencies (at a given range), allowing measurement at greater ranges. A redesign of the buoys themselves, minimizing the surface-piercing components, would reduce LF noise due to waves. High-frequency performance can be improved through the use of better quality hydrophones.

An examination was made of the potential effects of LM interference patterns, with the conclusion that *for this particular ship* there was no convincing evidence for these effects. A technique to account for the LM propagation effect was presented, including use of a vertically distributed source model and examination of the spectral ratio between two receiver depths. In the present measurements with the VECTOR, the expected LM pattern of peaks and nulls in frequency-time spectrograms was absent. Use of a vertically distributed source model can explain the lack of higher-order interference peaks and nulls; however, the first peak should be present. Furthermore, at frequencies below the first peak the predicted strong roll-off in uncorrected SSL estimates, expected to be particularly prominent in the shallow hydrophone at ranges beyond 100 m, was not observed. The reason for this lack of LM structure was believed to be blockage of the surface reflected path due to shadowing by the ships hull near bow aspect and acoustic absorption by the wake near stern aspect. This absence of a LM interference pattern should not be taken as universal, as it has been seen in other cases (e.g., Wales and Heitmeyer[4]) with larger vessels. The final approach taken was to not apply any LM corrections, and simply average SSL estimates between receiver depths and from buoys at different locations. The caveat in this approach being that SSL estimates below approximately 200 Hz may be underestimated by up to 6 dB.

In straight-line, constant-speed runs, the ship acoustic signature at broadside incidence was found to agree closely with well-established empirical relations. Specifically, above 150 Hz, the SSL estimates exhibited close to an inverse-frequency-squared dependence, in general agreement with models proposed by Ross[1] and Wales and Heitmeyer.[4] An examination of the broadband SL versus ship speed, over the range of 2.6–6.2 m s^{-1}, showed a dependence approximately as speed to the power of 3.8. This is significantly lower than

the 5.3 power dependence proposed by Ross, which was generated for an ensemble of much larger ships over a greater range of speeds ($4-12$ m s^{-1}). However, these present measurements included lower speed measurements than the Ross relations, which presumably included acoustic contributions from nonpropulsion machinery such as generators. A regression of broadband SL versus ship speed for data above 4 m s^{-1} was in complete agreement with the Ross result. In this context, a measurement of the ship SSL while stationary would be a useful supplement.

The VECTOR signature exhibited significant horizontal directionality. Relative to the broadside maximum, the bow aspect SSLs were reduced by 10–12 dB depending on frequency. Averaged over frequency, the aspect angle dependence was in approximate agreement with a cosine-squared model, with the main lobe directed 7° aft of beam (with assumed port-starboard symmetry). The lack of increasing directivity at higher frequencies suggested that there was no along-ship extent to this acoustic source. This directionality is hypothesized as due to a combination of true source (propeller) directionality, shadowing by the ship hull, and wake absorption effects. Arveson and Vendittis[3] showed similar directionality in near-horizontal aspect angle measurements on a larger cargo ship. Specifically, their measurements showed a 16 dB variation between bow and beam aspects at 350 Hz, with a roughly 6 dB bow-stern asymmetry. However, their measurements also showed more omnidirectional behavior in very LF (<30 Hz) narrowband propeller and machinery tones.

During turning maneuvers, the VECTOR exhibited signatures up to 20 dB above the beam aspect values from straight runs at the same speed. An important feature of these turning maneuvers was a simultaneous variation in aspect angle, speed, and turning rate. By correcting for the previously established variations in aspect angle and speed, a SSLE relative to broadside was estimated. This SSLE was mostly attributable to turn rate variations. At the initiation of the turn, there was a relatively rapid rise in SSLE of between 4 and 10 dB as the ship set its rudder and began to heel into the turn. The maximum SSLE occurred during a second peak, approximately 50 s after the start of the turn, coincident with point of minimum speed. When examined in terms of turn rate dependence, while there was a clear trend of increasing SSLE with TR, the behavior was complicated. At 160 and 250 Hz, there was a simple linear dependence on turn rate with maximum SSLE up to 10 dB. At 500 Hz and above, the SSLE showed additional signature output, up to 18 dB, near the point of minimum speed. This additional signature output was hypothesized as due to the fact that the commanded propeller rmp was set for a speed significantly above the actual ship speed. This is similar to an acceleration transient where the propeller advance ratio would be significantly below optimal. This present data set, in particular, with its lack of specific data on acceleration effects and lack of instantaneous machinery data (e.g., propeller rpm and torque), was deemed insufficient to fully separate variations due to the acceleration and turning rate.

It is unclear as to whether results from the VECTOR would be more generally applicable to large merchant and naval ships. During straight, constant-speed runs, VECTOR exhibited broadside SSL and speed dependencies in agreement with reference curves, generated from ensembles of larger ship signatures. This suggests that its signature was typical of this broader class of ship. Thus, it seems reasonable to expect that straight-course directionality patterns might be similar, as other measurements (for example, Arveson and Vendittis[3]) have shown. However, larger vessels (particularly merchants) would have some difficulty in executing the tight turning maneuvers demonstrated herein, and only the most powerful ships would be able to maintain speed during such turns. Clearly, one should expect detailed differences in SSL between ships of different sizes (particularly draft), different hull forms, different number of propellers and rudders, and greater maximum speed. Deeper draft vessels (5–8 m source depths) have been found to exhibit clear LM propagation effects.[4] Significant differences may be exhibited by twin-propeller ships during turns as the heel of the ship moves the outside propeller deeper and the inside propeller shallower. Also, at higher speeds, the vortex shedding and wake bubble generation effects may spread acoustic sources in the along-ship dimension, which would influence directionality.

The analysis on turn rate dependency implicitly required an assumption that the aspect angle and speed dependencies were similar during turns and straight runs. The turn rate dependence then became a catch-all parameter after the assumed directionality and speed dependence had been removed. Greater exploration of this assumption is recommended through a set of carefully controlled experiments, varying the turn rate through changing the ships turning radius while adding extra throttle to maintain speed. It has been suggested that coupling between speed and turn rate may be such that their variations could be condensed into the ratio U/TR. The present data set do not span sufficient variations in these variables to assess this. Also unassessed by these data was whether there was a difference in radiated acoustic signature between the inside and outside of the turn. For safety reasons during these sea trials, the buoys were always located on the inside of the turn.

B. Recommendations for portable acoustic ranging

This section summarizes practical lessons learned during the sea trials with the VECTOR and similar trials with other ships.

(1) In order to fully quantify the ship signature including maneuvers, the following types of measurements are recommended, with multiple trials of each type as time allows:
   (a) straight, constant-speed runs at a variety of speeds.
   (b) SSL while the ship is drifting (very low speed).
   (c) Straight accelerating runs from low to maximum speed, possibly varying the applied power level.
   (d) 180° or 360° turning maneuvers at variety of fixed speeds and turn rates. In practice this implies maintaining constant speed through variation in throttle and varying the turn rate through specific rudder angles at a given speed.

(2) Ship-to-buoy CPA should be between approximately 50 and 100 m, increasing slightly for higher speed runs. Small CPA have the potential to overload receivers, and (for straight runs) provide only a small number of measurements at aspect angles near broadside.
(3) Buoy placement should attempt to provide both port and starboard aspects in the case of straight runs, and inside versus outside of the turn measurements in the case of maneuvers. This requires careful buoy placement to allow the ship to maneuver safely through the buoys.
(4) A set of at least two buoys should be deployed in order to provide spatial diversity, effective coverage of port-starboard, or inside-outside turn geometries, and measurement redundancy in the event of buoy malfunction. A set of four buoys was found to be reasonably easy to deploy and recover.
(5) Hydrophone depths should be greater than 20–30 m to provide separation from noise created by wave action on the surface float. Larger receiver depths also push the LM interference pattern to lower frequencies and allow measurements at greater horizontal ranges in the case of downward-refracting near-surface sound-speed gradients. The hydrophone cables should be decoupled from motions of the surface buoys through the use of compliant, damped suspension gear.
(6) The buoys should have two independent hydrophones deployed at two different depths. This allows verification of surface-reflected interference effects and potentially allows estimation of the effective source depth distribution and sea-surface reflection coefficient.
(7) Good quality hydrophones with relatively flat frequency response up to at least 20 kHz should be utilized. Integral preamplifiers will ensure sufficient electronic noise immunity through the relatively long cables.
(8) Each buoy should be equipped with a differential GPS receiver, recording positions at 1 s intervals. Similar DGPS recorders should be installed on the source ship. This allows calculation of parameters, such as ship-to-buoy range and aspect angle. Using two DGPS systems on the ship, located at the bow and stern, allows calculation of the instantaneous ship heading. Alternately, recording the ship's gyrocompass output (if feasible) or use of short base line, multiple receiver GPS arrays will provide improved measurement of the ship heading.
(9) The buoys should be designed to minimize the surface-penetrating area, presumably through use of a spar-buoy concept. This would reduce wave impact noise against the buoy hull and reduce vertical excursions of the buoy. The only above-water requirement is for support of the GPS receiver and whatever aids to navigation are required (e.g., radar reflectors). Addition of a remote position relay system would enable the ship to better maneuver around the buoys, and find them quickly during the recovery phase.
(10) During the VECTOR trials, important ship's propulsion information, such as propeller rpm, pitch, torque, and rudder angle, were recorded manually and did not account for rapid changes during turns. It is highly desirable to automatically record ship machinery state during runs. A data recording rate of 1 sample/s is recommended for maneuvering runs.
(11) Measurements should be conducted in a sheltered, relatively deep-water location, preferably in water depths greater than 200 m and over soft, acoustically absorbent seabeds. This minimizes background noise and seabed reflection interference.
(12) In these trials, the various maneuvering runs were conducted quickly, with typically less than 15 min. intervals between successive runs. However, care must be taken to ensure that bubbly wakes from previous runs have had sufficient time to dissipate. This is particularly important for higher speed, aggressive maneuvers where there is strong wake generation. It is recommended that a minimum of 20–30 min be allowed between successive runs.

ACKNOWLEDGMENTS

This work was made possible by the Defence R and D Canada Technology Investment Fund program. CCGS VECTOR ship time was provided by Fisheries and Oceans Canada, with additional laboratory and boat facilities provided by the Institute of Ocean Sciences (Sidney, BC). The authors wish to thank Mr. Ron Teichrob and Nick Hall-Patch of the Institute of Ocean Sciences for their assistance in BURB technical issues, including hydrophone calibrations. The Master, Officers, and crew of the CCGS VECTOR are acknowledged for their expert ship-handling and instrumentation deployment skills during the sea trials in Saanich Inlet. Mr. Richard Johnson of DRDC Atlantic conducted the BURB hydrophone calibrations.

[1]D. Ross, *Mechanics of Underwater Noise* (Peninsula, Los Altos, CA, 1987).
[2]P. Scrimger and R. Heitmeyer, "Acoustic source level measurements for a variety of merchant ships," J. Acoust. Soc. Am. **89**, 691–699 (1991).
[3]P. Arveson and D. Vendittis, "Radiated noise characteristics of a modern cargo ship," J. Acoust. Soc. Am. **107**, 118–129 (2000).
[4]S. Wales and R. Heitmeyer, "An ensemble source spectra merchant ship-radiated noise," J. Acoust. Soc. Am. **111**, 1211–1231 (2002).
[5]L. Gray and D. Greeley, "Source level model for propeller blade rate radiation for the world's merchant fleet," J. Acoust. Soc. Am. **67**, 516–522 (1980).
[6]M. Trevorrow, S. Vagle, and N. Hall-Patch, "Description and field evaluation of the broadband underwater recording buoy system," DRDC Atlantic Report No. TM 2005-231, Defence Research and Development Canada Atlantic, of Dartmouth, Nova Scotia, 2005.
[7]G. Wenz, "Acoustic ambient noise in the ocean, spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956 (1962).
[8]R. Francois and G. Garrison, "Sound absorption based on ocean measurements Part II: boric acid contribution and equation for total absorption," J. Acoust. Soc. Am. **72**, 1879–1890 (1982).
[9]R. Young, "Image interference in the presence of refraction," J. Acoust. Soc. Am. **19**, 1–7 (1947).

Characterization of an elastic target in a shallow water waveguide by decomposition of the time-reversal operator

Franck D. Philippe,[a] Claire Prada, Julien de Rosny, Dominique Clorennec, Jean-Gabriel Minonzio, and Mathias Fink
Laboratoire Ondes et Acoustique, Université Denis Diderot Paris 7, UMRS CNRS 7587, ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 19 September 2007; revised 5 May 2008; accepted 8 May 2008)

This paper reports the results of an investigation into extracting of the backscattered frequency signature of a target in a waveguide. Retrieving the target signature is difficult because it is blurred by waveguide reflections and modal interference. It is shown that the decomposition of the time-reversal operator method provides a solution to this problem. Using a modal theory, this paper shows that the first singular value associated with a target is proportional to the backscattering form function. It is linked to the waveguide geometry through a factor that weakly depends on frequency as long as the target is far from the boundaries. Using the same approach, the second singular value is shown to be proportional to the second derivative of the angular form function which is a relevant parameter for target identification. Within this framework the coupling between two targets is considered. Small scale experimental studies are performed in the 3.5 MHz frequency range for 3 mm spheres in a 28 mm deep and 570 mm long waveguide and confirm the theoretical results.
 [DOI: 10.1121/1.2939131]

PACS number(s): 43.30.Vh, 43.60.Tj, 43.30.Bp [DRD] Pages: 779–787

I. INTRODUCTION

In a shallow water waveguide, active detection and characterization of a target is an active field of research. Recent developments in sonar technology are very promising for target detection but efficient techniques to characterize those targets with conventional Sonar techniques remain scarce. A major difficulty comes from the dispersive nature of the waveguide. Indeed an impulse signal evolves into a long wave packet before reaching the target and then the backscattered wave is once again modified by the transfer function of the waveguide. As a consequence, the frequency fluctuations of a target's response are dominated by the boundary reflections and highly depend on the receiver depth.[1,2] Therefore, in shallow water, even if the backscattered signals are free of bottom reverberation, the extraction of the target's signature is very challenging.

To overcome this problem, the most common method is to deconvolve the transfer function of the waveguide. However, it is well known that the deconvolution problem for one transducer is ill posed; in other words, the solution is nonunique for a given set of data. In order to solve this problem, a solution is to use a vertical source-receiver array (SRA) that spans the whole water column. For instance, Mignerey and co-workers successfully demonstrated passive multichannel deconvolution using estimated boundary conditions and environmental data in an oceanic waveguide.[3,4]

Several studies address the direct or inverse problem of target scattering in a waveguide using a SRA. In particular, the scattering from an extended object is discussed by Yang and Yates in terms of mode coupling between the incident field and the scattered field.[5] They show that the use of a vertical SRA is necessary to acquire a large set of modes in order to extract the angular dependence of the scattered field.

In this paper, a method is proposed to circumvent the complexity of the acoustic response of the propagating medium by applying the decomposition of the time-reversal operator (DORT) method with a vertical SRA. In a waveguide, this method, like other time reversal techniques,[6–8] takes advantage of the multipath propagation to achieve high resolution detection and focusing. It requires the acquisition of the full interelement response matrix of the SRA. From the singular value decomposition (SVD) of the array response matrix, a set of singular vectors and singular values is determined. The singular vectors are eigenvectors of the time-reversal operator (TRO) and lead to the localization of pointlike scatterers. It was shown that in a waveguide, the transmission of eigenvectors of the TRO allows selective focusing on different pointlike targets.[9,10] This decomposition is also a means to separate the echo of a target from ground reverberation[11] and it was used to achieve reverberation focusing and nulling.[12,13] Recent at sea experiments confirm the efficiency of the DORT method for detection and focusing in a shallow water environment.[14,15] However, the aforementioned studies only considered isotropic scatterers.

Until now, application of the DORT method to elastic anisotropic scatterers (cylinders or spheres) has been studied in free space by Chambers and co-worker[16,17] and Minonzio *et al.*[18,19] It was shown that the frequency dependence of the singular values permits characterizing an elastic target. In fact, the first singular value is proportional to the backscattering form function used in underwater acoustics.[20]

This paper presents the results of the study into the effect of the waveguide on the singular values of the array response matrix and for the particular case of a Pekeris waveguide, analytical results are presented. It is organized as

[a]Electronic mail: franck.philippe@espci.fr

follows: first in Sec. II a theory is derived, showing that for one isotropic scatterer, if the array spans the whole water column, the backscattering form function of the target is given by the first singular value despite the reflections from the boundaries. Then this result is generalized for two scatterers and for large scatterers.

Finally, in Sec. III, ultrasonic waveguide experiments mimicking at sea underwater experiments are presented that confirm the theoretical results.

II. EXTRACTION OF THE SIGNATURE OF AN ISOTROPIC SCATTERER: THEORY

When using the DORT method, multiple paths from one target to the array are exploited allowing high resolution focusing. The resolution is increased because the effective aperture of the array in the waveguide is similar to a larger aperture virtual array in free space. In 1987, Ingenito[2] derived an expression for the acoustic field scattered by an elastic sphere using a normal mode decomposition. In the following, the extraction of the target signature is derived in a similar manner. Furthermore, it is demonstrated that the first singular value of the array response matrix is proportional to the target form function, that is to say, the monostatic frequency response in free space.

A. One isotropic scatterer

The analysis is performed in the frequency domain and for simplicity, the frequency dependence is kept implicit. The vector $\mathbf{h}_s$ has its ith element $[\mathbf{h}_s]_i$ equal to the Green's function between the scatterer at position $\mathbf{r}_s$ of cylindrical coordinates (z_s, r_s) and the transducer number i at position $\mathbf{r}_i$ of cylindrical coordinates (z_i, r_i). Thanks to the reciprocity principle, the vector describing the propagation for the wave from the array to the scatterer is given by the transpose $\mathbf{h}_s$ denoted $\mathbf{h}_s^t$.

Let us first consider an isotropic scatterer characterized by $S(0)$, the scattering coefficient. The expression of the array response matrix $\mathbf{K}$ is

$$\mathbf{K} = \mathbf{h}_s S(0) \mathbf{h}_s^t. \tag{1}$$

The matrix $\mathbf{K}$ is symmetrical and its rank equals 1 by construction. Thus the SVD of $\mathbf{K}$ is written as

$$\mathbf{K} = \mathbf{U}_1 \sigma_1 \mathbf{V}_1^\dagger, \tag{2}$$

where σ_1 is the unique singular value larger than 0, and $\mathbf{U}_1$ and $\mathbf{V}_1$ are the normalized singular vectors. The Hermitian transpose is noted.[†]

By identification of Eq. (1) with Eq. (2), the first singular value is $\sigma_1 = |S(0)| \, \|\mathbf{h}_s\|^2$. Hence, the singular value does not only depend on the scatterer form function but also on the propagation phenomena through $\|\mathbf{h}_s\|$. Using the Green's functions G, the norm of $\mathbf{h}_s$ is given by

$$\|\mathbf{h}_s\|^2 = \sum_{i=1}^{N} |G(z_i, z_s, r_s - r_i)|^2. \tag{3}$$

In a waveguide that is range independent, the Green's function only depends on three position variables, the depth of the source z_i, the depth of the scatterer z_s, and the range between the source and the probed position $R = |r_s - r_i|$ (Fig. 1). Assuming that the SRA samples sufficiently well the water column, the discrete sum in Eq. (3) is replaced by a continuous integral as follows:

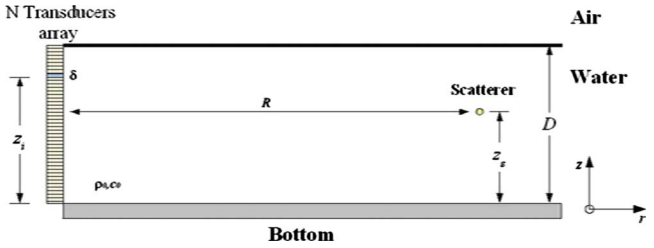


FIG. 1. (Color online) Geometry of the experiment and parameters.

$$\|\mathbf{h}_s\|^2 = \sum_{i=1}^{N} |G(z_i, z_s, r_s - r_i)|^2 \approx \frac{1}{\delta} \int_0^D |G(z_i, z_s, r_s - r_i)|^2 dz, \tag{4}$$

where δ is the array pitch and D is the water depth. The Green's function can be decomposed into the normal modes of the waveguide, i.e.,

$$G(z, z_s, R) = \frac{i}{4\rho} \sum_{m=0}^{\infty} \Psi_m(z_s) \Psi_m(z) H_0^{(1)}(k_{rm} R), \tag{5}$$

where $\Psi_m(z)$ is the normal mode associated with the vertical wave number k_{zm}. For an isospeed problem, the radial wave number k_{rm} is linked to k_0 and k_{zm} by the relation $k_{rm} = \sqrt{k_0^2 - k_{zm}^2}$. When R is large enough, replacement of the Hankel function of first order $H_0^{(1)}$ by its asymptotic approximation transforms the Green's function expression into

$$G(z, z_s, R) \approx \frac{i}{\rho \sqrt{8\pi R}} e^{-i\pi/4} \sum_{m=0}^{\infty} \Psi_m(z_s) \Psi_m(z) \frac{e^{ik_{rm}R}}{\sqrt{k_{rm}}}. \tag{6}$$

In a Pekeris waveguide, the impedance mismatch between the water and the bottom is finite. One can show that k_{rm} is purely real only when $k_b < k_{rm} < k_0$, where k_b is the bottom wave number and k_0 is the water wave number. In other words, when the propagation angle θ_m with the horizontal line associated with the mode m $[\cos(\theta_m) = k_{rm}/k_0]$ is lower than the Brewster angle, θ_b $[\cos(\theta_b) = c_0/c_b]$, the mode propagates with very little attenuation (other than geometrical spreading loss). For $\theta_m > \theta_b$, modes leak into the sea bottom and they decrease exponentially with range.[21] Moreover, the SRA transducers are not pointlike and thus have a directivity angle denoted θ_d. As a consequence, the transducer array is not sensitive to modes with $\theta_d < \theta_m$.

Finally, at long range, only the modes with $\theta_m < \theta_{\max}$ $[\theta_{\max} = \min(\theta_d, \theta_b)]$ contribute to the field recorded by the array. The Green's function is well approximated by replacing the infinite sum by a finite sum with a maximum index equal to $M-1$, where $M-1$ is such that $k_{rM1} = k_0 \cos(\theta_{\max})$.

Using Eqs. (4) and (6) and the orthogonality property between normal modes, the norm of $\mathbf{h}_s$ is given by

$$\|\mathbf{h}_s\|^2 = \frac{1}{\delta\rho 8\pi R}\sum_{m=0}^{M-1}\frac{\Psi_m^2(z_s)}{k_{rm}}. \tag{7}$$

Hence from the last equation, the expression for the singular value of a pointlike scatterer is deduced,

$$\sigma_1 = \frac{|S(0)|}{\rho 8\pi R k_0 \delta}\sum_{m=0}^{M-1}\frac{\Psi_m^2(z_s)}{\sqrt{1-\sin^2(\theta_m)}}, \tag{8}$$

where k_0 is the intrinsic wave number given by $k_0=2\pi f/c$ and $\sin(\theta_m)=k_{zm}/k_0$.

In the case of a Pekeris waveguide, the expressions for the lossless modes Ψ_m are well approximated by[21]

$$\Psi_m(z) = \sqrt{\frac{2\rho}{D}}\sin(k_{zm}z) \quad \text{and } k_{zm} = \frac{\pi}{2D}+\frac{m\pi}{D}.$$

Replacing $\Psi_m(z)$ by the above expression, Eq. (8) becomes

$$\sigma_1 = \frac{|S(0)|}{4\pi RDk_0\delta}\sum_{m=0}^{M-1}\frac{\sin^2(k_{zm}z_s)}{\sqrt{1-\sin^2(\theta_m)}}. \tag{9}$$

In most experiments, $\theta_{\max}$ is small and as $\theta_m > \theta_{\max}$, a series expansion in terms of $\sin^2(\theta_m)$ is justified,

$$\sigma_1 = \frac{|S(0)|}{4\pi RDk_0\delta}\sum_{m=0}^{M-1}\sin^2(k_{zm}z_s)\left(1+\frac{\sin^2(\theta_m)}{2}+\cdots\right). \tag{10}$$

Keeping only the first term of the expansion, σ_1 is approximated by

$$\sigma_1 \approx \frac{|S(0)|}{4\pi RDk_0\delta}\sum_{m=0}^{M-1}\sin^2(k_{zm}z_s). \tag{11}$$

Substitution of the expression of k_{zm} in Eq. (11) yields (Appendix A)

$$\sigma_1 \approx \frac{|S(0)|}{4\pi RDk_0\delta}\frac{M}{2}\left[1-\frac{1}{M}f_M\left(\frac{z_s}{D}\right)\right], \tag{12}$$

where $f_M(\xi)=\cos(\pi M\xi)[\sin(\pi M\xi)/\sin(\pi\xi)]$.

The function f_M is maximum at $\xi=0[f_M(0)=M]$ and minimum at $\xi=1$ $[f_M(1)=-M]$. Between these two extremes, the function is close to 0. The characteristic width (first zero) $\delta\xi$ associated with the two extremes equals $1/M$.

The expression for σ_1 deserves several comments. First, when the target is far from the boundaries, the second term in Eq. (12) is negligible so that σ_1 is constant with depth and equal to $|S(0)|M/(8\pi RDk_0\delta)$. Considering that $M=2\sin(\theta_{\max})D/\lambda$, the minimum distance between the target and the boundaries can be written as $\Lambda=\lambda/2\sin(\theta_{\max})$ and as $\theta_{\max}$ weakly depends on frequency, the only frequency dependence of σ_1 comes from $|S(0)|$. Note that this result can be generalized to other types of waveguides, provided k_{zm} is roughly proportional to m. Second, when the target is close to the surface, σ_1 decreases until it reaches 0. Indeed, due to the pressure release boundary condition, the pressure field is zero close to the surface and the target is invisible. In contrast, close to the bottom, the pressure field is maximum and the singular value is twice its value in the middle of the waveguide. This analytical formulation of the first singular value is in good agreement with the experimental results presented by Mordant *et al.* [9] Indeed, in this paper, the authors applied the DORT method to the detection of a 0.2 mm diameter nylon wire in a waveguide. The experiment was done with a 60-element transducer array with a central frequency of 1.5 MHz and an array pitch of 0.58 mm. The water waveguide was delimited by air and a steel bottom, it was 35 mm deep and the nylon wire was at range 400 mm. The wire was moved from the bottom to the surface, and for each position, the array response matrix was recorded, and the first singular value was plotted as a function of depth (Fig. 2).

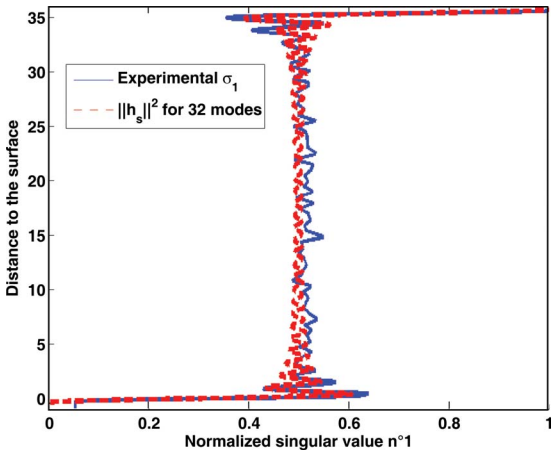


FIG. 2. (Color online) Comparison between the first singular value σ_1 given by Eq. (11) and the experimental singular value measured by Mordant *et al.* (Ref. 9) as a function of the target's depth for a frequency of 1.5 MHz.

Mordant *et al.* proposed a numerical model using the virtual images of the array with respect to the waveguide interfaces. They found that the experimental results were well fitted considering only the first 11 reflections. For such a number of reflections, Eq. (D4) in Appendix D gives a maximum of 32 modes. In Fig. 2, the first singular value given by Eq. (11) is plotted for $M=32$. The agreement between theoretical and experimental values is excellent.

B. Two isotropic scatterers

The case of two pointlike and isotropic scatterers can now be addressed. Neglecting multiple scattering, the array response matrix can be written as the sum of the transfer matrices of each scatterer alone [22]

$$\mathbf{K} = \mathbf{h}_{s1}S_1(0)\mathbf{h}_{s1}^t + \mathbf{h}_{s2}S_2(0)\mathbf{h}_{s2}^t, \tag{13}$$

where $\{\mathbf{h}_{s1}\}_i=G(z_{s1},z_i,r_{s1}-r_i)$ and $\{\mathbf{h}_{s2}\}_i=G(z_{s2},z_i,r_{s2}-r_i)$.

A coupling between the two scatterers will introduce complications in the singular value decomposition of K. Its formulation will not be as easily deduced as the two vectors $\mathbf{h}_{s1}$ and $\mathbf{h}_{s2}$ will not be orthogonal anymore. As a consequence, the same approach can be used on each term only if the inner product $\mathbf{h}_{s2}^\dagger\mathbf{h}_{s1}$ is small enough to be negligible. This product can be calculated using the modal decomposition

$$\mathbf{h}_{s2}^{\dagger}\mathbf{h}_{s1} = \frac{1}{\delta 4\pi R}\sum_{m=0}^{M-1}\frac{\Psi_m(z_{s1})\Psi_m(z_{s2})}{k_{rm}}e^{ik_{rm}(R_1-R_2)}. \quad (14)$$

Here, for simplicity, we assume that the two targets are at the same range ($R_1=R_2$). The inner product becomes

$$\mathbf{h}_{s2}^{\dagger}\mathbf{h}_{s1} \approx \frac{1}{\delta 4\pi R k_0}\sum_{m=0}^{M-1}\Psi_m(z_{s1})\Psi_m(z_{s2}). \quad (15)$$

In Appendix B, we show that this product can be expressed in terms of the f_M function

$$\mathbf{h}_{s2}^{\dagger}\mathbf{h}_{s1} \approx \frac{1}{\delta\rho 8\pi R k_0}\left[f_M\left(\frac{z_1+z_2}{2D}\right)-f_M\left(\frac{z_1-z_2}{2D}\right)\right]. \quad (16)$$

Far from the boundaries, the coupling between the two targets can be neglected when $|z_2-z_1|>D/M$. However, again $M\sim 2\sin(\theta_{\max})D/\lambda$ so $|z_2-z_1|>\Lambda$. This expression is consistent with the Rayleigh criterion taking into account the angular aperture $2\theta_{\max}$. Nevertheless, the modal approach is more general because it is still valid near the boundaries.

C. One extended axisymmetrical scatterer

An extended scatterer is characterized by a frequency dependent form function S depending on the incoming angle θ_1 and the outgoing angle θ_2. For axisymmetric scatterers at one frequency, the form function only depends on the angle difference $\theta_2-\theta_1$ and is even. Assuming small angles one can write the second order Taylor expansion as

$$S(\theta_2-\theta_1) = \left[S(0)+\frac{(\theta_2-\theta_1)^2}{2}S''(0)\right]. \quad (17)$$

A plane wave is written as $\varphi_\theta(r,z)=\exp[-ik_0(\sin\theta z+\cos\theta r)]$, where k_0 is the wave number, r is the range, and z is the height. Introduction of the following derivation operators ∂_z^L and ∂_z^R such that $\partial_z^L(\varphi,\psi)=(\partial\varphi/\partial z)\psi$ and $\partial_z^R(\varphi,\psi)=(\partial\psi/\partial z)\varphi$ will be useful for the following derivation. Indeed using those operators on two plane waves ϕ_{θ_1} and ϕ_{θ_2} yields $\partial_z^L(\varphi_{\theta_1},\varphi_{\theta_2})=-ik_0\sin\theta_1\varphi_{\theta_1}\varphi_{\theta_2}\approx -ik_0\theta_1\varphi_{\theta_1}\varphi_{\theta_2}$ and similarly $\partial_z^R(\varphi_{\theta_1},\varphi_{\theta_2})\approx -ik_0\theta_2\varphi_{\theta_1}\varphi_{\theta_2}$. Here the small angle approximation was made. As a consequence, Eq. (17) can be expressed in terms of φ_{θ_1}, φ_{θ_2} instead of θ_1, θ_2,

$$S(\theta_2-\theta_1) = S(0)\varphi_{\theta_1}\varphi_{\theta_2}|_{z,r=0} - S''(0) \times\frac{(\partial_z^L-\partial_z^R)^2}{2k_0^2}(\varphi_{\theta_1},\varphi_{\theta_2})\Big|_{z,r=0}. \quad (18)$$

As this approximation is valid for all pairs of plane waves with small angles, it can be generalized to any pair of fields satisfying the paraxial approximation. So, assuming that the Green's function $\{\mathbf{h}_s\}_i$ from any of the transducers to the scatterer center satisfies the parabolic approximation, the array response matrix can be written as

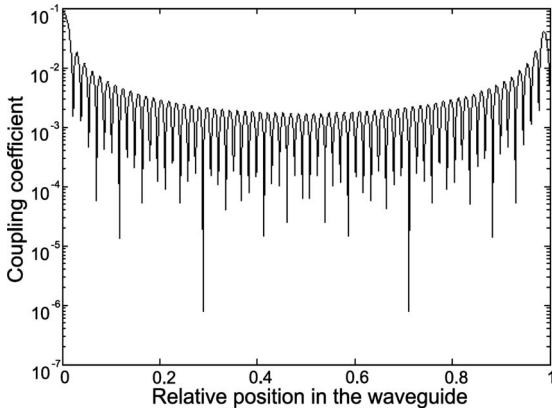


FIG. 3. Scalar product Σ of the first two singular vectors of the array response matrix $\mathbf{K}$ with respect to the relative depth position of the target in the waveguide.

$$\mathbf{K} = S(0)\mathbf{h}_s\mathbf{h}_s^t + \frac{S''(0)}{k_0^2}\partial_z\mathbf{h}_s\partial_z\mathbf{h}_s^t - \frac{S''(0)}{2k_0^2}[\mathbf{h}_s(\partial_z^2\mathbf{h}_s^t) + (\partial_z^2\mathbf{h}_s)\mathbf{h}_s^t]. \quad (19)$$

Finally, assuming that $\mathbf{h}_s^{\dagger}\cdot\partial_z\mathbf{h}_s\approx 0$, $\partial_z^2\mathbf{h}_s^{\dagger}\cdot\partial_z\mathbf{h}_s\approx 0$, and $\|\mathbf{h}_s\|\gg\|\partial_z^2\mathbf{h}_s\|/k_0^2$ (paraxial approximation), the first singular vector is $\mathbf{V}_1=\mathbf{h}_s/\|\mathbf{h}_s\|$ with the singular value $\sigma_1=S(0)\|h\|^2$ and the second singular vector is $\mathbf{V}_2=\partial_z\mathbf{h}_s/\|\partial_z\mathbf{h}_s\|$ with the singular value $\sigma_2=S''(0)\|\partial_z h\|^2/k_0^2$.

Hence the two first singular values in the waveguide are simply related to the backscattering form function $S(0)$ and its second derivative $S''(0)$ only when the scalar product

$$\Sigma = \frac{\mathbf{h}_s^{\dagger}}{\|\mathbf{h}_s\|}\cdot\frac{\partial_z\mathbf{h}_s}{\|\partial_z\mathbf{h}_s\|} \quad (20)$$

is small compared to 1.

An analytical expression for Σ is derived in Appendix C. The coupling term Σ is plotted with respect to the scatterer depth in Fig. 3. This coupling decreases from the maximum near the boundaries (about 10%) to the minimum (about 0.2%) over a characteristic distance given by Λ. These results are consistent with the study made by Gaunaurd and Huang for a sphere near a hard flat boundary.[20]

In Appendix D the ratio between the singular values obtained with and without the waveguide interfaces shows that the waveguide enhances the first singular value by a factor of $2R\sin(\theta_{\max})/D$ which is the number of reflections from the boundaries while the enhancement reaches $[2R\sin(\theta_{\max})/D]^3$ for the second singular value.

This result shows that the second singular value can be a good complementary parameter for the identification of a target in a waveguide.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental setup

The data have been obtained with an array made of 64 transducers working at 3.5 MHz using 1 μs pulses ($\lambda=0.43$ mm). The array pitch is $\delta=0.417$ mm leading to an array aperture equal to 26.7 mm. The waveguide interfaces are air/water and water/Plexiglas and the depth is D

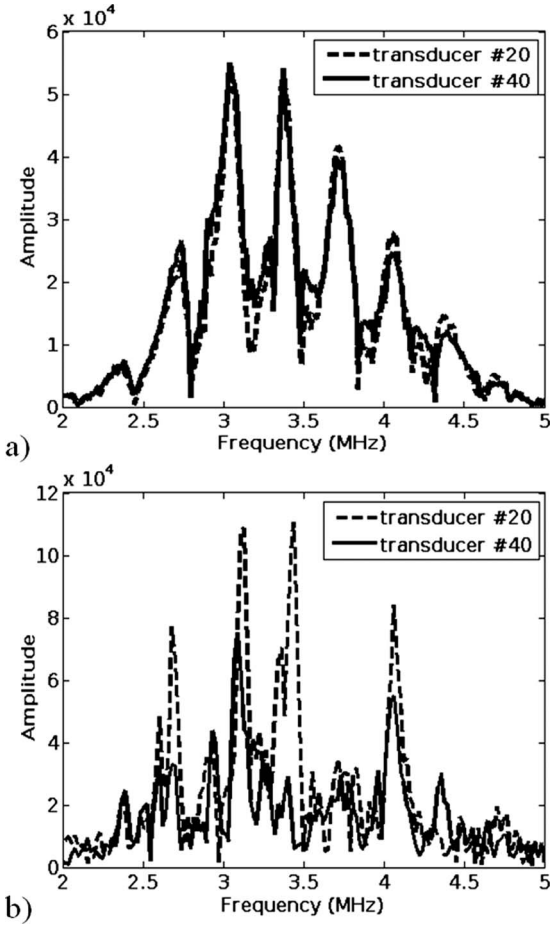


FIG. 4. Monostatic frequency response of 3 mm steel sphere for two transducers (a) in free space and (b) in a 28.1-mm-deep water waveguide.

=28.1 mm (65λ). The guide's width is 60 cm which is sufficiently large to avoid lateral reflections. The distance between the array and the targets is $R=575$ mm, i.e., 1340λ.

These small scale experiments give an idea of what would happen for at sea experiments. For a SRA working at 12.5 kHz, the equivalent waveguide would be 160 m long and 8 m deep.

B. Resonant scatterer

1. Backscattering form function

Two examples of the monostatic frequency response from a 3-mm-diameter steel sphere in free space are plotted in Fig. 4(a). As $2a > \lambda R/D$, the sphere can be considered as a large object. It clearly appears that the backscattered signal weakly depends on the transducer's position and the oscillations are related to the sphere resonances. In order to provide a comparison, two monostatic responses in the waveguide are plotted in Fig. 4(b). In free space, the backscattered response is directly proportional to the target backscattering form function while in the waveguide, the response is much more complex and strongly depends on the transducer position. These effects are due to the complex interference pattern generated by the multiple reflections at the waveguide boundaries.

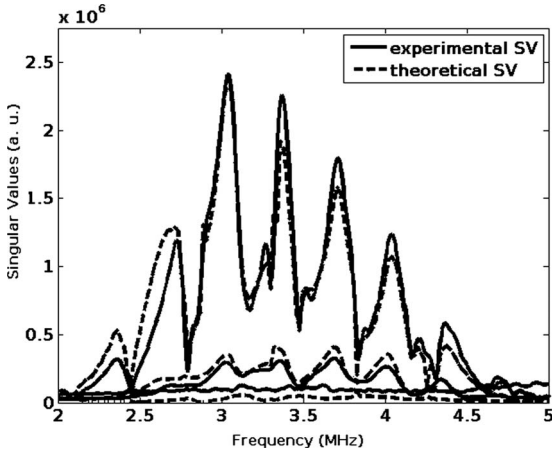


FIG. 5. First three singular values of a 3 mm steel sphere in a 28.1-mm-deep waveguide at distance $R=575$ mm. The experimental data (continuous line) are obtained in the waveguide. The theoretical curves (dash-dotted line) are computed in a free space configuration (Ref. 19) with a virtual SRA corresponding to three reflections at the surface and the bottom.

2. Form function and singular values

The scatterer is a 3-mm-diameter steel sphere in the waveguide at a depth of 13 mm. The first three singular values obtained in the waveguide are plotted in Fig. 5 versus frequency. These curves are compared to the theoretical ones obtained in a free space configuration [19] accounting for the corresponding virtual array. For the computation, the size of the virtual SRA, D', is chosen such as $D'=2R\sin(\theta_{\max})$ with $\theta_{\max}=20°$ corresponding to a virtual aperture seven times larger than the real array aperture.

This result shows that DORT facilitates the extraction of the backscattering signature and its second derivative without frequency distortion. Even more, the waveguide, by a virtual aperture effect, increases the singular values, leading to a better target, characterization in noisy environment. This effect is particularly significant with the second singular value which is proportional to $\sin^3(\theta_{\max})$.

The first two singular vectors are numerically backpropagated with the range-dependent acoustic model (RAM) numerical simulation [23]. To that end, the RAM configuration takes into account the waveguide geometry, in particular, the weak bottom slope of 0.4°. At one frequency, the experimental singular vector is phase conjugated and each element of the vector is the phase and amplitude excitation for the elements of the simulated array. After numerical propagation, the acoustic field value is stored in computer memory. The simulation is repeated for several frequencies between 2 and 5 MHz. The averaged field around the focus is plotted in Fig. 6 in logarithmic scale.

As in free space the first and the second singular values lead, respectively, to symmetric and antisymmetric focusing. However, this is only true in the vicinity of the sphere because unlike in free space, far from the focus, the focal spot symmetry is lost due to interaction with the boundaries. The focal spot is 1.4 mm wide; thus the virtual array obtained is at least six times larger than the real one, thanks to the re-

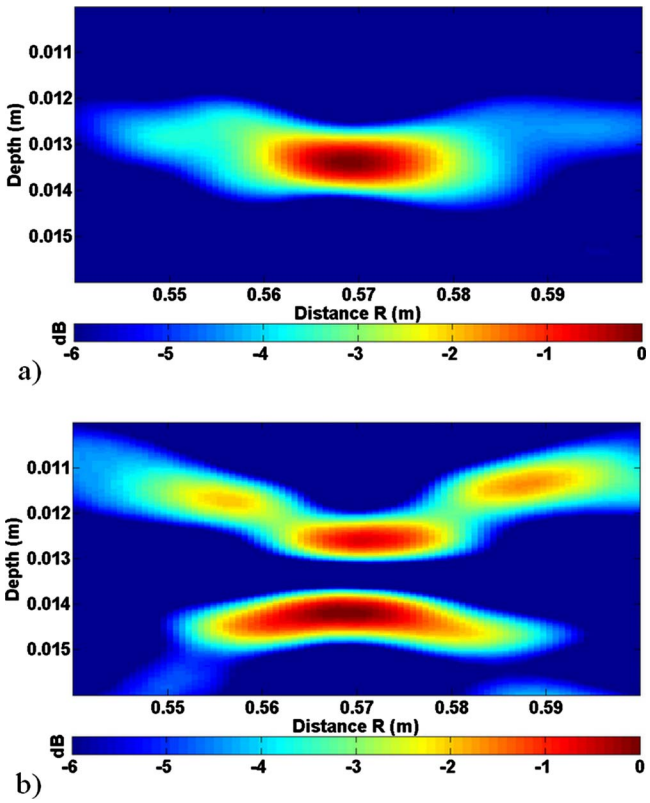


FIG. 6. (Color online) Backpropagation with RAM code of the first two singular vectors. (a) and (b) represent the acoustical energy averaged over all frequencies between 2 and 5 MHz in logarithmic scale.

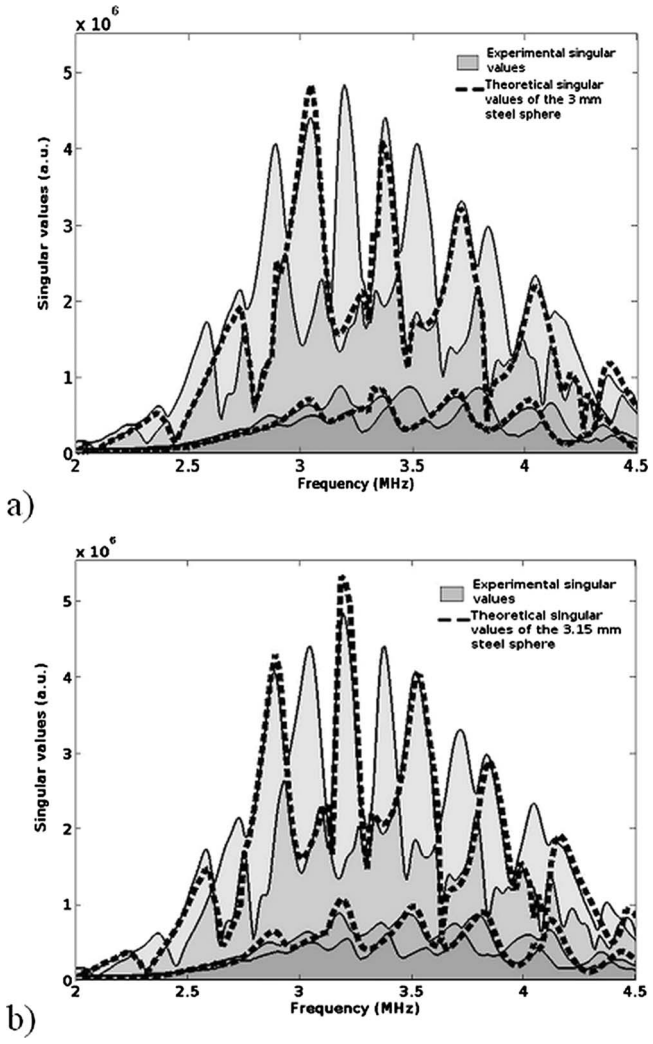


FIG. 7. Two different spheres in a waveguide: comparison between the first four experimental singular values and the first two theoretical singular values of the (a) 3 mm steel sphere and (b) 3.15 mm steel sphere.

flections on the interfaces of the waveguide. This is in agreement with the virtual aperture $D' = 197$ mm introduced previously.

C. Separation of two elastic spheres

The following experiment illustrates two aspects of the DORT method: the detection of two targets at the same range and the extraction of their form functions. The experiment was performed in the same waveguide as in the previous sections with two steel spheres of radii differing by 5%. This radius mismatch induces a 5% difference in the resonance frequencies. The distance between the two spheres is 10 mm which is much larger than $D/\sin(\theta_{\max})$, so that, according to the results of Sec. II B, the responses of the two spheres weakly interact.

The first four singular values as a function of the frequency are displayed in Fig. 7. There is no simple way to associate a singular value with a target, in particular, the first singular value may not correspond to the same target at all frequencies. One target may have the highest reflectivity at one frequency (for example, at a resonance frequency) while the other has the highest reflectivity at another frequency. This is why here the first two singular values corresponding to the two backscattering form functions of the spheres are intertwined due to the slight radius mismatch. The target corresponding to the first singular value is either the 3 mm sphere or the 3.15 mm sphere. The frequencies where the switchings occur are emphasized in Figs. 8(a) and 9(a) by vertical lines. The two backscattering form functions can be assigned to their corresponding spheres by comparing the singular value to the theoretical backscattering form function of each sphere taken alone. Indeed, as shown in Sec. II B, for distant scatterers, there is one singular value associated with each scatterer. Each singular value is the one obtained at the same conditions but without the other scatterer. Similar effects occur for the other two singular values.

The backpropagation of the singular vectors as a function of frequency is shown in Figs. 8 and 9 and confirms this result. The backpropagated field associated with the first singular value focuses at the depth of the sphere with the highest backscattering form function. The second singular vector focuses at the position of the other sphere. The same effect occurs when the third singular vector is backpropagated; the dipolar focal spot is centered on the strongest target. The switches from one sphere position to the other correspond to singular value crossings that are shown by vertical lines in Fig. 9. Thus the backpropagation provides a means to reconstruct the frequency dependence of the backscattering form function for each target.

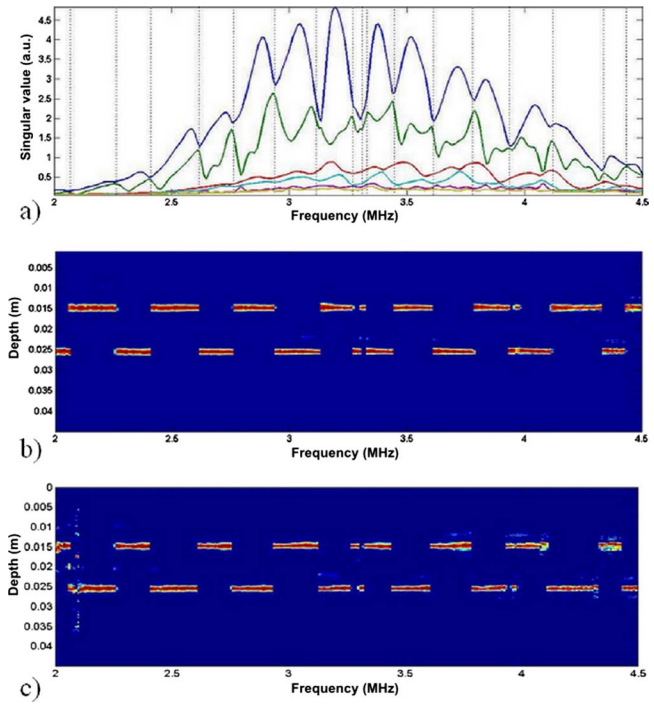


FIG. 8. (Color online) (a) Measured singular values of the two spheres as a function of frequency. A vertical line designates a frequency where a switch in the relative strength of the targets occurs. (b) Backpropagation with RAM code of the first eigenvector as a function of frequency and depth at a range of 575 mm. (c) Backpropagation with RAM code of the second eigenvector at the same distance.

IV. CONCLUSION

In this paper, the authors have presented a theoretical and experimental study of the eigenvectors of the time reversal operator for a scatterer in shallow water with a modal approach to wave propagation. By using a normal mode expansion and introducing the form function of an isotropic target within the paraxial approximation, the length $\Lambda = \lambda/2 \sin(\theta_{\max})$ was demonstrated to correspond to the Rayleigh criterion. Indeed, it was shown that two targets at the same range are weakly coupled when the depth difference is larger than Λ and that when one target is at least Λ away from the boundaries, the frequency dependence of the first singular value is nearly identical to the backscattering target signature. Moreover it was shown that the second singular value is proportional to the second derivative of the angular form function. The second singular value is an interesting complementary identification parameter especially because guided propagation enhances it by a factor N_{reflex}^3 which is the number of reflections to the power 3.

The small scale experiments presented in this paper are in good agreement with the theory. The extended targets used were steel spheres, each having several associated singular values. Their backscattered frequency signatures were successfully extracted from the impulse response matrix without any evaluation of the waveguide parameters. The DORT method is able to detect, separate, and discriminate between two spheres with diameters that only differ by 5%. Such precision is very promising for further at sea experiments.

Hence it was shown that the DORT method enables target characterization without the need of precise knowledge of the waveguide when the target is at least Λ away from any other element of the medium.

Furthermore, additional knowledge can be obtained when an accurate numerical model of the waveguide is available: localization of the targets is possible and backpropagation of the singular vectors provides a means to associate a form function with a target without any ambiguity.

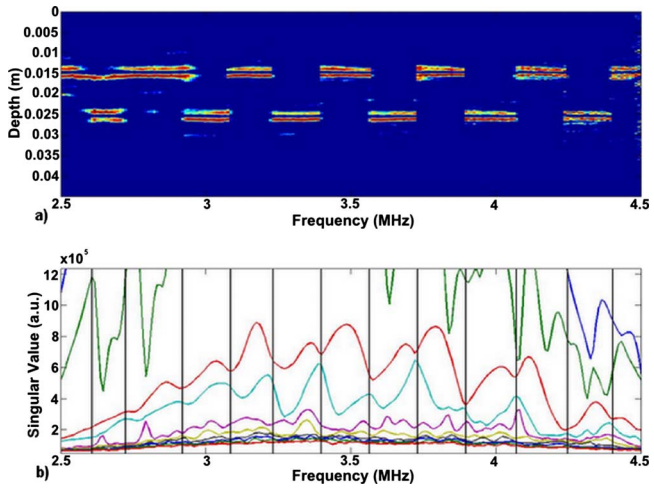


FIG. 9. (Color online) (a) Backpropagation with RAM code of the third eigenvector as a function of frequency and depth at a range of 575 mm. The focusing is antisymmetric with respect to the scatterer center as predicted by theory. (b) Zoom on the third and fourth singular values as a function of frequency.

APPENDIX A: FIRST SINGULAR VALUE FOR AN ISOTROPIC SCATTERER IN A WAVEGUIDE

The singular values σ_1 is given by

$$\begin{aligned}\sigma_1 &= \frac{|S(0)|}{4\pi RDk_0\delta}\sum_{m=0}^{M-1}\sin^2(k_{zm}z_s)\\ &= \frac{|S(0)|}{4\pi RDk_0\delta}\sum_{m=0}^{M-1}\frac{1-\cos(2k_{zm}z_s)}{2}\\ &= \frac{|S(0)|}{8\pi RDk_0\delta}\left[M-\Re e\left(\sum_{m=0}^{M-1}e^{i2(m+1/2)\pi/Dz_s}\right)\right]\\ &= \frac{|S(0)|}{8\pi RDk_0\delta}\left[M-\Re e\left(e^{i(\pi/D)z_s}\sum_{m=0}^{M-1}(e^{i2(\pi/D)z_s})^m\right)\right].\end{aligned} \tag{A1}$$

Moreover, the geometric series with common ratio $e^{i2(\pi/D)z_s}$ has the following exact solution:

$$\sum_{m=0}^{M-1}(e^{i2(\pi/D)z_s})^m = \frac{1-e^{i2(\pi/D)Mz_s}}{1-e^{i2(\pi/D)z_s}}. \tag{A2}$$

Thus Eq. (1) becomes

$$\sigma_1 = \frac{|S(0)|}{8\pi RDk_0\delta}\left[M - \cos\left(\frac{\pi M z_s}{D}\right)\frac{\sin\left(\frac{\pi M z_s}{D}\right)}{\sin\left(\frac{\pi z_s}{D}\right)}\right]. \quad \text{(A3)}$$

For convenience the function f_M is defined as follows:

$$f_M(\xi) = \left(\cos(\pi M\xi)\frac{\sin(\pi M\xi)}{\sin(\pi\xi)}\right). \quad \text{(A4)}$$

The expression of the first singular value can then be concisely written as

$$\sigma_1 = \frac{|S(0)|}{8\pi RDk_0\delta}\left[M - f_M\left(\frac{z_s}{D}\right)\right]. \quad \text{(A5)}$$

APPENDIX B: DERIVATION OF THE COUPLING TERM FOR TWO SCATTERERS

At the same range, the inner product between $\mathbf{h}_{s1}$ and $\mathbf{h}_{s2}$ is given by

$${}^{\dagger}\mathbf{h}_{s2}\mathbf{h}_{s1} = \frac{1}{\delta\rho 8\pi R}\sum_{m=0}^{\infty}\Psi_m(z_{s1})\Psi_m(z_{s2}). \quad \text{(B1)}$$

Again we consider a Pekeris waveguide. Moreover, assuming that the array is only sensitive to small incident angles, the inner product can be approximated by

$${}^{\dagger}\mathbf{h}_{s2}\mathbf{h}_{s1} \approx \frac{1}{\delta 4\pi RDk_0}\sum_{m=0}^{M-1}\sin(k_{zm}z_{s1})\sin(k_{zm}z_{s2}). \quad \text{(B2)}$$

Using a well-known trigonometric relation, it becomes

$${}^{\dagger}\mathbf{h}_{s2}\mathbf{h}_{s1} \approx \frac{1}{\delta 8\pi RDk_0}\sum_{m=0}^{M-1}[\cos(k_{zm}[z_{s1}-z_{s2}]) - \cos(k_{zm}[z_{s1}+z_{s2}])]. \quad \text{(B3)}$$

The sum of the cosines is calculated as in Appendix A.

$${}^{\dagger}\mathbf{h}_{s2}\mathbf{h}_{s1} \approx \frac{1}{\delta\rho 8\pi RDk_0}\left[f_M\left(\frac{z_1+z_2}{2D}\right) - f_M\left(\frac{z_1-z_2}{2D}\right)\right]. \quad \text{(B4)}$$

APPENDIX C: DERIVATION OF THE COUPLING TERM BETWEEN $\mathbf{h}_s$ AND $\partial_z\mathbf{h}_s$

Here we want to compute the norm of the first derivative of $\mathbf{h}$ over z and the inner product between $\mathbf{h}_s$ and $\partial_z\mathbf{h}_s$, i.e.,

$$\|\partial_z\mathbf{h_s}\|^2 = \frac{1}{\delta\rho 8\pi R}\sum_{m=0}^{\infty}\frac{[\partial_z\Psi_m(z_s)]^2}{k_{rm}}, \quad \text{(C1)}$$

$${}^{\dagger}\mathbf{h_s}\partial_z\mathbf{h_s} = \frac{1}{\delta\rho 8\pi R}\sum_{m=0}^{\infty}\frac{\Psi_m(z_s)\partial_z\Psi_m(z_s)}{k_{rm}}. \quad \text{(C2)}$$

Assuming a perfect waveguide and a cutoff for the mode summation, Eqs. (C1) and (C2) can be written as

$$\|\partial_z\mathbf{h_s}\|^2 = \frac{1}{\delta 4\pi RDk_0}\sum_{m=0}^{M-1}k_{zm}^2\cos^2(k_{zm}z_s), \quad \text{(C3)}$$

$${}^{\dagger}\mathbf{h_s}\partial_z\mathbf{h_s} = \frac{1}{\delta 4\pi RDk_0}\sum_{m=0}^{M-1}k_{zm}\sin(k_{zm}z_s)\cos(k_{zm}z_s). \quad \text{(C4)}$$

These two expressions can be worked out using the first and second derivatives of f_M. Indeed,

$${}^{\dagger}\mathbf{h_s}\partial_z\mathbf{h_s} = \frac{1}{\delta 4\pi RDk_0}\frac{1}{2D}f'_M\left(\frac{z_s}{D}\right), \quad \text{(C5)}$$

and

$$\|\partial_z\mathbf{h_s}\|^2 = \frac{1}{\delta 8\pi RDk_0}\left(\frac{(4M^3-M)\pi^2}{12D^2} + \frac{f''_M\left(\frac{z_s}{D}\right)}{4D^2}\right). \quad \text{(C6)}$$

$$\partial_\xi f_M(\xi) = \left(\frac{2M\sin(\pi\xi)\cos(2\pi M\xi) - \cos(\pi\xi)\sin(2\pi M\xi)}{2\sin^2(\pi\xi)}\right), \quad \text{(C7)}$$

$$\Sigma = \frac{f'_M\left(\frac{z_s}{D}\right)}{\sqrt{\frac{M - f_M\left(\frac{z_s}{D}\right)}{4}\left[\frac{(4M^3-M)\pi^2}{3} + f''_M\left(\frac{z_s}{D}\right)\right]}}. \quad \text{(C8)}$$

APPENDIX D: ENHANCEMENT OF THE SINGULAR VALUES DUE TO THE WAVEGUIDE

In free space, the first singular value is easily calculated:

$$\sigma_1^{\text{free}} = S(0)\|h_{\text{free}}\|^2 \quad \text{(D1)}$$

with Eq. (4) and $G(z,z_s,r_s-r_i)=(1/4\pi)(e^{-ikR}/R)$ the first singular value in free space becomes

$$\sigma_1^{\text{free}} = \frac{S(0)D}{\delta(4\pi R)^2}, \quad \text{(D2)}$$

where D is the array aperture and δ is the array pitch.

In the same manner the second singular value is

$$\sigma_2^{\text{free}} = \frac{S''(0)\|\partial_z h_{\text{free}}\|^2}{k_0^2} \approx \frac{S''(0)D^3}{48\delta R^4\lambda^2 k_0^2}. \quad \text{(D3)}$$

It was shown in Sec. II C that in a waveguide, the expression of the singular values is

$$\sigma_1^{\text{guide}} = \frac{|S(0)|M}{8\pi Rk_0D\delta} \quad \text{and} \quad \sigma_2^{\text{guide}} = \frac{S''(0)\|\partial_z h_{\text{guide}}\|^2}{2k_0}.$$

As a consequence

$$\frac{\sigma_1^{\text{guide}}}{\sigma_1^{\text{free}}} = \frac{|S(0)|M}{8\pi Rk_0D\delta} \times \frac{\delta(4\pi R)^2}{|S(0)|D}$$

with

$$M = \frac{N_{\text{reflex}} D^2}{\lambda R}. \tag{D4}$$

The ratio becomes $\sigma_1^{\text{guide}} / \sigma_1^{\text{free}} = N_{\text{reflex}}$.

Keeping only the M^3 term, Eq. (C6) becomes

$$\sigma_2^{\text{guide}} = \frac{S''(0) M^3 \lambda}{48 \delta R D^3 k_0^2} = \frac{S''(0)(2 \sin \theta_d)^3}{48 \delta R k_0^2 \lambda^2}, \tag{D5}$$

with $M = 2 \sin \theta_d D / \lambda$.

$$\frac{\sigma_2^{\text{guide}}}{\sigma_2^{\text{free}}} = \frac{(2 \sin \theta_d)^3 R^3}{D^3} = N_{\text{reflex}}^3. \tag{D6}$$

[1] R. H. Hackman and G. S. Sammelmann, "Multiple scattering analysis for a target in an oceanic waveguide," J. Acoust. Soc. Am. **84**, 1813–1825 (1988).

[2] F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).

[3] P. Mignerey and S. Finette, "Multichannel deconvolution of an acoustic transient in an oceanic waveguide," J. Acoust. Soc. Am. **92**, 351–364 (1992).

[4] S. Finette, P. Mignerey, J. Smith III, and C. Richmond, "Broadband source signature extraction using a vertical array," J. Acoust. Soc. Am. **94**, 309–318 (1993).

[5] T. C. Yang and T. W. Yates, "Scattering from an object in a stratified medium. II. Extraction of scattered signature," J. Acoust. Soc. Am. **96**, 1020 (1994).

[6] 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).

[7] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, "A long-range and variable focus phase conjugation experiment in shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1998).

[8] P. Roux and M. Fink, "Time reversal in a waveguide: Study of the temporal and spatial focusing," J. Acoust. Soc. Am. **107**, 2418–2429 (2000).

[9] N. Mordant, C. Prada, and M. Fink, "Highly resolved detection and selective focusing in a waveguide using the D.O.R.T. method," J. Acoust. Soc. Am. **105**, 2634–2642 (1999).

[10] B. Pinçon and L. Ramdani, "Selective focusing on small scatterers in acoustic waveguides using time reversal mirrors," Inverse Probl. **23**, 1–25 (2007).

[11] T. Folégot, C. Prada, and M. Fink, "Resolution enhancement and separation of reverberation from target echo with the time reversal operator decomposition," J. Acoust. Soc. Am. **113**, 3155–3160 (2003).

[12] J. F. Lingevitch, H. C. Song, and W. A. Kuperman, "Time reversed reverberation focusing in a waveguide," J. Acoust. Soc. Am. **111**, 2609–2614 (2001).

[13] H. C. Song, W. S. Hodgkiss, W. A. Kuperman, P. Roux, T. Akal, and M. Stevenson, "Experimental demonstration of adaptive reverberation nulling using time reversal," J. Acoust. Soc. Am. **118**, 1381–1387 (2005).

[14] C. Prada, J. De Rosny, D. Clorennec, J. G. Minonzio, A. Aubry, M. Fink, L. Bernière, S. Hibral, P. Billand, and T. Folégot, "Experimental detection and focusing in shallow water by decomposition of the time reversal operator," J. Acoust. Soc. Am. **122**, 768–761 (2007).

[15] C. F. Gaumond, D. M. Fromm, J. F. Lingevitch, R. Menis, G. F. Edelmann, D. C. Calvo, and E. Kim, "Demonstration at sea of the decomposition-of-the-time-reversal-operator technique," J. Acoust. Soc. Am. **119**, 976–990 (2006).

[16] D. H. Chambers and A. K. Gautesen, "Time reversal for a single spherical scatterer," J. Acoust. Soc. Am. **109**, 2616–2624 (2001).

[17] D. H. Chambers, "Analysis of the time-reversal operator for scatterers of finite size," J. Acoust. Soc. Am. **112**, 411–419 (2002).

[18] J. G. Minonzio, C. Prada, D. Chambers, D. Clorennec, and M. Fink, "Characterization of subwavelength elastic cylinders with the decomposition of the time-reversal operator: Theory and experiment," J. Acoust. Soc. Am. **117**, 789–798 (2005).

[19] J. G. Minonzio, F. D. Philippe, C. Prada, and M. Fink, "Characterization of an elastic cylinder and an elastic sphere with the time-reversal operator: application to the sub-resolution limit," Inverse Probl. **24**, 025014 (2008).

[20] G. C. Gaunaurd and H. Huang, "Acoustic scattering by a spherical body near a plane boundary," J. Acoust. Soc. Am. **96**, 2526–2536 (1994).

[21] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), Chap. 2, pp. 123–127.

[22] J. G. Minonzio, C. Prada, A. Aubry, and M. Fink, "Multiple scattering between two elastic cylinders and invariants of the time-reversal operator: Theory and experiments," J. Acoust. Soc. Am. **102**, 875–883 (2006).

[23] M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

Bayesian geoacoustic inversion of ship noise on a horizontal array

Dag Tollefsen[a)]
Norwegian Defence Research Establishment (FFI), Box 115, 3191 Horten, Norway

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

(Received 11 March 2008; accepted 12 May 2008)

This paper applies geoacoustic inversion to low-frequency narrow-band acoustic data from a quiet surface ship recorded on a bottom-moored horizontal line array in shallow water. A Bayesian matched-field inversion method is employed which quantifies geoacoustic uncertainties and allows for meaningful comparison of inversion results from different data sets. Geoacoustic inversion results for ship-noise data are compared with inversion results for multitone data from a towed controlled source collected in the same experiment, and with independent geophysical measurements. To increase the information content of low-level ship-noise data, the effect of including multiple, independent data segments in the inversion is investigated and shown to significantly reduce geoacoustic parameter uncertainties. Geoacoustic uncertainties are also shown to depend on ship range and orientation, with increased uncertainties for long ranges and for the ship stern oriented away from the array. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2940581]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 788–795

I. INTRODUCTION

Knowledge of geoacoustic parameters of the seabed is required for many sonar applications. Matched-field inversion (MFI) is a method to infer seabed parameters from acoustic field data measured in the water column. Much work on MFI has been performed using high-level acoustic sources and vertical line arrays of sensors.[1–7] Recent interest has expanded to include applications to data from towed[8–11] and bottom-moored[12–14] horizontal line arrays (HLAs), and the use of alternative sound sources such as noise from the tow ship received on a towed array,[9,10] or from ships of opportunity received on moored arrays.[15,16,13] The use of ship noise in both cases reduces environmental impact over a controlled source. The use of a towed HLA offers an advantage over moored arrays in terms of system mobility. However, factors such as low signal-to-noise ratio (SNR) of tow-ship noise[9] and limitations on source–receiver separations (hence limited sampling of the acoustic angular spectrum) can limit the information in such data. The use of moored arrays and noise from ships of opportunity allows for unobtrusive geoacoustic characterization, and has further advantages in terms of simplicity and economy (i.e., a dedicated ship is not required).

Geoacoustic inversion using noise from ships of opportunity and a vertical line array has been reported by Chapman *et al.*,[15] who applied MFI to broadband ship-noise data, and by Nicholas *et al.*,[16] who applied MFI to narrow-band noise from a research ship. Koch and Knobles[13] used broadband noise from a ship end fire to a bottom-moored HLA for geoacoustic inversion, and provided a qualitative discussion of the shape of the distributions of model samples collected via simulated annealing. However, no work to date has provided a quantitative assessment of the information content of ship noise for geoacoustic inversion in terms of rigorous parameter uncertainty estimation, including how information content varies with ship range and orientation, and with number of observations; these issues are addressed here.

This paper considers geoacoustic inversion of ship-noise data from a quiet surface ship collected at a bottom-moored HLA in shallow water. A Bayesian matched-field inversion method is employed to estimate model parameters and to quantify their uncertainty distributions. This allows for meaningful comparisons of the geoacoustic information content of different data sets. We demonstrate that including multiple, independent time segments in the inversion can significantly reduce uncertainties in the geoacoustic parameter estimates. The effects of ship orientation and ship range on inversion results are also quantified and discussed. Finally, results from inversion of ship noise are compared with results from inversion of controlled-source data as well as supporting geophysical data from the experiment region.

II. THEORY

In Bayesian MFI, information regarding the model of geoacoustic parameters $\mathbf{m}$ is obtained from the posterior probability density (PPD), $P(\mathbf{m}|\mathbf{d})$, according to Bayes' rule,

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m}). \quad (1)$$

In Eq. (1), $P(\mathbf{m})$ is the prior distribution and, for (fixed) measured data $\mathbf{d}$, $P(\mathbf{d}|\mathbf{m})$ is interpreted as a function of $\mathbf{m}$, the likelihood function, which can generally be expressed $L(\mathbf{m}) \propto \exp[-E(\mathbf{m})]$ for an appropriate data misfit function E (considered in the following). Defining a generalized misfit

[a)]Electronic mail: dag.tollefsen@ffi.no

$\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m})$, the PPD may be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m})]}{\int \exp[-\phi(\mathbf{m}')]d\mathbf{m}'}, \tag{2}$$

where the integration spans the model space. The multidimensional PPD is interpreted in terms of properties defining parameter estimates, uncertainties, and interrelationships, such as the maximum *a posteriori* (MAP) estimate, the posterior mean estimate, the model covariance matrix, parameter mean deviations (MD), and marginal probability distributions defined, respectively, as

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max}\{P(\mathbf{m}|\mathbf{d})\}, \tag{3}$$

$$\bar{\mathbf{m}} = \int \mathbf{m} P(\mathbf{m}|\mathbf{d}) d\mathbf{m}, \tag{4}$$

$$\mathbf{C} = \int (\mathbf{m} - \bar{\mathbf{m}})(\mathbf{m} - \bar{\mathbf{m}})^T P(\mathbf{m}|\mathbf{d}) d\mathbf{m}, \tag{5}$$

$$\mathrm{MD}_i = \int |m_i - \bar{m}_i| P(\mathbf{m}|\mathbf{d}) d\mathbf{m}, \tag{6}$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i - m_i') P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{7}$$

where δ is the Dirac delta function. Parameter correlations are quantified by normalizing the covariance matrix to produce the correlation matrix,

$$S_{ij} = C_{ij}/\sqrt{C_{ii}C_{ij}}. \tag{8}$$

For nonlinear problems such as geoacoustic inversion, the integrals in Eqs. (4)–(7) can be solved numerically using the Markov-chain Monte Carlo method of fast Gibbs sampling.[4–6] Computing the MAP estimate requires minimizing the generalized misfit ϕ, which can be performed using a numerical optimization algorithm, e.g., the adaptive simplex simulated annealing method.[17]

The ship-noise data considered in this paper consist of complex acoustic fields measured at an N-sensor array and F frequencies corresponding to J distinct segments of the recorded acoustic pressure time series, with each time segment divided into K subsegments, i.e., $\mathbf{d} = \{\mathbf{d}_{fjk}, f = 1, F; j = 1, J; k = 1, K\}$. The source–receiver range is considered to be fixed over the K subsegments comprising each time segment, but range varies between segments (the subsegments, referred to as snapshots, allow for data averaging to improve SNR). The source spectrum (amplitude and phase) is considered unknown over frequency and time. The data errors are assumed to be complex, circularly symmetric Gaussian-distributed random variables, uncorrelated in space, frequency, and time, with unknown variances which depend on frequency and time segment but are considered constant over snapshots, i.e., ν_{ij}, $i = 1, F$; $j = 1, J$. In this case, the likelihood function is given by

$$L(\mathbf{m}) = \prod_{f=1}^{F} \prod_{j=1}^{J} \prod_{k=1}^{K} \frac{1}{(\pi \nu_{fj})^N} \times \exp[-|\mathbf{d}_{fjk} - A_{fjk} e^{i\theta_{fjk}} \mathbf{d}_{fjk}(\mathbf{m})|^2 / \nu_{fj}], \tag{9}$$

where $\mathbf{d}(\mathbf{m})$ represents modeled (replica) data, and A and θ represent source amplitude and phase. Maximizing the likelihood over unknown variance and source spectrum by setting $\partial P(\mathbf{d}|\mathbf{m})/\partial \nu_{fj} = \partial P(\mathbf{d}|\mathbf{m})/\partial A_{fjk} = \partial P(\mathbf{d}|\mathbf{m})/\partial \theta_{fjk} = 0$ leads to the data misfit (log-likelihood) function

$$E(\mathbf{m}) = N \sum_{f=1}^{F} \sum_{j=1}^{J} \log_e B_{fj}(\mathbf{m}), \tag{10}$$

where $B_{fj}(\mathbf{m})$ is the Bartlett mismatch defined by

$$B_{fj}(\mathbf{m}) = \mathrm{Tr}\{\mathbf{C}_{fj}\} - \frac{\mathbf{d}_{fj}^{\dagger}(\mathbf{m}) \mathbf{C}_{fj} \mathbf{d}_{fj}(\mathbf{m})}{|\mathbf{d}_{fj}(\mathbf{m})|^2}. \tag{11}$$

In Eq. (11), $\mathrm{Tr}\{\cdot\}$ represents the matrix trace, the dagger represents conjugate transpose, and $\mathbf{C}_{fj}$ is the data cross-spectral density matrix (CSDM) for the fth frequency and jth time segment defined by the ensemble average over the K snapshots:

$$\mathbf{C}_{fj} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{d}_{fjk} \mathbf{d}_{fjk}^{\dagger}. \tag{12}$$

It is important to note the distinction between including more data in the form of more snapshots in Eq. (12) or more time segments in Eq. (10). Averaging over more snapshots generally increases the SNR; however, for a moving source, there is a limit on the number of snapshots that can be considered to correspond to a fixed range. Including more time segments adds data at different ranges which increases information content, but also increases computational effort.

III. EXPERIMENT, DATA, AND MODEL

The acoustic data and supporting environmental measurements considered here were collected in an experiment conducted in the Barents Sea in 2003, which is described in detail in Ref. 14. The experiment region and ship tracks selected for analysis are illustrated in Fig. 1. An 18-element, 900 m long HLA was deployed on the seabed at a water depth of approximately 282 m, oriented north–south with the north end at the origin of the coordinate system of Fig. 1. The array consisted of seven sensors spaced at 10 m intervals at the north end followed by eleven sensors at increasing spacing to a maximum of 160 m at the south end. The ship tracks were radial and oriented at angles of approximately 30° with respect to the array length axis (end fire). The ship was outbound with the stern oriented toward the array during the east (outbound) track, and inbound with the bow oriented toward the array during the west (inbound) track. The ship-to-array ranges for the ship-noise data considered here are 1.49–1.70 and 5.09–5.30 km along the outbound track and 1.46–1.67 km along the inbound track. All acoustic data were collected within a 3 h time interval.

Supporting oceanographic measurements consisted of water-column temperature and salinity profiles measured us-

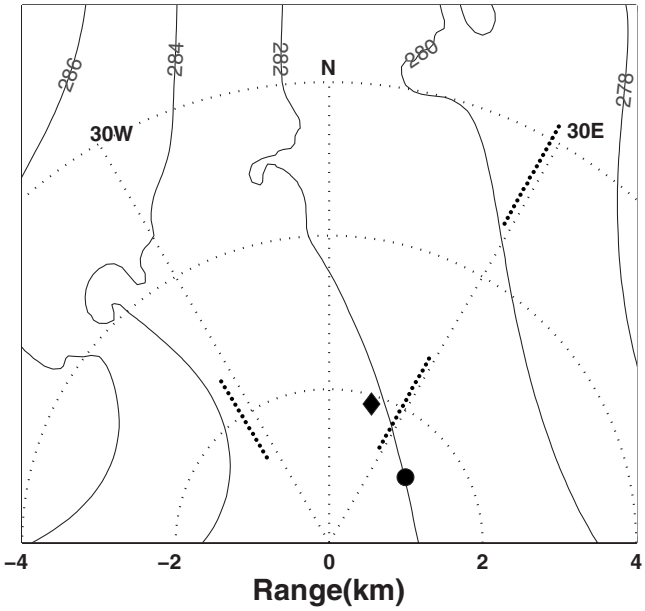


FIG. 1. Area of experiment. The HLA was laid north–south with the north end at the origin of the coordinate system. Solid lines are water depth (m) contours. Dotted lines indicate range (km) and azimuth (°) to the array; heavy dotted lines indicate ship tracks used for analysis. The diamond indicates the location of the gravity core, and the closed circle indicates the receiver location for the seismic refraction survey.

ing a conductivity-temperature-depth (CTD) probe at the start of the outbound track and the end of the inbound track, and expendable bathythermograph casts along the tracks. The calculated sound-speed profiles showed little variation with time and position. Sound speed was nearly constant at 1475 m/s in an upper layer of 40 m depth, followed by an abrupt decrease to 1470 m/s at 100 m depth and increase to 1473 m/s at the seabed. The seas were calm (sea state 1) during both tracks. Supporting geophysical measurements consisted of seismic-reflection and subbottom profiling sonar data along a track parallel to the outbound track, a shallow gravity core, and a wide-angle bottom refraction survey. The positions of the latter two measurements are indicated in Fig. 1. The seismic-reflection data indicated an upper layer of unconsolidated sediment of thickness 120–140 m overlying layers of consolidated material. The subbottom profiling sonar indicated some internal structure near the sea floor with a possible weak reflector at 10–20 m depth. Analysis of the gravity core indicated a sound speed of 1500–1520 m/s and density of 2.0–2.1 g/cm^3 over the upper 1 m of the core. The seismic refraction survey gave an averaged value of 1745 m/s for sound speed in unconsolidated sediment.

Previous work[14] considered inversion of data from a towed controlled source at five frequency components within 30–160 Hz at relatively high SNR of 9–15 dB. Considered here are selected narrow-band frequency components emanating from the tow-ship R/V H U SVERDRUP II recorded concurrently with the controlled-source data. The R/V H U SVERDRUP II (55 m length overall, 400 ton displacement, 5.2 m draft), a relatively quiet research ship, was moving at a constant speed of 5.2 kn along both tracks.

The processing sequence for the ship-noise data consisted of forming CSDM estimates from $K=10$ consecutive

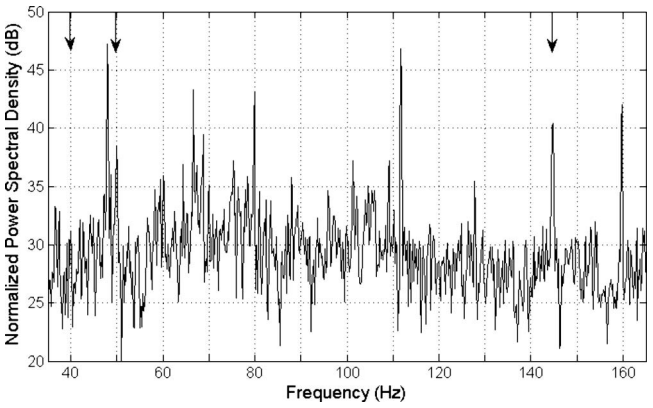


FIG. 2. Normalized power spectral density from one hydrophone of the HLA (short-range outbound track data). Arrows indicate the three ship-noise frequency lines used for inversion.

50% overlapping data snapshots of length 3.3 s, converted to the frequency domain using a fast Fourier transform with a frequency bin width of 0.3 Hz, with a Hamming time windowing function applied. The total averaging time is 18.2 s, over which the ship moved approximately 50 m. (This processing scheme differs slightly from that used for the controlled-source data;[14] however, the same total averaging time was used for both data types.) A simple frequency tracker was applied to follow shifts in the frequency lines of the ship between data segments. A plot of the normalized power spectral density from one hydrophone of the array (for data from the short-range outbound track) is provided in Fig. 2. The arrows indicate the three frequency components of the ship noise selected for inversion (the first frequency component is relatively weak in this plot). These frequency components were chosen because they are within the frequency band used for the controlled-source inversions,[14] are known to emanate from the research ship (identified using near-field acoustic data from ship passages close to the array), and were found to be relatively stable (i.e., not of transient nature) over the period of data collection. The five highest-level frequency components in Fig. 2 are tones or overtones from the controlled source; additional relatively strong frequency components are either less prominent or unstable tones of the tow ship, or could possibly originate from distant ships.

The estimated average SNR of the ship-noise data is given in Fig. 3. The SNR estimates are based on an average of power spectral densities (after snapshot averaging) over the elements of the array and over five data segments. For the

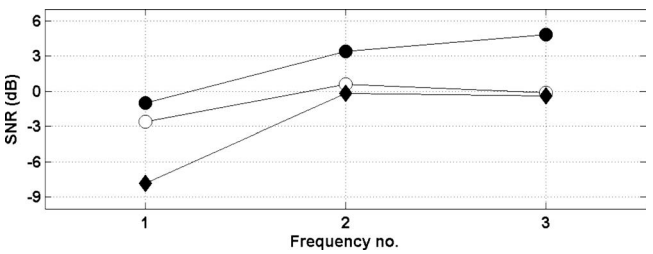


FIG. 3. Average SNR for ship noise at three frequencies. The short-range outbound track, long-range outbound track, and short-range inbound track are indicated by closed circles, closed diamonds, and open circles, respectively.

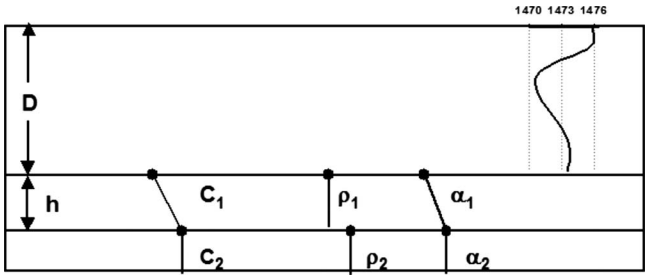


FIG. 4. Geoacoustic model. The upper layer represents the water column [inset: the sound-speed profile (m/s)]; the seabed is divided into two layers. See Table I for symbols and parameter units.

outbound track, the SNR varies from −1 to +5 dB for the three processed frequencies for the short-range data, while at long range the SNR varies from −8 to 0 dB. For the short-range inbound-track data, the SNR varies from −3 to 0 dB. The lower SNR for these data compared to the outbound track data at similar range, and the fact that the frequency-to-frequency variation in SNR differs from that of the two outbound tracks is assumed related to directionality of these ship-noise frequency components, with generally higher noise levels in the direction of the stern. (Measurements for a larger ship reported in Ref. 18 indicate similar directionality of ship noise at low frequencies.)

A simple geoacoustic model consistent with the geophysical data was developed for the purposes of inversion. (The same model was used in the controlled-source inversions of Ref. 14.) The model (Fig. 4) consists of a water column of depth D with a known sound-speed profile over a two-layer seabed. The seabed model consists of an upper layer with depth-dependent properties over a homogeneous basement half-space. The geoacoustic parameters used to describe the upper seabed model layer are the layer thickness (h), sound speed at the top and bottom of the layer (c_1 and c_2), attenuation at the top and bottom of the layer (α_1 and α_2), and a depth-independent density (ρ_1). The lower layer is described by sound-speed and attenuation values that are identical to those at the base of the upper layer, and by an independent density value (ρ_2). This parametrization provides continuous profiles for the sound speed and attenuation, with gradients assumed in the upper layer (restricted to positive gradients for sound speed). In addition to the seven geoacoustic model parameters, small corrections to the nominal values of water depth (D), and to source range (r), source depth (z), and source bearing (b) were also included in the inversions (the latter three parameters are repeated for each data segment). The model parameters and the prior search bounds applied in the inversions are given in Table I.

The assumptions on the data error statistics in Sec. II were checked by performing *a posteriori* tests on data residuals[6] [difference between measured data and modeled data for the optimal model computed by minimizing Eq. (10), for each snapshot and data segment]. The Lilliefors test was applied to test for Gaussianity, with no evidence against the hypothesis of Gaussian-distributed errors at a 95% confidence level in 96% of test cases. The runs test provided no evidence against spatial randomness (i.e., between elements of the HLA) at a 95% confidence level in 98% of test cases.

TABLE I. Model parameters and search bounds used in inversions.

Parameter and units	Lower bound	Upper bound
h (m)	1.0	40
c_1 (m/s)	1450	1900
c_2 (m/s)	c_1	c_1+30h
ρ_1 (g/cm^3)	1.40	3.00
ρ_2 (g/cm^3)	1.40	3.00
α_1 (dB/m kHz)	0.01	1.00
α_2 (dB/m kHz)	0.01	1.00
D (m)	278	286
dr (km)	−0.10	+0.10
z (m)	3.0	6.0
b (°)	27	30

Tests for residual correlation over frequency and over time (i.e., between data segments) are not meaningful in this context due to the small number of samples (three and five, respectively) available for testing. However, with these small numbers of samples, the effect of potential error correlations is substantially reduced.

IV. RESULTS

A. Multiple data segments

The following presents results from inversion of ship-noise data from the outbound track at ranges of 1.49–1.70 km and considers the effect of increasing data content, in the form of multiple data segments, in the inversion. Each data segment is represented by the CSDM from data averaged over a time interval of 18.2 s. Data segments are combined under the assumption of a range-independent environment. Replica acoustic fields were generated using the normal-mode propagation model ORCA,[19] using a complex-plane mode search to account for possible near-field effects due to continuous modes. The normalized Bartlett mismatch was within 0.34–0.81 for these data. [These estimates were obtained by evaluating Eq. (11) for each data segment and frequency at the maximum-likelihood model obtained by minimizing the data misfit function defined in Eq. (10). Note that this misfit function implicitly weights the data at each frequency according to its estimated uncertainty.]

Geoacoustic inversion results for the cases of one, two, and five data segments are shown in Fig. 5 in terms of marginal PPDs for the seven geoacoustic model parameters described in Sec. III, and an additional geoacoustic parameter (c_{ave}) describing the average sound speed in the upper sediment layer (computed from inversion parameters c_1, c_2, and h).[14] In addition to the seven geoacoustic parameters and water depth, each inversion included three geometric parameters for each data segment (e.g., a total of 23 parameters for inversions with five data segments). First consider the results for a single data segment in the upper distributions of Fig. 5. The most sensitive parameters are those determining the sound-speed profile in the seabed (defined by c_1, c_2, h). Densities and attenuations are relatively insensitive parameters with essentially flat distributions. However, even for the sen-

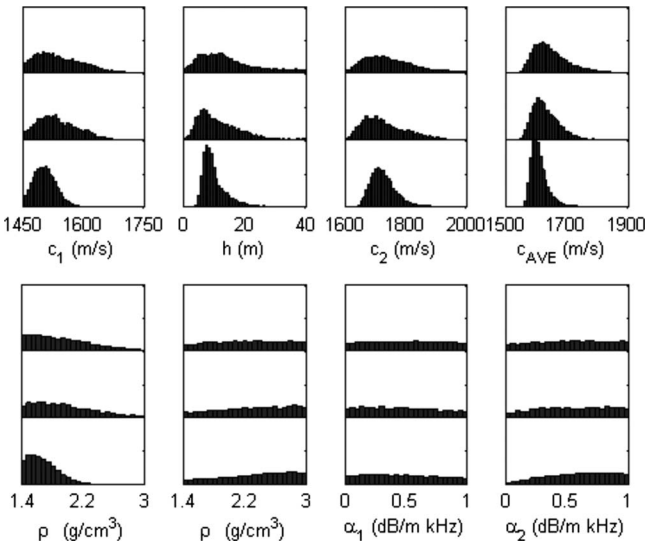


FIG. 5. Marginal PPDs for different number of data segments included in inversion of ship-noise data: One segment (upper distributions), two segments (middle distributions), and five segments (lower distributions).

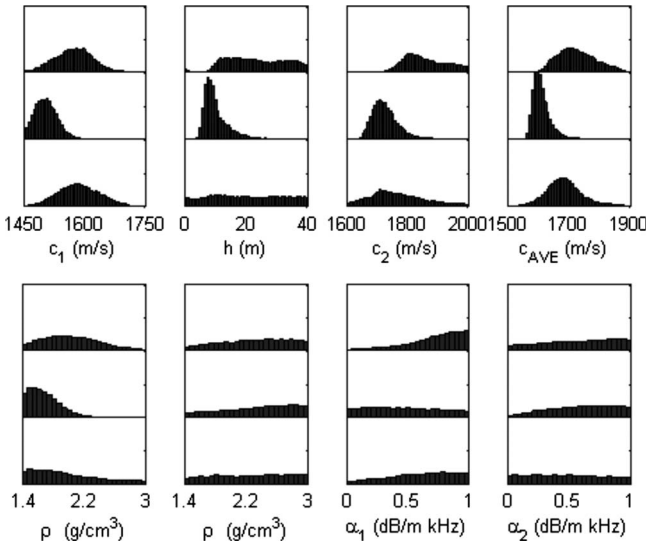


FIG. 7. Marginal PPDs for short-range inbound (upper distributions), short-range outbound (middle distributions), and long-range outbound (lower distributions) ship-noise data.

sitive parameters, the marginal distributions are relatively wide, indicating limited ability to resolve the seabed sound-speed profile. By including two and five data segments in the inversions (middle and lower distributions in Fig. 5), the uncertainty distributions for the parameters defining the seabed sound-speed profile narrow significantly. This is further quantified in Fig. 6, where MDs for the three inversions are compared for each geoacoustic parameter. Reductions in MD by a factor of approximately 2 from one data segment to five data segments are observed for the sensitive parameters. For example, the mean deviation for the top sound speed c_1 is reduced from 50 m/s for one data segment to 40 m/s for two data segments, and to 23 m/s for five data segments. For sound speed c_2 in the lower layer, the MD is reduced from 111 m/s for one data segment to 43 m/s for five segments. The uncertainty of upper-layer density ρ_1 is also reduced; however, for the five-segment case the marginal distribution

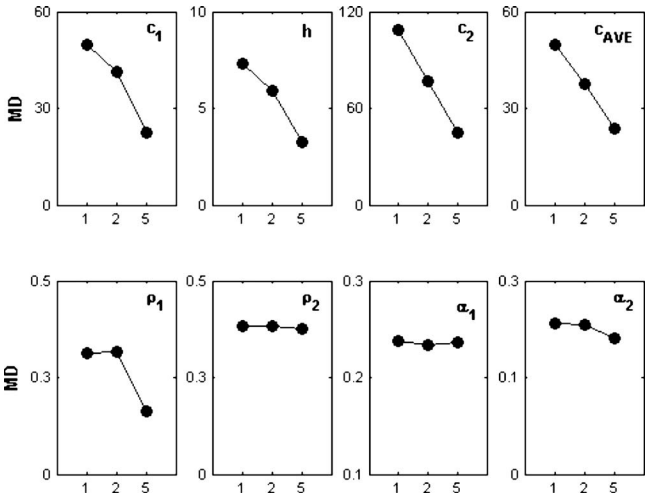


FIG. 6. Posterior uncertainty estimates of geoacoustic parameters quantified in terms of mean absolute deviations for: 1, 2, and 5 data segments included in the inversion. Parameter units vary between plots and are given in Table I.

for this parameter is significantly constrained by the lower *a priori* bound (Fig. 5), and hence the MD can underestimate the uncertainty. There is no significant reduction in MDs for the insensitive parameters α_1, α_2, and ρ_2.

B. Ship range and orientation

We now examine the effects of range and ship orientation on ship-noise geoacoustic inversion by considering additional data from the outbound track at 5.09–5.30 km range and from the inbound track at 1.46–1.67 km range. The normalized Bartlett mismatch was within 0.54–0.95 and 0.56–0.89 for these data, respectively. Geoacoustic inversion results for the two additional data sets are shown in Fig. 7 in terms of marginal PPDs for geoacoustic parameters, with results for the short-range outbound data considered in Sec. IV A included for reference (middle distributions). Five data segments are included in all inversions. Figure 7 shows that there is general consistency between the three results (i.e., the distributions have considerable overlap). The upper sound speed c_1 is reasonably well determined for the short-range inbound and long-range outbound data, although at lower resolution than for the short-range outbound data. However, distributions for h and c_2 (and c_{ave}) are considerably wider; thus, the detailed structure of the sound-speed profile in the seabed is not well resolved by the long-range outbound and short-range inbound data. The distribution widths for c_1 and h are slightly narrower for the short-range inbound data than for the long-range outbound data. The mean and MD of c_1 from the short-range inbound and long-range outbound data are 1586 ± 43 and 1573 ± 39 m/s, respectively, while the values for the short-range outbound data are 1507 ± 23 m/s. The apparent shift to higher c_1 values can be interpreted as a consequence of a loss of resolution of the structure of the seabed sound-speed profile (i.e., the distinction between upper and lower sound speeds becomes smeared).

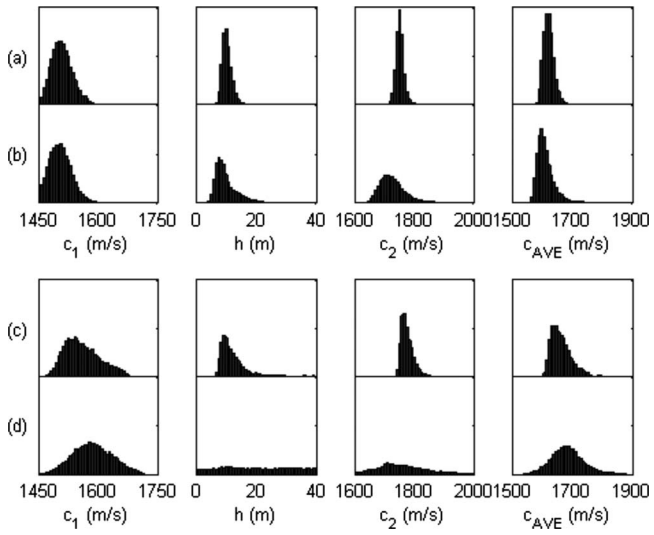


FIG. 8. Marginal PPDs for: (a) Short-range outbound controlled-source data, (b) short-range outbound ship-noise data, (c) long-range outbound controlled-source data, and (d) long-range outbound ship-noise data.

The average SNR for the short-range inbound data is 2–5 dB lower than the short-range outbound data as shown in Fig. 3, likely due to directional effects (Sec. III). This lower SNR could explain the poorer results for the inbound data when compared with the outbound data from similar range. The estimated SNR for the long-range outbound data is 4–7 dB lower than the short-range outbound data due to the loss of signal strength with range (since the ship orientation with respect to the HLA is unchanged). This decrease in SNR degrades the long-range inversion results, although modal attenuation likely also contributes to degraded results for this case (discussed in the following section).

C. Ship-noise versus controlled-source inversions

The following compares results of inversion of ship-noise data from the outbound track with previous results from inversion of controlled-source data,[14] and also with reference geophysical measurements. The SNR of the controlled-source data used in the inversions (five frequency components) varied from 9 to 15 dB (for both short-range and long-range data, since the source level was changed with range). Figures 8(a) and 8(b) show marginal PPDs for the parameters defining the seabed sound-speed profile for short-range controlled-source and ship-noise data, respectively. There is excellent agreement between the results obtained from the two data sets, but with wider distributions (increased uncertainty) for the ship-noise data. For both data sets, the detailed structure of the seabed sound-speed profile is well resolved. Results are quantified in terms of parameter mean estimates with MD uncertainties in Table II. Values for controlled-source and ship-noise data are, respectively, 11.2 ± 1.4 and 11.2 ± 3.1 m for h, 1510 ± 21 and 1507 ± 23 m/s for c_1, and 1753 ± 13 and 1737 ± 43 m/s for c_2. Table II also lists approximate values for these parameters obtained from reference geophysical measurements[14] ($h = 10–20$ m, $c_1 = 1500–1520$ m/s, and $c_2 = 1745$ m/s), which agree well with the geoacoustic inversion results.

TABLE II. Geoacoustic parameter estimates (mean with mean-deviation uncertainties) from inversion of ship-noise and controlled-source data (Ref. 14) at range of 1.5 km (outbound track). Also included are approximate values from supporting geophysical measurements.

Parameter and units	Ship noise	Controlled source	Geophysical data
h (m)	11.2 ± 3.1	11.2 ± 1.4	10–20
c_1 (m/s)	1507 ± 23	1510 ± 21	1500–1520
c_2 (m/s)	1737 ± 43	1753 ± 13	1745
ρ_1 (g/cm^3)	1.69 ± 0.16	2.03 ± 0.20	2.0–2.1
ρ_2 (g/cm^3)	2.31 ± 0.38	2.06 ± 0.33	
α_1 (dB/m kHz)	0.49 ± 0.23	0.32 ± 0.18	
α_2 (dB/m kHz)	0.60 ± 0.21	0.21 ± 0.13	

Figures 8(c) and 8(d) show geoacoustic inversion results for long-range controlled-source and ship-noise data, respectively. The seabed sound-speed profile parameters are reasonably well determined by the higher-SNR controlled-source data, while ship-noise data determine only c_1. Estimates for c_1 are 1586 ± 43 m/s for controlled-source data and 1559 ± 37 m/s for ship-noise data; in both cases these values are higher than for the corresponding short-range inversions.

The degradation in seabed resolution from the short- to long-range ship-noise data [Figs. 8(b) and 8(d)] is likely not due solely to decreased SNR. Some degradation is also observed between the short- and long-range controlled-source data [Figs. 8(a) and 8(c)], even thought the SNRs are almost identical for these data. In Ref. 14 the degradation for the controlled-source data was attributed to attenuation of higher-order modes. This has two possible effects: loss of resolution of deeper seabed structure (due to loss of information from modes propagating at higher grazing angles, hence with deeper bottom penetration), and loss of resolution of shallower structure (due to loss of information from the short vertical wavelengths of higher-order modes).

To further investigate the effects of SNR and range on seabed resolution, we consider inversions of synthetic ship-noise data. These data were generated for the model environment described in Sec. III, with geoacoustic parameters taken to be the mean values for the controlled-source inversion (Table II). Noise was represented by an error variance added to the main diagonal of the CSDM (e.g., Ref. 7), with variances representative of the misfit of the outbound ship-noise data.[20] The results from inversion of short- and long-range synthetic data in Figs. 9(a) and 9(b) closely resemble those

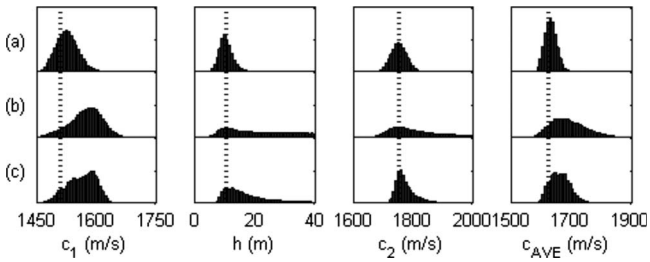


FIG. 9. Marginal PPDs for: (a) Short-range synthetic ship-noise data, (b) long-range synthetic ship-noise data, and (c) long-range synthetic high-SNR ship-noise data. Vertical dashed lines indicate true parameter values.

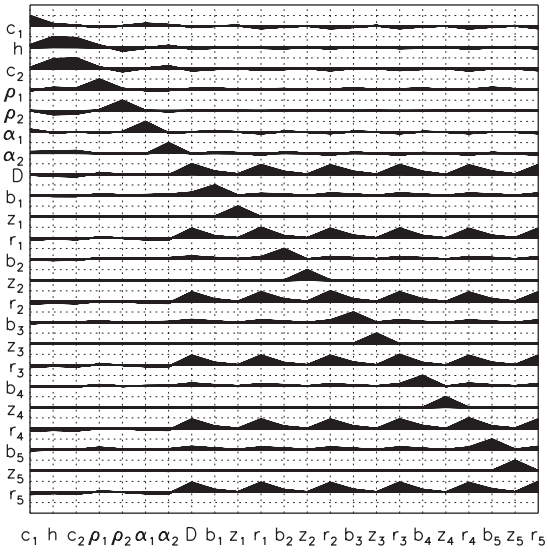


FIG. 10. Parameter correlation matrix for short-range outbound ship-noise data.

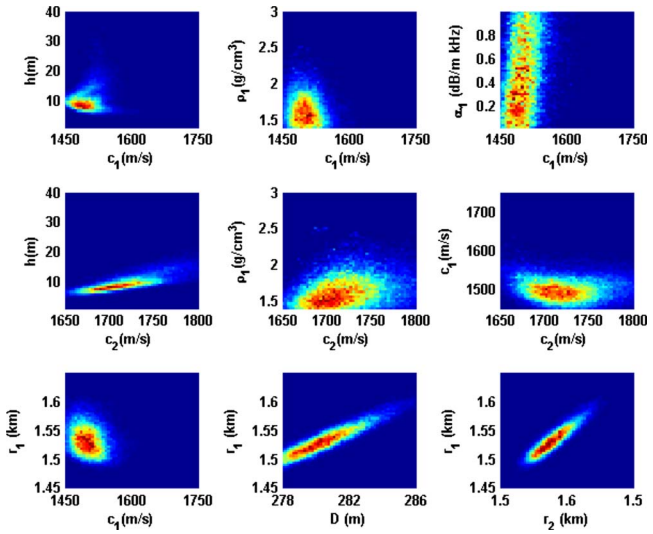


FIG. 11. (Color online) Selected joint marginal PPDs for short-range outbound ship-noise data.

obtained from inversion of experimental ship-noise data in Figs. 8(b) and 8(d), although parameter distributions are slightly narrower for the synthetic data. Of note is the fact that the marginal distributions for c_1 for synthetic data appear shifted to higher values relative to the true value, and that the shift increases with range. To isolate the effect of range on inversion results, inversion of long-range synthetic data was repeated with the SNR equal to that of short-range (ship-noise) data. Results are presented in Fig. 9(c): The wider distributions for c_1, h, and c_2, and the shift in c_1 to higher values when compared with Fig. 9(a) are due to the range effect; the further widening of distributions in Fig. 9(b) can be interpreted as an additional effect of lower SNR.

D. Parameter interrelationships

Finally, it is of interest to examine interparameter correlations in ship-noise inversion. Figure 10 shows the correlation matrix for the short-range outbound ship-noise data. Since five data segments are included in the inversion, the model included 15 source parameters (five sets of b, z, and r). The positive correlations between layer thickness h and layer sound speeds c_1 and c_2 indicate that the data have limited ability to distinguish between models with thicker, faster layers and thinner, slower layers (since both can produce similar acoustic transit times through the layer). Correlations between geometric and geoacoustic parameters are in general small, with weak negative correlations between each range parameter r_i and sound speeds c_1 and c_2. Within the geometric parameters, there are positive correlations between range parameters and water depth D; this corresponds to a known environmental mismatch effect in matched-field processing,[21] and was also obtained from inversion of controlled-source data.[14] Finally, there are strong positive correlations between each of the range parameters r_i. This indicates that the range corrections in these inversions could likely be treated by a single offset parameter that applies to all source ranges, thus reducing the dimensionality of the inversion problem.

Figure 11 shows joint marginal PPDs for selected combinations of parameters. The joint marginals for the correlated parameters, specifically for h and c_1, h and c_2, and r_1 and D (similar results for each r_i and D) illustrate how the correlations increase parameter uncertainties. In the joint marginal distribution for h and c_1, two distinct modes of higher probability indicate strongly nonlinear behavior: one indicating a weak negative correlation between the two parameters, while the other (weaker mode) indicates a positive correlation between the two parameters. The net result is the weak positive correlation indicated in Fig. 10.

V. SUMMARY

This paper examined geoacoustic inversion of narrowband ship-noise data collected at a bottom-moored horizontal line array in shallow water. A Bayesian formulation was applied to quantify the information content of ship noise to resolve seabed parameters in terms of rigorous uncertainty distributions. A general problem with ship-noise data can be low SNR and relatively few suitable frequency components for processing. This problem was partially mitigated by combining multiple data segments in the inversions, leading to significantly reduced uncertainties of geoacoustic parameter estimates. Inversions using multiple-segment ship-noise data were compared for three cases: inbound and outbound ship at short range, and outbound ship at long range. The SNR was highest for the short-range outbound data, and decreased with range and with change in the ship orientation presumed due to directional effects (i.e., higher SNR with ship stern toward the array). The short-range outbound ship noise provided the most informative inversion results, and resolved the seabed sound-speed profile (parametrized by an upper layer with a gradient over a homogeneous lower layer). The long-range outbound and short-range inbound data were unable to resolve the seabed sound-speed profile, likely due to the lower SNR and for the long-range data attenuation of higher-order modes. The geoacoustic parameters estimated from the inversion of short-range outbound ship noise agree

well with results from inversion of controlled-source data, and with reference geophysical measurements. The results presented in this paper are for inversion of noise from a relatively quiet ship; results could differ for a different ship or environment, although the method developed here should be applicable to other experiment situations.

[1]M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[2]S. E. Dosso, M. L. Yeremy, 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 seismoacoustic data using genetic algorithms and a posteriori probability distributions," J. Acoust. Soc. Am. **92**, 2770–2783 (1994).

[4]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).

[5]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).

[6]S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).

[7]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 197–207 (2006).

[8]M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).

[9]D. J. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and M. Siderius, "Geoacoustic inversion of tow-ship noise via near-field matched-field processing," IEEE J. Ocean. Eng. **28**, 454–467 (2003).

[10]D. J. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).

[11]L. T. Fialkowski, T. C. Yang, K. Yoo, E. Kim, and D. C. Dacol, "Consistency and reliability of geoacoustic inversions with a horizontal line array," J. Acoust. Soc. Am. **120**, 231–246 (2006).

[12]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).

[13]R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," J. Acoust. Soc. Am. **117**, 626–637 (2005).

[14]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).

[15]N. R. Chapman, R. M. Dizaji, and R. L. Kirlin, "Geoacoustic inversion using broad band ship noise," in *Proceedings of the Fifth European Conference on Underwater Acoustics*, edited by M. E. Zakharia, Lyon, France, 2006, pp. 787–792.

[16]M. Nicholas, J. S. Perkins, G. J. Orris, and L. T. Fialkowski, "Environmental inversion and matched-field tracking with a surface ship and L-shaped receiver array," J. Acoust. Soc. Am. **116**, 2891–2901 (2004).

[17]S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).

[18]P. T. Arveson and D. J. Vendittis, "Radiated noise characteristics of a modern cargo ship," J. Acoust. Soc. Am. **107**, 118–129 (2000).

[19]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).

[20]S. E. Dosso and M. J. Wilmut, "Data uncertainty estimation in matched-field geoacoustic inversion," IEEE J. Ocean. Eng. **31**, 470–479 (2006).

[21]D. R. Del Balzo, C. Feuillade, and M. R. Rowe, "Effects of water-depth mismatch on matched-field localization in shallow water," J. Acoust. Soc. Am. **83**, 2180–2185 (1988).

Classification of live, untethered zooplankton from observations of multiple-angle acoustic scatter

Paul L. D. Roberts[a] and Jules S. Jaffe
Marine Physical Lab, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 4 July 2007; revised 13 May 2008; accepted 14 May 2008)

A broadband, multiple-angle acoustic array was used to classify millimeter to centimeter sized live zooplankton in a laboratory tank. Reflections in the frequency range from 1.5 to 2.5 MHz were recorded from untethered 1–4 mm calanoid copepods and 8–12 mm mysids over an angular range of 0°–47°. A synchronized, coregistered video system recorded animal location and orientation. To highlight differences between animals, a frequency correlation matrix was computed from the observed wide-band power spectra of the scattered sound. Significant differences in the slopes and shapes of the eigenvalue spectra of this matrix were found for mysids versus copepods. These results support the idea that broadband, multiple-angle scatter can be used to classify organisms of different sizes and shapes. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945114]

PACS number(s): 43.30.Sf, 43.30.Ft, 43.60.Fg [KGF] Pages: 796–802

I. INTRODUCTION

Crustacean zooplankton play a major role in the ocean's ecosystem, so it is important to develop noninvasive methods to measure their abundance and behavior. Instruments deployed in the laboratory and field have measured sound scatter from a wide range of animals (McNaught, 1968; Holliday *et al.*, 1989; Wiebe *et al.*, 1990; Stanton *et al.*, 1998a; Lavery *et al.*, 2002; Lavery *et al.*, 2007). In addition, scattering models and classification algorithms have been formulated (Martin *et al.*, 1996; Stanton *et al.*, 1998b; Lavery *et al.*, 2002; Lawson *et al.*, 2006) with the ultimate goals of quantification of animal size and abundance, identification of different taxa (Holliday *et al.*, 1989; Holliday *et al.*, 2003; McGehee *et al.*, 2004), and measurements of *in situ* behavior (De Robertis *et al.*, 2000; Genin *et al.*, 2005).

The fundamental challenge to achieving these goals arises from the vast diversity of zooplankton in the ocean and the confounding influence of size, shape, orientation, and material properties on acoustic scatter (McGehee, 1998; Martin Traykovski *et al.*, 1998; Warren *et al.*, 2002). Variations of these factors lead to substantial ambiguities in using acoustics to both identify and count animals. Reducing these ambiguities, while retaining a system that is practical for fieldwork, would be of great value.

One potential solution is to observe sound that has been reflected at different angles. In the context of diffraction theory, Jaffe (2006) demonstrated that swim-bladder size could be accurately inferred from multiple-angle acoustic reflections observed from a single fish. Roberts and Jaffe (2007) used numerical methods to demonstrate that individual copepods and euphausiids could be classified using multiple-angle, wide-band reflections. The study indicated that the multiple-angle method was more accurate than other techniques using either narrow- or wide-band sound with a single transceiver.

Here, the multiple-angle technique was applied to two types of marine zooplankton: copepods and mysids. Copepods are an order of crustacean zooplankton (typically 1–4 mm in length) distributed throughout the world's marine and fresh waters. Mysids are common coastal inhabitants similar in size, shape, and composition to euphausiids (typically 5–10 mm in length). Copepods and euphausiids are dominant taxa of marine ecosystems and there is great interest in quantifying their distributions with remote, noninvasive methods.

A laboratory scattering apparatus was constructed to record simultaneous reflections from zooplankton at multiple observation angles. To be compatible with available hardware, eight receivers were evenly spaced on a line, forming a 2-m-long array. The length of the array was chosen so that it could eventually be deployed on an autonomous underwater vehicle (AUV) or glider. The even spacing of receivers sampled the available aperture uniformly. No optimization of receiver position or array length was attempted, but the final geometry was consistent with simulated multiple-angle experiments (Roberts and Jaffe, 2007). During recording of acoustic reflections, two video cameras simultaneously recorded the size, position, and orientation of animals. Multiple-angle observations were analyzed using a correlation matrix approach designed to highlight changes in scattering across the array. Eigenvalue analyses of these matrices demonstrate that, in our laboratory setup, multiple-angle acoustic data can be used to accurately discriminate between copepods and mysids.

II. EXPERIMENTAL METHODS AND DATA PROCESSING

A. Acoustic scattering apparatus

The scattering apparatus consisted of a linear array of ten disk transducers (Panametrics, Waltham, MA). Eight were used as receivers and two as transmitters. All receivers and one transmitter were 19 mm diameter, 2.25 MHz broadband transducers (V305-SU). The other transmitter had a di-

[a]Electronic mail: plrobert@ucsd.edu

 0001-4966/2008/124(2)/796/7/$23.00

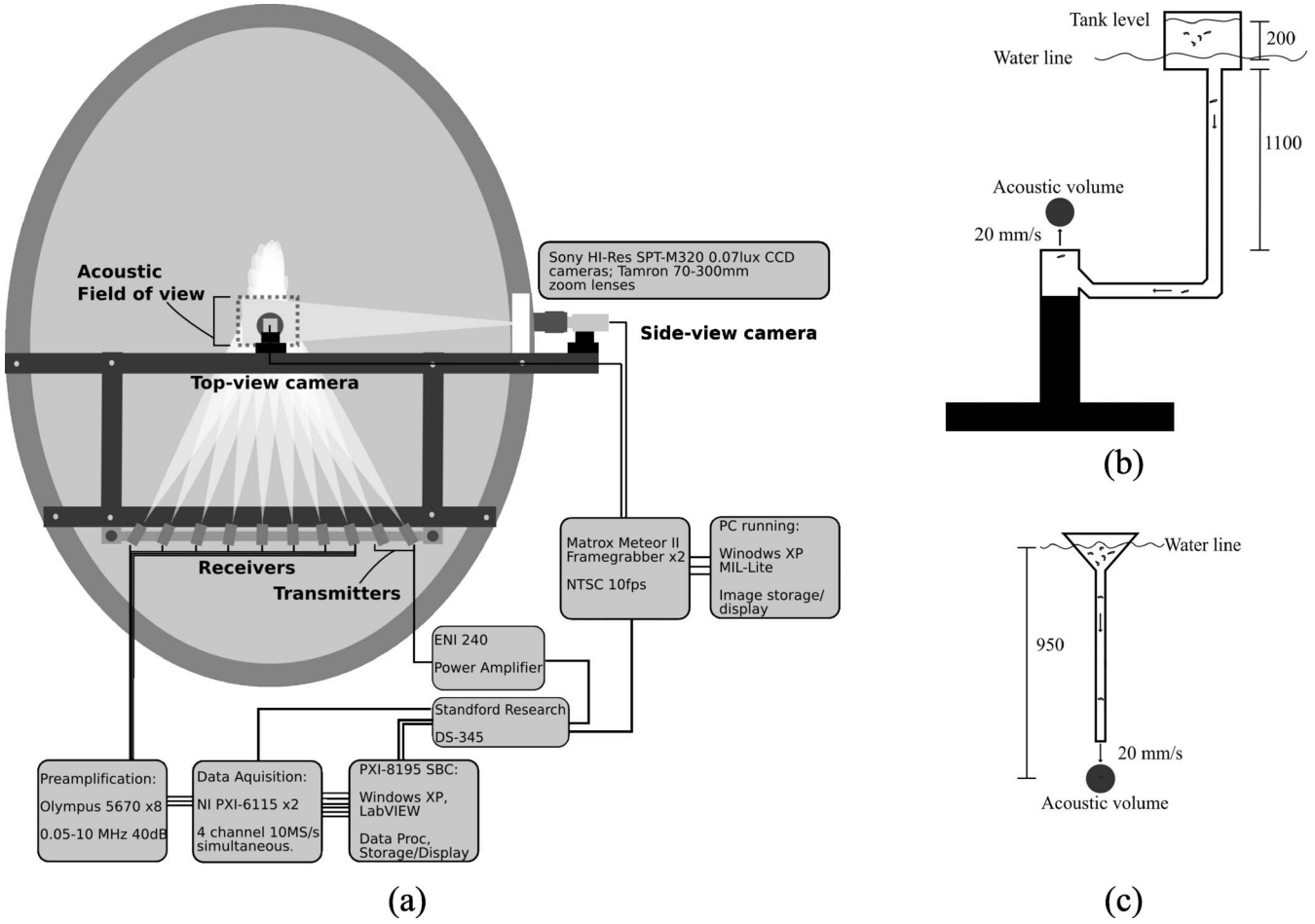


FIG. 1. (a) The multiple-angle scattering apparatus (viewed from above) showing the acoustic and optic elements, Unistrut (Unistrut Corporation) frame, experimental tank, and associated data acquisition components. (b) The bottom-up pump. Animals are drawn out of a tank above the surface and injected below the field of view. (c) The top-down pump. Animals are sedated and allowed to sink through the pipe into the FOV. All distances are in millimeters.

ameter of 38 mm (V395-SU). The larger transmitter was used to increase the source level, improving the signal to noise ratio (SNR) of echoes from smaller animals. The field of view (FOV), defined by the overlap among all acoustic beams, was 42 ml with the 38 mm transmitter and 328 ml with the 19 mm transmitter. Optical images within the FOV were acquired using two, high-sensitivity VGA cameras (Sony SPT-M320). The acoustic array and cameras were rigidly connected by a Unistrut (Unistrut Corporation, Wayne, MI) system to ensure minimal relative movement during experiments [Fig. 1(a)]. A custom-built rail system was used to slide the array in and out of the water. This allowed the array to be kept dry when not being used, and then precisely positioned and locked in place for experiments. All experiments were performed in a 3.0-m-wide, 4.2-m-long, 2.4-m-deep elliptical tank with view ports 1.2 m above the bottom. The tank was filled with chilled, filtered seawater maintained at a temperature of 14.9 °C throughout the experiments.

Acoustic data were acquired by a National Instruments (Austin, TX) PXI-8195 controller running WINDOWS XP (Microsoft, Redmond, WA) with two PXI-6115, 10 MHz, 12 bit, four-channel simultaneous sampling boards with 64 Mbytes of on-board memory. The output from each receiver was fed through a Panametrics 5670 broadband preamplifier prior to digitization. The transmitter was driven by an ENI (Bell Electronics, Kent, WA) AP400B 400 W power amplifier. Waveforms sent to the power amplifier were generated by a Stanford Research (Sunnyvale, CA) DS345 arbitrary waveform generator. Software developed in LABVIEW (National Instruments) controlled the acquisition, recording, and real-time display of data.

A single personal computer (PC) running WINDOWS 2000 controlled the stereo video system. Images were "grabbed" from each camera—at an adjustable rate controlled by the acoustic transmissions—using synchronized Matrox (Matrox, Dorval, Canada) Meteor II frame grabbers. Software developed in C++ used the Matrox Imaging Library to read images from the boards, bundle them together, display them in real-time, and save them to disk.

To obtain high-contrast images of animals, a 200 mW laser was used for illumination (Aixiz 200 mW, 650 nm module). A wavelength of 650 nm was selected as it is almost invisible to the animals yet suffers limited attenuation through the medium. The laser beam was spread with a diverging lens to yield a cone of light that intersected the FOV.

B. Experimental setup

Preliminary experiments revealed that scatter from tethers maintaining zooplankton in the FOV dominated observa-

tions at these high frequencies. Therefore, a substantial challenge was to position live, untethered animals in the FOV. Copepods were pumped from a small holding tank through a system of hoses and out through a 75 mm dimeter pipe positioned directly under the FOV [Fig. 1(b)]. Copepods typically stayed near the FOV for several seconds after exiting the pipe, whereas mysids quickly swam away from the FOV. To mitigate this problem, mysids were sedated by placing them into a dilute (1% by volume) water bath of clove oil and filtered seawater. They were kept in the bath until they ceased swimming (typically 30–60 s) but retained leg movement. Sedated mysids were then transferred immediately to a funnel system [Fig. 1(c)] that guided them into the FOV while they sank slowly. These mysids eventually recovered as inferred by their swimming behavior.

C. System alignment and calibration

Alignment and calibration of the multiple-angle system were more complicated than for a typical monoangle system. Transducers were aligned using a long strand of nylon monofilament suspended perpendicular to the array at a point that was designated the system's origin. The monofilament's diameter of 75 μm resulted in an acoustic scattering pattern that was omnidirectional in the horizontal plane, but narrow in the vertical plane. Pointing angles for all transmitters and receivers were iteratively adjusted to maximize reflections.

Three-dimensional transducer positions were estimated using the time of arrival at each receiver of the echo scattered from the monofilament. With the array elements located on a line, their Euclidean offsets from the origin were determined by computing the mean square error between the observed and predicted arrival times of echoes at each receiver, for all possible offsets. A clear minimum was found [Fig. 2(a)] and the model output, compared to measured data, yielded excellent agreement [Fig. 2(b)]. The array was off center with a total angular span of roughly 46.7° (Fig. 3). The horizontal distance from the center of the array to the origin was 1.14 m and the angular spacing between array elements was on average 5.8° ± 1°.

The video system was aligned with the acoustic system using a small sphere suspended in the middle of the FOV. One camera was mounted so that it looked through the tank's view port (side view). The second camera was mounted in a waterproof housing with a small view port and submerged almost directly above the FOV (top view). Both cameras were mounted to the Unistrut frame using heavy-duty, three-axis telescope mounts (Losmandy DCM2, Los Angeles, CA). These mounts allowed each camera to be rotated until the target was in the center of each frame. They were then locked in place.

Before and after experiments, acoustic calibrations were performed using a 1 mm diameter tungsten carbide sphere, following the calibration procedure outlined by Foote (1990). Echoes were collected while the sphere was translated within the FOV. Since the echo spectrum varied only slightly for displacements of the sphere within the FOV, the beam pattern was assumed to be constant in that region. The echo spectrum from the sphere was then used to convert

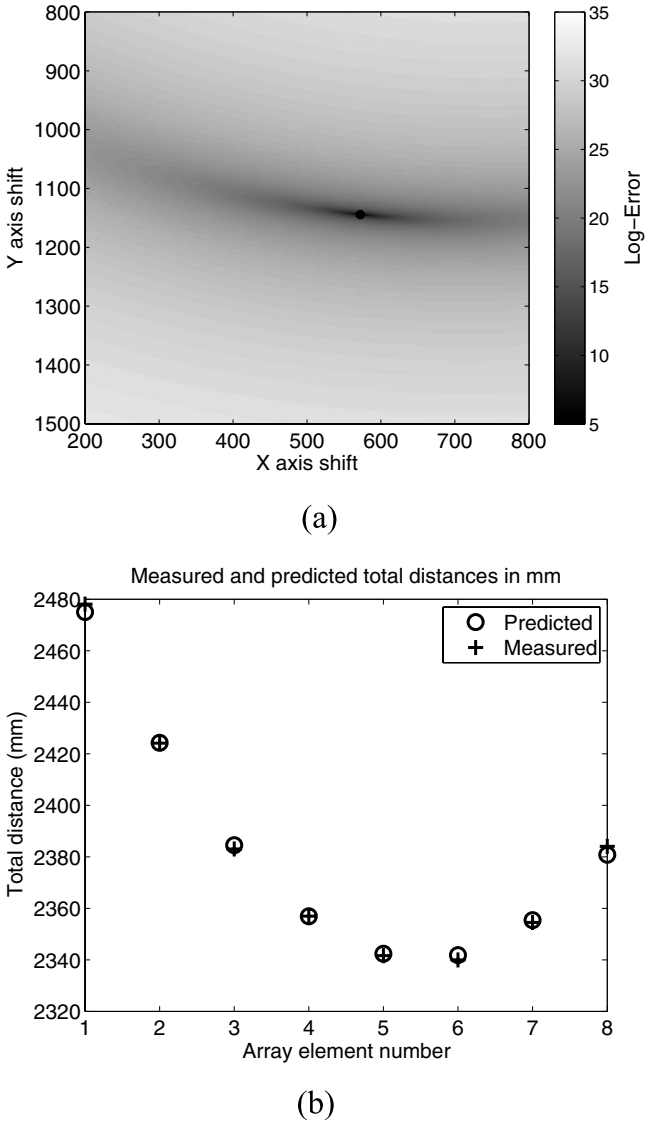


FIG. 2. Calibration results for the array. (a) The error surface for the x and y shift parameters using the calibration data, the inferred sound speed, and the measured element spacing. (b) The modeled and measured total acoustic path distances for each receiver. The total acoustic path is the path from the transmitter, to the origin, and then to the receiver.

recorded echo spectra to target strength. This process was repeated with the nylon monofilament using the scattering model described by Minonzio *et al.* (2005). To check consistency of the calibration, echo spectra—converted to target strength using each calibration method—were compared and found to be within 3 dB. This was adequate for the subsequent analyses as the method only relies on the relative calibration among receivers.

D. Data acquisition and processing

Multiple-angle data were collected during a series of experiments spanning several months. All zooplankton were collected in the La Jolla Cove area (San Diego, CA) by small boat and immediately brought back to the laboratory. Mysids were collected by gently dragging a mesh butterfly net across the kelp at the surface of the kelp forest. This procedure typically resulted in 10–100 mysids ranging from length

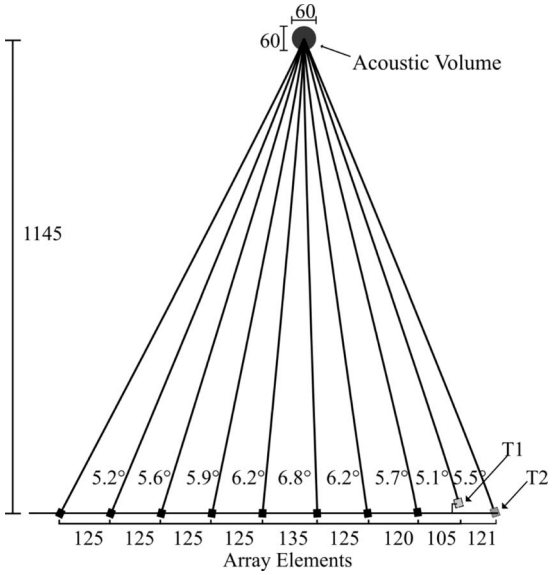


FIG. 3. The geometry of the acoustic array. All distances are in millimeters. The two transmitters are at the right end of the array. Receiver element 1 is at the left end of the array and element 8 is at the right end of the array.

8 to 12 mm. Copepods were collected in a series of net tows using a 250 μm cod end, 1 m diameter net. This yielded several thousand copepods (primarily calanoid) ranging in length from 1 to 4 mm. Animals were allowed to equilibrate with seawater siphoned directly from the FOV for 30 min prior to being injected into the tank. During experiments, between one and ten copepods were injected toward the FOV at a time. Only one mysid was injected at a time. Of those animals injected, roughly 10% actually passed through the FOV.

The transmit signal was a linear frequency-modulated (LFM) chirp from 1.5 to 2.5 MHz with a cosine-squared envelope and a duration of 500 μs. The data-acquisition system recorded ten sequential echoes with 100 ms delay between them. The large tank and relatively small volume of intersection between transmitter and receivers allowed a nearly reverberation-free echo. However, in data from the mysid experiments, there were some small artifacts caused by the injection pipe. These were coherently removed in postprocessing using data recorded from the pipe alone.

Raw acoustic data were matched filtered with a synthetic model of the transmit signal (Chu and Stanton, 1998; Kay, 1998; Warren *et al.*, 2002). The matched-filter output was then windowed to localize the echo from the animal. A window length of 250 time samples (25 μs at 10 MHz sample rate) was selected to capture the longest possible echo for the largest animal insonified. Due to the extended length of the transmit signal, the matched-filter processing gave a SNR improvement of roughly 23 dB over a very short pulse of equivalent power. This processing gain was critical for obtaining good SNR from these weakly scattering animals. The same matched filter was also applied to calibration data. A 4800-point fast Fourier transform (FFT) was used to estimate power spectra of echoes recorded by each receiver. The FFT of each received echo was multiplied by the power spectrum predicted by the calibration model and divided by the power spectrum of the calibration echo. This corrected for the shape of the transmit pulse and the small variation in element sensitivity across the array.

To highlight fundamental differences between echoes from each class of scatterers, a frequency correlation algorithm was developed. Let the M-point FFT of the windowed, matched-filter output for the jth element be $F_j[k]$. Then define the cross correlation between the positive frequency coefficients of the FFTs of two elements a and b as

$$X_{a,b}[m] = \sum_{k=0}^{k=N} F_a^H[((k))_N] F_b[((k-m))_N] \quad \text{for } 1 \le a \le 8 \text{ and } 1 \le b \le 8, \tag{1}$$

where $((x))_N$ denotes $x \bmod N$ and $N=(M+2)/2$. The cross correlation with the maximum magnitude was then used to form an approximately Hermitian, positive-semidefinite matrix

$$G_{a,b} = X_{a,b}[m^*], \tag{2}$$

where

$$m^* = \arg\max_m |X_{a,b}[m]|. \tag{3}$$

As explained in Roberts and Jaffe (2007), the matrix G will be nearly rank 1 if there is an equal correlation between all pairs of receivers. In contrast, if there is a weak correlation between nonidentical pairs of receivers G will be similar to a scaled identity matrix.

To quantify the degree of correlation, the eigenvalue decomposition (Moon and Stirling, 2000) was computed as

$$\Lambda = Q^H G Q, \tag{4}$$

where $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_8)$ is a diagonal matrix of eigenvalues. In practice, there was a substantial amount of correlation among echoes across the array, and the eigenvalues deceased logarithmically. Therefore, the log-eigenvalue spectrum was used for analysis.

III. RESULTS

A. Scattering apparatus and experiment analysis

The transducer array, video system, and data-acquisition hardware worked well for collecting repeatable, multiple-angle scattering measurements when animals were positioned in the FOV. The combination of a rigid frame, with robust rotation mounts for all sensors, allowed simple and straightforward alignment that remained unchanged throughout the experiments. The rail system for translating the array in and out of the water proved invaluable and was constructed using standard, off-the-shelf parts at small cost. The camera system's low resolution, coupled with the interlaced video signal and poor laser-beam characteristics, yielded images of only moderate quality, though they were adequate for this study.

Positioning live specimens in the FOV without causing artifacts in recorded data was the most challenging aspect of these experiments. Numerous methods were evaluated during the development of the system (including a wide variety

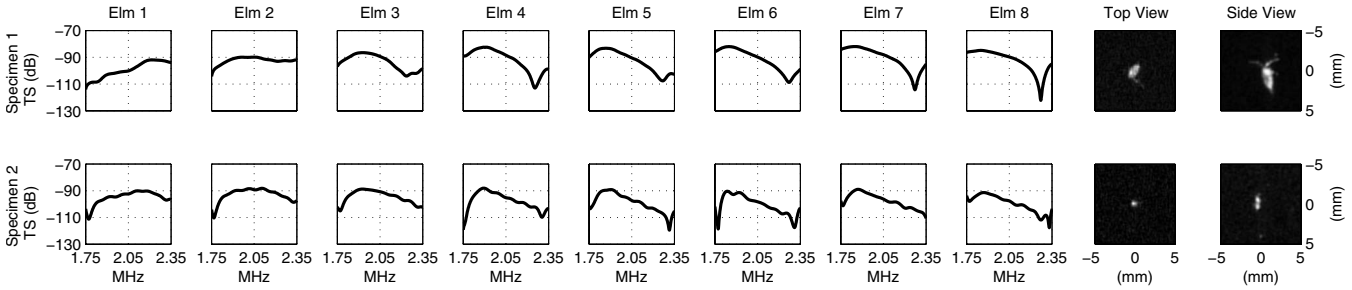


FIG. 4. Example data for two different copepods. Data are plotted for each receiver elements 1–8 as calibrated target strength vs frequency. Images on the far right show top and side views of the copepods at the time of insonification. These images are oriented so that the array is at the bottom of the top-view image with element 1 on the left. For the side-view image, element 1 is oriented into the page, and element 8 is out of the page.

of tethers) but the injection method provided the only artifact-free data. Unfortunately, the injection frequently added bubbles to the acoustic field, or the animal would move out of the FOV before reflections were recorded. Therefore, few artifact-free echoes were obtained during experiments. Data are presented here from eight individual copepods and eight individual mysids. The small size of the data set is solely a consequence of the lack of an efficient means for repeatedly placing untethered, live animals in the FOV.

B. Multiple-angle data analysis

Multiple-angle data require additional processing to characterize target animals. In the first step, spectra of echoes at each angle were computed (Figs. 4 and 5). Despite the small sample size, patterns in target strength data were clear. These highlighted the influence of animal orientation on echo spectra, motivating the frequency correlation processing.

For copepods, target strength curves (Fig. 4) were very similar among receivers. In addition, target strength was slowly varying across frequencies. Video observations indicated that both specimens shown in Fig. 4 were nearly broadside to the array. However, a small tilt can be inferred from a null in power spectra moving from elements 1–8. Multiple-angle data can therefore offer enhanced insight into the animal's orientation.

For mysids, target strength is less similar among receivers than with copepods, and varies more as a function of frequency (Fig. 5). Video data showed that, for the two specimens shown in Fig. 5, orientation in the horizontal plane differed by roughly 90° (Fig. 5, top-view images). Specimen 1 was nearly end on to element 1 and nearly broadside to element 8. Specimen 2 was nearly broadside to element 1 and nearly end on to element 8. Orientations can also be inferred from acoustic data: there is a decrease in spectral complexity as the animal orientation becomes closer to broadside. Near broadside, spectra are smoother with a well-defined null. This null is likely a result of interference between sound reflected from the sides of the mysid's body closest to and farthest from the array element.

To highlight differences among data sets, multiple-angle spectra for each specimen were combined together to form an image. The spectrum for each angle was first normalized by its standard deviation to remove relative differences in the average reflected energy. When plotted as a function of frequency and angle (Fig. 6), this normalized spectral magnitude shows substantial similarity among frequency and angle for copepods [Fig. 6(a)] with more variability for mysids [Fig. 6(b)].

Frequency correlations among spectra of angular data were computed using Eqs. (1) and (2). Log-eigenvalue spectra (Fig. 7) indicate that echoes at each angle are more correlated for copepods than for mysids: the slope of the average spectrum for copepods is roughly twice that for mysids. Furthermore, the average spectrum for copepods decreases steadily, whereas for mysids it is rather flat up to the sixth eigenvalue at which point it steadily decreases.

IV. DISCUSSION AND CONCLUSIONS

Experiments using a multiple-angle acoustic receiver array and live copepods and mysids have shown that it is possible to use the scattered acoustic signal to distinguish between these zooplanktonic taxa. Preliminary experiments

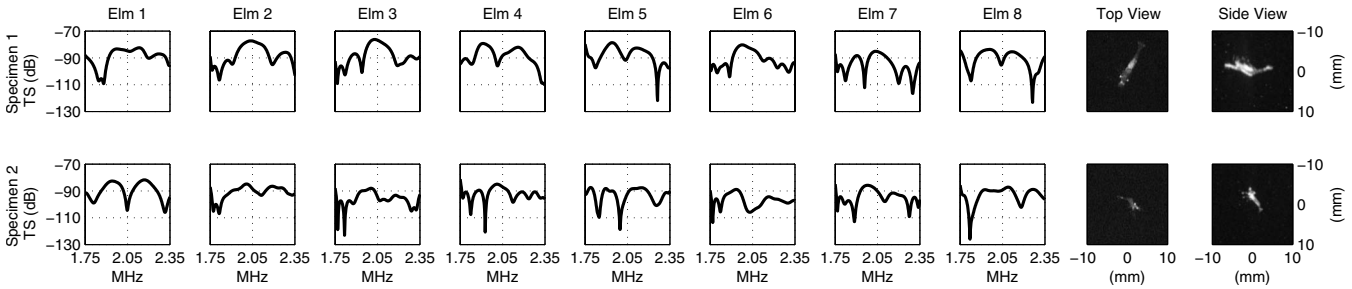


FIG. 5. Example data for two different mysids. Data are plotted for each receiver elements 1–8 as calibrated target strength vs frequency. Images on the far right show top and side views of the copepods at the time of insonification. Images are oriented identical to Fig. 4. Note that the image scale is changed from Fig. 4.

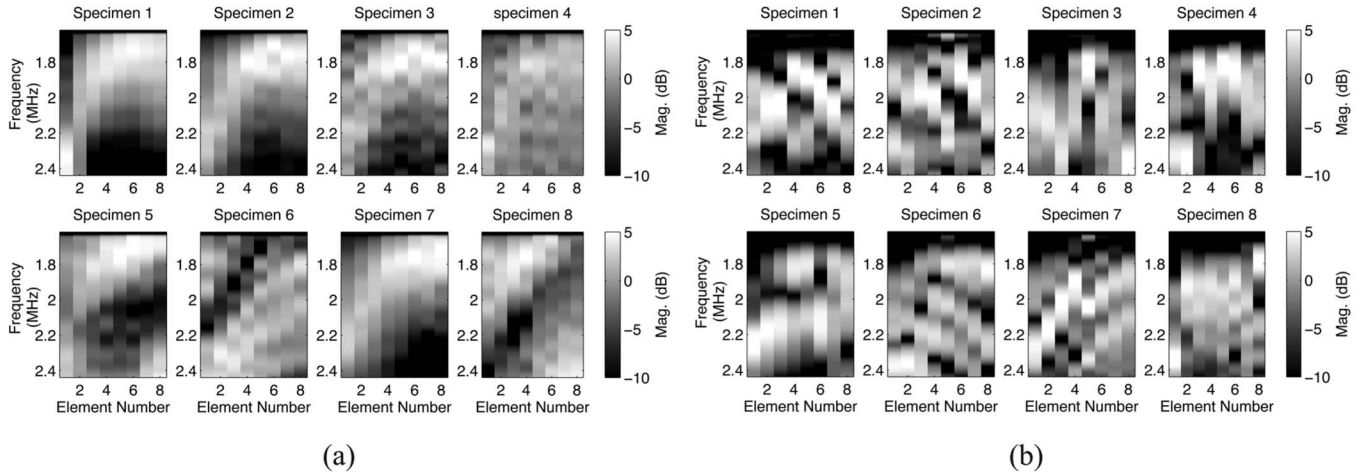


FIG. 6. (a) Normalized magnitude of the scattering spectrum as a function of frequency and array element number for all eight copepod specimens. These data highlight the slowly varying nature of target strength as a function of frequency and look angle. (b) Normalized magnitude of the scattering spectrum as a function of frequency and element number for all eight mysid specimens. These data highlight the relative increase in complexity of target strength as a function of frequency and look angle when compared to (a).

showed that signals from animals tethered in the FOV were dominated by scatter from the tether. Techniques were developed to introduce live, untethered animals into the FOV; however, data quantity was limited by the low success rate in positioning animals in the FOV. In a real pelagic environment these factors would not be an issue.

Multiple-angle data (Figs. 4–6) exemplify a fundamental principle of sound scattering from weak scatterers: the scattered sound field in the immediate vicinity of a target and its radiated pattern in the far field are Fourier transform pairs (Morse and Ingard, 1968). Therefore, variability in target strength is controlled by the size and shape of the scatterer. A thicker scatterer at a given orientation permits more variability over frequency for a fixed bandwidth. Likewise, a more elongate scatterer permits more variability over angle. This can complicate the interpretation of single-angle, wide-band scatter when animal orientation is unknown (Martin Traykovski *et al.*, 1998). A comparison of data from single angles (Fig. 6) reveals that single broadband echoes from copepods and mysids can be quite similar depending on the orientation of the animal relative to the system. However, when multiple angles are considered, this similarity is dramatically reduced.

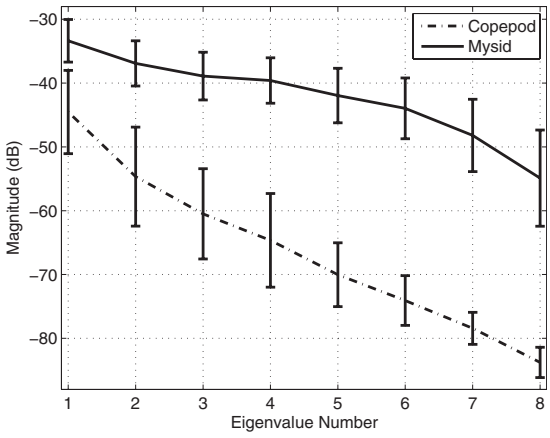


FIG. 7. Average log-eigenvalue spectrum for copepod (dashed) and mysid (solid) data. Error bars denote one standard deviation. The slope of the log-eigenvalue spectrum for copepods is roughly twice that for mysids demonstrating a greater degree of correlation among echoes at each angle for copepods.

A further advantage of the method presented here is that transducers only need to be intercalibrated and not absolutely calibrated. This method is therefore relatively immune to biofouling, so long as the biofouling occurs equally for all array elements. A potential disadvantage is that multiple elements require more sophisticated hardware and computer processing than a single-element system. However, several existing systems use multiple transducer configurations for measuring Doppler shifts to infer currents, so this is not a problem—even in a battery-powered instrument. Furthermore, the computational burden is modest.

While this study has demonstrated the utility of a multiple-angle array for zooplankton identification, there remain several research issues requiring further consideration. One concerns optimizing the array geometry: the number of elements, the distance between them, and the overall length of the array. Another concerns extension of the processing methods, once the data are in hand. The eigenvalue method described here works well in analyzing the presented laboratory data; however, alternate analyses should be explored.

Naturally, the real challenge in zooplankton sensing is to implement the method in the ocean. Beyond the basic issue of putting electronic instruments in a corrosive, high-pressure environment, there is the added problem of the large diversity of animals: the many compositional types and body morphologies make enumeration and identification difficult. One approach might be to combine the use of broadband, multiple-angle scattered sound with target-strength observations of individuals. Such a system would generate information about both the size and reflectivity of the animals, with reduced sensitivity to their absolute orientation. This will greatly reduce the number of candidate animals corresponding to a measured set of reflections, enhancing our ability to discriminate among them. Incorporating these ideas into a sea-going system will increase our knowledge about the ma-

rine planktonic ecosystem and the role of zooplankton in the regulating the dynamics and fluxes through that ecosystem.

ACKNOWLEDGMENTS

The authors would like to thank the SIO machine shop for assistance with construction of the scattering apparatus, Eddie Kisfaludy, Erdem Karakoylu, Fernando Simonet, Ben Maurer, and Robert Glatts for technical consulting on the scattering apparatus design and help with experiments, two anonymous reviewers and Peter Franks for helpful comments on the manuscript, and the California Sea Grant for funding this research.

Chu, D. Z., and Stanton, T. K. (**1998**). "Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton," J. Acoust. Soc. Am. **104**, 39–55.

De Robertis, A., Jaffe, J. S., and Ohman, M. D. (**2000**). "Size-dependent visual predation risk and the timing of vertical migration in zooplankton," Limnol. Oceanogr. **45**, 838–1844.

Foote, K. G. (**1990**). "Spheres for calibrating an 11-frequency acoustic measurement system," ICES J. Mar. Sci. **46**, 284–286.

Genin, A., Jaffe, J. S., Reef, R., Richter, C., and Franks, P. J. S. (**2005**). "Swimming Against the Flow: a Mechanism of Zooplankton Aggregation," Science **308**, 860–862.

Holliday, D. V., Donaghay, P. L., Greenlaw, C. F., McGehee, D. E., McManus, M. M., Sullivan, J. M., and Miksis, J. L. (**2003**). "Advances in defining fine- and micro-scale pattern in marine plankton," Aquat. Living Resour. **16**, 131–136.

Holliday, D. V., Pieper, R. E., and Kleppel, G. S. (**1989**). "Determination of zooplankton size and distribution with multi-frequency acoustic technology," ICES J. Mar. Sci. **46**, 51–62.

Jaffe, J. S. (**2006**). "Using multiple-angle scattered sound to size fish swim bladders," ICES J. Mar. Sci. **63**, 1397–1404.

Kay, S. M. (**1998**). *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice-Hall, Upper Saddle River, NJ), Vol. **2**.

Lavery, A. C., Stanton, T. K., McGehee, D. E., and Chu, D. Z. (**2002**). "Three-dimensional modeling of acoustic bascattering from fluid-like zoolankton," J. Acoust. Soc. Am. **111**, 1197–1210.

Lavery, A. C., Wiebe, P. H., Stanton, T. K., Lawson, G. L., Benfield, M. C., and Copley, N. (**2007**). "Determining dominant scatterers of sound in mixed zooplankton populations," J. Acoust. Soc. Am. **122**, 3304–3326.

Lawson, G. L., Wiebe, P. H., Ashjian, C. J., Chu, D. Z., and Stanton, T. K. (**2006**). "Improved parameterization of Antarctic krill target strength models," J. Acoust. Soc. Am. **119**, 232–242.

Martin, L. V., Stanton, T. K., Wiebe, T. K., and Lynch, J. F. (**1996**). "Acoustic classification of zooplankton," ICES J. Mar. Sci. **53**, 217–224.

Martin Traykovski, L. V., O'Driscoll, R. L., and McGehee, D. E. (**1998**). "Effect of orientation on broadband acoustic scattering of Antarctic Krill Euphuasia superba: Implications for inverting zooplankton spectral signatures for angle of orientation," J. Acoust. Soc. Am. **104**, 2121–2135.

McGehee, D. E., Demer, D. E., and Warren, J. D. (**2004**). "Zooplankton in the Ligurian Sea: Part I. Characterization of their dispersion, relative abundance and environment during summer 1999," J. Plankton Res. **26**, 1409–1418.

McGehee, D. E., O'Driscoll, R. L., and Traykovski, L. V. M. (**1998**). "Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz," Deep-Sea Res., Part II **45**, 1273–1294.

McNaught, D. C. (**1968**). "Acoustical determination of zooplankton distributions," in The 11th Annual Conference on Great Lakes Research, pp. 76–84.

Minonzio, J. G., Prada, C., Chambers, D., Clorennec, D., and Fink, M. (**2005**). "Characterization of subwavelength elastic cylinders with the decomposition of the time-reversal operator: Theory and experimen," J. Acoust. Soc. Am. **117**, 789–798.

Moon, T. K., and Stirling, W. C. (**2000**). *Mathematical Methods and Algorithms for Signal Processing* (Prentice-Hall, Upper Saddle River, NJ).

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ), Chap. 8, pp. 407–418.

Roberts, P. L. D., and Jaffe, J. S. (**2007**). "Multiple angle acoustic classification of zooplankton," J. Acoust. Soc. Am. **121**, 2060–2070.

Stanton, T. K., Chu, D. Z., Wiebe, P. H., Martin, L. V., and Eastwood, R. L. (**1998a**). "Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms," J. Acoust. Soc. Am. **103**, 225–235.

Stanton, T. K., Chu, D. Z., and Wiebe, P. H. (**1998b**). "Sound scattering by several zooplankton groups. II. Scattering models," J. Acoust. Soc. Am. **103**, 236–253.

Warren, J. D., Stanton, T. K., McGehee, D. E., and Chu, D. Z. (**2002**). "Effect of animal orientation on acoustic estimates of zooplankton properties," IEEE J. Ocean. Eng. **28**, 271–282.

Wiebe, P. H., Greene, C. H., Stanton, T. K., and Burczynski, J. (**1990**). "Sound scattering by live zooplankton and micronekton—empirical-studies with a dual-beam acoustical system," J. Acoust. Soc. Am. **88**, 2346–2360.

Acoustic characterization of panel materials under simulated ocean conditions using a parametric array source

Victor F. Humphrey[a)]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

Stephen P. Robinson
National Physical Laboratory, Teddington, Middlesex TW11 0LW, United Kingdom

John D. Smith
DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ, United Kingdom

Michael J. Martin
QinetiQ Ltd., Cody Technology Park, Ively Road, Farnborough, Hampshire GU14 0LX, United Kingdom

Graham A. Beamiss and Gary Hayman
National Physical Laboratory, Teddington, Middlesex TW11 0LW, United Kingdom

Nicholas L. Carroll
QinetiQ Ltd., Winfrith Technology Park, Winfrith Newburgh, Dorchester, Dorset DT2 8XJ, United Kingdom

(Received 30 September 2007; revised 16 May 2008; accepted 16 May 2008)

A technique for evaluating the underwater acoustic performance of panels under simulated ocean conditions in a laboratory test facility is described. The method uses a parametric array as a source of sound within a test vessel capable of simulating ocean depths down to 700 m and water temperatures from 2 to 35 °C. The reflection loss and transmission loss of the test panel may be determined at frequencies from a few kilohertz to 50 kHz. The use of the parametric array enables wideband measurements to be undertaken with short-duration pulses and reduces the effects of diffraction from the panel edges. An acoustic filter is used to truncate the array in order to provide a source-free measurement region and to simplify the measurement process. The difficulties of establishing a parametric array in the confined space of the vessel are outlined, and the experimental procedures adopted are described. The techniques were validated by undertaking measurements on two test objects that have predictable behavior. The potential of the technique is also illustrated with experimental results for test panels for hydrostatic pressures up to 2.8 MPa. An extensive discussion of the measurement limitations is included.
[DOI: 10.1121/1.2945119]

PACS number(s): 43.30.Xm, 43.58.Vb, 43.30.Ky, 43.30.Lz [KGF] Pages: 803–814

I. INTRODUCTION

Various elastomeric materials are used in underwater acoustics for encapsulation of transducers, housing of arrays, coating of structures, and lining of test tanks (Bobber, 1988). The acoustic properties of these materials can vary dramatically with frequency, temperature, and pressure (depth) (Capps *et al.*, 1981). This is especially true when using a viscoelastic material close to its glass-rubber transition (Ferry, 1980). Consequently, the accurate characterization of the acoustic properties of such materials is of crucial importance when selecting materials and assessing their performance.

Techniques exist to determine properties such as the reflection loss (the reflection coefficient in decibels) and transmission loss (transmission coefficient in decibels) from measurements made on panels of the material under test (Bobber, 1988; Rudgers and Solvold, 1984; Mikeska and Behrens, 1976). In the most common configuration for such measurements, a panel of the material is ensonified with the acoustic field produced by a conventional linear source transducer, and the reflected and transmitted fields are sampled with hydrophones positioned just in front of and just behind the panel. Either discrete frequency bursts or broadband pulses may be used to drive the source transducer (Thibieroz and Giangreco, 1991).

Panel testing suffers from problems that arise due to the finite size of the panel under test. In particular, diffraction of the acoustic field from the edges of the panel leads to signals that arrive at the hydrophones at times only marginally different from those of the reflected and transmitted signals. Consequently, it is difficult to separate the diffracted signal from the reflected (or transmitted) signal; this limits the accuracy of the technique at low frequencies. A number of methods that attempt to overcome this limitation have been reported, such as the use of novel signal processing techniques (Trivett and Robinson, 1981; Piquette, 1987; Piquette 1996). In addition, attempts have been made to reduce the

[a)]Electronic mail: vh@isvr.soton.ac.uk.

effects of diffraction or to quantify them in order to derive corrections (Piquette, 1986; Piquette, 1991; Piquette, 1994). The use of hydrophones or sensors placed on the surface of the test panel has also been reported (Audoly and Giangreco, 1990). The effect of diffraction is minimized if the hydrophone is placed close to the panel surface, but in the case of the reflection measurement, it then becomes difficult to resolve the incident and reflected signal. This has led to methods where only the transmitted signal is measured, the other properties being derived from this measurement (Audoly, 1992).

The panel measurements described above are typically undertaken in open laboratory tanks. Their use limits the range of testing that is possible since such tanks cannot generally provide the range of environmental conditions that exist during deployment in the sea. As an alternative, full sea trials can be undertaken, but these are often prohibitively expensive and are best used for final testing of complete systems. In any case, although an increased depth may be achieved, sea trials offer little control of temperature. This increases the motivation to make maximum use of laboratory experiments for an acoustic characterization of materials. The use of some laboratory tanks that offer environmental control of temperature and hydrostatic pressure has been reported in the literature (Hagelberg and Corsaro, 1985; Piquette and Paolero, 2003; Esward *et al.*, 2002; Humphrey *et al.*, 2003).

In this paper, a description is given of a method of characterizing panel materials which makes use of a parametric array as the acoustic source (Humphrey, 1985; Humphrey and Berktay 1985). This has the advantage that a relatively narrow beam may be generated at low frequencies, which helps to minimize unwanted signals due to both edge diffraction and boundary reflections, and provides short-duration pulses that assist in resolving the desired signals in the time domain. To enable the properties of the material under test to be determined as a function of temperature and depth, the method has been implemented in the Acoustic Pressure Vessel (APV) facility at the UK National Physical Laboratory (NPL) where the environmental conditions can be modified.

In the next section of the paper, a description is given of the operation of the parametric array. This is followed by a report of the experimental procedure, including a description of the theoretical and practical difficulties and how these were addressed. For the purposes of validation, a number of test objects were designed to exercise the method. These test objects are described, and the results of their characterization are presented. Finally, a discussion is given of the limitations of the method.

II. THE PARAMETRIC ARRAY

A. Array characteristics

A parametric array uses the nonlinear propagation of a primary wave field to generate additional lower-frequency (secondary) components that are then used to make measurements. The principle of a parametric array was first proposed by Westervelt in 1963 and has since been used in a range of sonar applications. An important fundamental characteristic

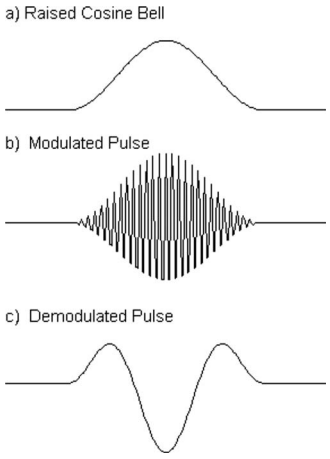


FIG. 1. Drive wave forms for the parametric array showing the ideal form of the demodulated signal.

of an ideal parametric array is its relatively narrow acoustic beam without sidelobes. In practice the sidelobe levels will be extremely low unless the array is truncated, as was done here. Even with truncation the narrow beam characteristics at low frequencies make a parametric array a useful source in confined laboratory conditions (Humphrey, 2002).

A succinct account of the operation of parametric arrays can be found in Hamilton (Hamilton, 1998). The principle of operation requires the nonlinear interaction of two coaxial high amplitude primary waves, resulting in the generation of secondary waves that appear to come from "pseudosources" distributed throughout the interaction region of the primaries. The distribution of these pseudosources, and their phases, forms an end-fire array with a characteristically narrow beam. The beam width is determined by the length of the interaction region rather than the size of the source transducer. For example, to produce a single frequency secondary signal, two single frequencies f_1 and f_2 may be transmitted from the source transducer, with the parametric array used to generate a difference frequency beam at a frequency f_2-f_1.

Although it is possible to operate the parametric array using two primary single frequencies (as described above), the primary transducer can also be driven with a short pulse consisting of a modulated carrier frequency. The drive signals used for the work described here are shown in Fig. 1. A 300 kHz carrier is amplitude modulated with a raised cosine bell envelope $e(t)$ where

$$e(t) = \frac{(1+\cos 2\pi f_e t)}{2}, \quad -\frac{1}{2f_e} < t < \frac{1}{2f_e}, \tag{1}$$

and f_e is the envelope frequency. In this case the low-frequency secondary wave form generated on axis can be shown to be proportional to the second derivative, with respect to time, of the square of the transmitted pulse envelope (Berktay, 1965; Humphrey, 1985). The generated wave form shape and spectrum can easily be adjusted by altering the length of the envelope function by changing f_e.

For the work reported here envelope frequencies of 20, 10, 5, and 3 kHz were used. The spectra of the resulting acoustic signals in the secondary beam are shown in Fig. 2. These illustrate the range of signal frequencies possible and

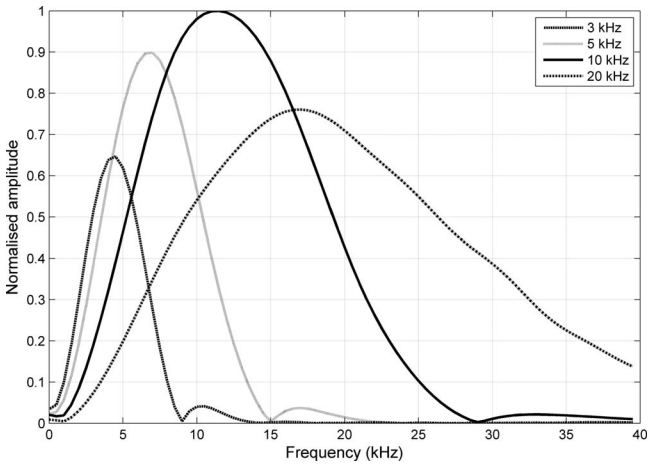


FIG. 2. Difference frequency spectra obtained using raised cosine bell modulated pulses with envelope frequencies (f_e) of 20, 10, 5, and 3 kHz. The amplitudes are normalized to the maximum of the 10 kHz spectra.

the extent to which the spectra can be controlled to investigate the panel performance over different frequency ranges. For example, the raised cosine bell envelope based on a 10 kHz sinusoid results in a secondary signal that has a −6 dB bandwidth extending from 5 to 18 kHz. In general, measurements are made using a range of pulses with different center frequencies to ensure that the signal-to-noise ratio is satisfactory over the entire frequency range of interest. The performance of the array for higher-frequency pulses is limited by the bandwidth of the transmitting transducer.

The parametric array has been used as the acoustic source for the work described here because it can produce a relatively narrow beam at low kilohertz frequencies. This helps to minimize both diffraction from the edges of the material sample and unwanted reflections from the boundaries of the measurement tank. The measured beam profiles of the secondary frequency signals are shown in Fig 3. The data show that for 10 kHz, the amplitude will be about 4 dB down at the edge of an 800 mm wide test panel compared to the amplitude at the panel center. For higher frequencies, the

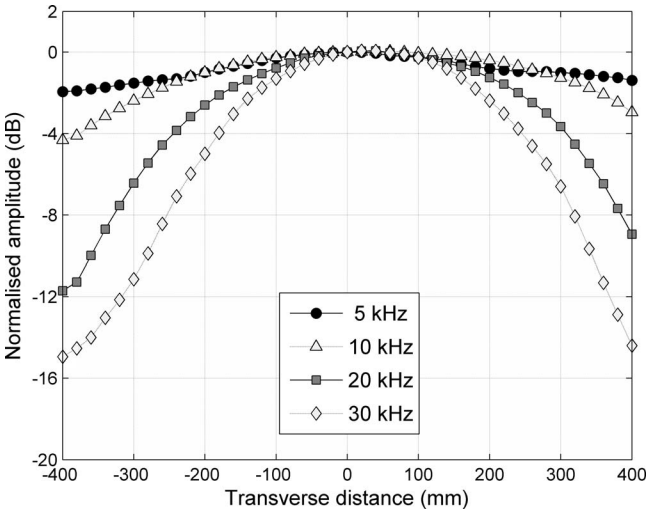


FIG. 3. Difference frequency beam profiles for frequencies of 5, 10, 20, and 30 kHz at a range of 2.75 m from the transducer for an array truncated at 1.88 m from the transducer.

reduction is even greater. At frequencies lower than 10 kHz, the benefit is not so significant. However, another feature of the parametric array is that the broadband acoustic pulses generated have a short duration. This improves the ability to discriminate between the desired signals and unwanted signals by appropriate windowing of the wave form in the time domain.

B. Acoustic filter

One of the principal difficulties of using a parametric array for such laboratory measurements is ensuring that other nonlinear effects do not complicate or invalidate the measurement process. Hence, it is important to check that effects such as hydrophone nonlinearity (Humphrey and Hsu, 1980; Moffet and Henriquez, 1982) or nonlinearity of the receiving electronics are not important (Humphrey, 1985; Humphrey and Berktay 1985b). It is also important to ensure that the required measurement region can be isolated from nonlinear propagation effects. These issues are especially important for parametric array measurements in finite-sized tanks where the ranges available mean that the primary beam is not significantly attenuated by absorption and spreading losses. As a result, the primary levels can still be relatively high and other nonlinear effects potentially significant. These difficulties can be safely avoided by truncating the interaction region using an acoustic low pass filter. This is formed from a material that transmits the secondary waves but attenuates the higher-frequency primary waves. The filter may be considered as a baffle that delineates one end of the end-fire array source region. This arrangement produces a source-free region beyond the filter in which measurements may be easily performed without any of the above complications (Humphrey, 1988).

The acoustic filter used in this work is made from a 30 mm thick polyurethane panel. The polyurethane material is filled with gas-filled microspheres and is designed to strongly attenuate the 300 kHz primary waves and to transmit the secondary waves with little attenuation.

During the initial measurements using the system, the filter performance was found to exhibit hysteresis after the removal of an applied hydrostatic pressure, with the properties taking time to return to their original values. This is thought to be caused by the behavior of the gas-filled microspheres within the filter material. This effect is not important for measurements at ambient pressure (essentially equivalent here to atmospheric pressure). However, when measurements are attempted at elevated hydrostatic pressure (to simulate depth), the hysteresis could potentially lead to differences in the amplitude of the secondary wave transmitted through the filter on successive hydrostatic pressure cycles. Since measurements are typically required on two pressure cycles, with and without the panel, the hysteresis in the filter performance could lead to errors in the calculation of transmission loss or reflection loss.

To study this effect, the amplitude of the first peak of the secondary signal was monitored as the hydrostatic pressure was steadily increased and then decreased. Figure 4 shows the results of these measurements for pulse center frequen-

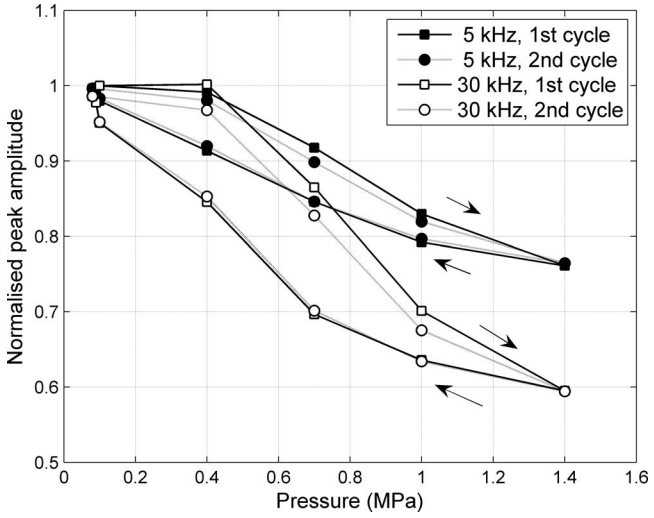


FIG. 4. Variation of generated difference frequency pulse amplitude as a function of hydrostatic pressure through two hydrostatic pressure cycles at 26 °C for pulse center frequencies of 5 and 30 kHz. The sequence of results moves clockwise around the hysteresis loops as the pressure is varied, with the two upper lines representing the rising pressure phase and the two lower lines representing the falling pressure phase.

cies of 5 and 30 kHz. In each case, two pressure cycles are shown, with the measured values proceeding clockwise around the hysteresis loops as the pressure is first increased and then decreased. As can be seen, the amplitude of the signal passing through the filter decreases as the pressure is applied. On reducing the pressure the signal level increases but does not follow the curve for increasing pressure, resulting in a net reduction in output on returning to ambient pressure. As a result, when the pressure is increased on the second pressurization cycle, the signal amplitude does not follow the first cycle. However, it is clear that the pressure amplitudes are the same on the descending part of the pressure cycles, and it was also observed that after the first pressure cycle the performance of the filter stabilized, taking some hours to relax back to its initial state after first being pressurized.

An experimental procedure was therefore adopted to minimize the effect of the above behavior. This included (i) always precycling the hydrostatic pressure in the vessel at the start of each day's measurements to precondition the filter and (ii) keeping the same pressure history and (as much as possible) the same timing of the pressure cycles for each pair of measurements in a set (including approaching the required pressure from the same direction in the cycle).

III. EXPERIMENTAL IMPLEMENTATION

A. Experimental configuration in the test tank

The APV facility utilized for this work consists of a cylindrical steel tank of external dimensions 7.6 m long by 2.5 m in diameter. The water temperature can be controlled in the range of 2–35 °C, and the hydrostatic pressure can be controlled from atmospheric pressure to 6.9 MPa (equivalent to a depth of approximately 680 m) (Preston and Robinson, 1998; Beamiss *et al.*, 2001). A programable logic controller under computer control provides automatic control of the

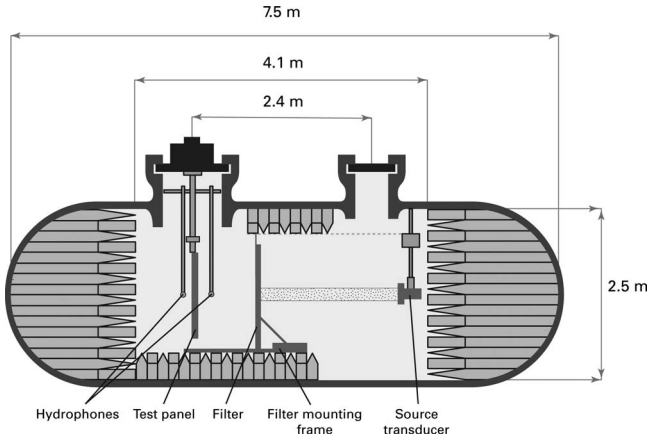


FIG. 5. Schematic diagram of the parametric array in APV showing the location of the transducer, acoustic filter, test panel, and hydrophones (all dimensions in meters). Note that both receiving hydrophones are depicted, although only one is used at a time. The diagram is drawn to scale.

facility. There are two access ports, the centers of which are 2.4 m apart. The larger port is 0.84 m in diameter, and the smaller port is 0.50 m in diameter. The inside of the vessel is lined with an absorbent lining (Darner, 1954), leaving a free internal volume 4.1 m long and 1.9 m in diameter. A schematic diagram showing the arrangement within the APV during panel measurements is provided in Fig. 5.

For the work described here, the large port was used to provide access for the panel under test, which was typically 0.8 m wide and 0.9 m tall. In the measurement configuration, the source transducer, the acoustic filter, the panel under test, and two hydrophones were mechanically supported so that their centers were aligned with the major axis of the vessel. The panel under test was supported by the central shaft extending from the large port lid (maximum load of 500 kg). The measurement hydrophones were small spherical hydrophones with resonance frequency of 95 kHz. They were mounted at the end of free-flooding mounting poles suspended from two horizontal slotted beams that were also attached to the central mounting shaft. The slots allowed the hydrophone-to-panel separation to be varied from zero to a maximum of 400 mm from the panel surface. In addition, with the configuration used it was possible to insert or remove the test panel without disturbing the hydrophones. Note that in Fig. 5, both hydrophones are shown in position. However, to avoid reflected signals from one hydrophone being received by the other, only one hydrophone was used at a time. Depending on the measurement to be made, the hydrophone was mounted either in front of or behind the test panel. The source transducer was mounted below the small port. For the measurements reported here the separation between the source transducer and the panel under test was 2.75 m. The acoustic filter was located 1.88 m from the transducer.

The most significant engineering challenge in implementing the measurement system in the vessel was presented by the acoustic filter. The aim was to utilize as much of the available cross section of the vessel as possible, while still allowing the filter's position to be changed if required. In addition, the mount for the filter had to be of light but robust

construction, which could be quickly erected while the vessel was drained of water. The solution was a hexagonal space frame made from extruded aluminium components from the MB Building Kit System range manufactured by Industrietechnik und Maschinenbau GmbH. The frame was partly assembled outside the pressure vessel, with the final assembly being undertaken inside the vessel. The hexagonal filter support is 1.7 m from one angle to the diametrically opposite angle, and the base is 2.0 m long. The combined weight of the hexagon and base is 80 kg and the complete filter panel assembly weighs approximately 150 kg. To overcome the manual handling problems associated with working in the confined space of the vessel, the filter is built up from three separate contiguous panels (Esward *et al.*, 2002; Humphrey *et al.*, 2003).

B. Experimental procedure

The source transducer used for the parametric array was a 300 kHz square single element piezoelectric piston with pressure compensation designed by the University of Birmingham, UK. This transducer was driven by a modulated 300 kHz signal. The modulation frequencies of the raised cosine bell envelopes used for the work reported here were 20, 10, 5, and 3 kHz. The drive signals were generated in software and downloaded from a computer to the signal source, an Agilent HP33120A arbitrary wave form generator. The output of the signal source was amplified using a 6 kW power amplifier manufactured by Cooknell Electronics Ltd., UK. Typical peak drive voltage amplitudes were 1 kV, and the pulse repetition rates between 5 and 10 Hz were used.

The hydrophone signals were amplified using a Stanford Research SR560 low noise voltage preamplifier and digitized using an Agilent HP89410A vector signal analyzer at a sampling rate of 1 MHz. Coherent averaging was used onboard the signal analyzer to improve the signal-to-noise ratio, with typically 50 repeated acquisitions being averaged. The wave forms were windowed in the time domain to include the desired acoustic pulses and exclude unwanted diffracted and reflected signals. A Tukey window was applied for which the uniform region extended to 90% of the length of the windowed signal, the two outer portions (each of 5%) being shaded with a cosine function. Fast Fourier transforms were performed on the windowed signals with zero padding used to extend the wave form such that the record length was a power of 2 (1024 or 2048 points). The resulting spectra for the incident and reflected (or transmitted) signals were used to calculate the loss as a function of frequency. The reflection loss, R, and transmission loss, T, were derived from the incident, reflected, and transmitted pressure waves (p_i,p_r, and p_t) by

$$R(f) = -20\log_{10}\left(\left|\frac{p_r(f)}{p_i(f)}\right|\right), \tag{2}$$

$$T(f) = -20\log_{10}\left(\left|\frac{p_t(f)}{p_i(f)}\right|\right), \tag{3}$$

where f is frequency.

For transmission loss measurements, two hydrostatic pressure cycle measurement runs were performed. On the first run, the transmitted signal was measured with the panel in place and the hydrophone behind the panel. On the second, a measurement was made of the signal, with the panel removed. In removing the test panel, care was taken not to alter the position of the receiving hydrophone. For the measurement of the transmitted signal, the hydrophone was placed close to the panel to maximize the time delay for diffracted signals originating from the panel edge when compared to the transmitted signals traveling through the panel. Typical values of hydrophone-panel separation were between 20 and 65 mm. In addition, for elastomeric panels attached to steel mounting plates, to reduce the size of the diffracted signal from the panel's rear edges, the panels were reversed such that the softer material of the panel was facing the rear (so that the incident sound wave now passed through the metal plate first). This also eliminated the possibility of the hydrophone sensing reradiation of sound into the water by any Lamb waves excited in the metal backing plate (Humphrey, 1988; Cao *et al.*, 1995; Hurrell, 2002). This panel reversal has no effect on the magnitude of the transmission loss, which is independent of the direction of the acoustic wave through the panel (Rudgers and Solvold, 1984; Piquette, 1991).

For reflection loss measurements, the receiving hydrophone is placed in front of the panel under test. In this case, the hydrophone may in principle be used to record both the incident and reflected wave in the same wave form. This approach may be used when the pulses are short enough to resolve the incident and reflected signals without overlap. This applies when a relatively high pulse modulation frequency is used. However, at lower kilohertz frequencies and for panels with several internal multipath signals, this approach is not possible since the pulses are of longer duration and in general will overlap in time. Therefore, as for the transmission loss measurements, two measurement runs were generally performed. The first run was with the test panel in place, and the second run was a reference measurement, with the panel removed from the vessel but with the hydrophone remaining undisturbed. The two wave forms obtained were then subtracted to obtain the wave form corresponding to the reflected wave form alone. Figure 6 shows examples of wave forms obtained during reflection loss measurements with and without the panel, and the result of subtracting the two wave forms.

The adopted subtraction procedure has the potential to introduce errors since the hydrophone is required to return to the same position in the acoustic field when the panel is removed. This is not a trivial requirement because in order to remove the test panel, the large lid for the vessel must be unlocked, raised, and then lowered again (now without the panel in place). If the test panel is heavy (some of those used here weighed between 70 and 100 kg), the stress introduced to the rigging by the presence of the panel must not be allowed to alter the hydrophone position. In general, it was observed that although the repeatability of hydrophone positioning was good, it was not perfect and a small time delay could be introduced between the two wave forms recorded

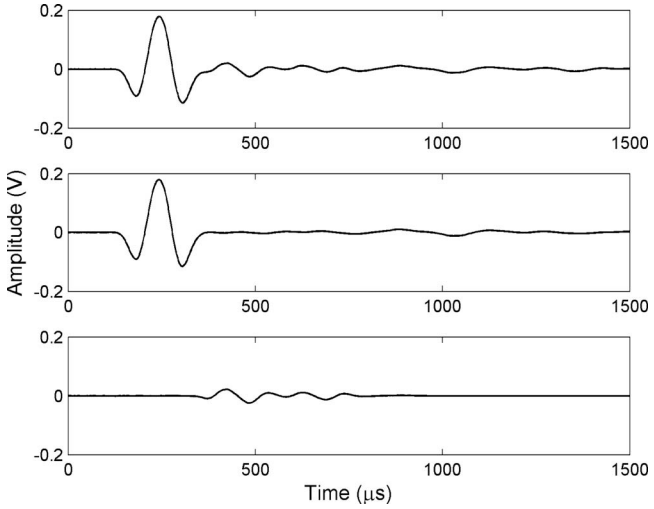


FIG. 6. Examples of received hydrophone signals for a measurement of reflection loss. The wave forms shown were recorded for a pulse modulation frequency of 5 kHz. Top: signal with test panel present; center: reference signal with test panel absent; bottom: the signal obtained by subtracting the two signals, showing the reflected signal only.

by the hydrophone with and without the panel in place. This time delay was typically less than one-half of the wave form sampling interval (less than 0.5 μs) and was removed by applying an appropriate time shift to one of the wave forms so that the subtraction process removed the start of the incident signal as completely as possible. This is equivalent to a positioning error of approximately 0.75 mm. In addition, a small amplitude correction factor was applied, typically less than $\pm 1\%$.

In choosing the hydrophone-panel separation for reflection loss measurements, there is a tradeoff between being close to the panel surface to provide better isolation of diffracted and stray reflected signals and being further away to allow enough time to resolve the incident signal and reflected signal from the panel without excessive overlap. For the results presented here, typical separations of 125–250 mm were used.

It should also be noted that for reflection loss measurements, the reflected wave has traveled further than the incident wave by twice the separation between the receiving hydrophone and the test panel. If the incident wave is not a plane wave, it is necessary to correct the measured reflection loss for the reduction in signal amplitude due to spreading losses (Humphrey, 1985; Humphrey *et al.*, 2003).

An additional problem that was observed was that small air bubbles would sometimes adhere to the hydrophone or its mount. Although these bubbles would only cause small scattered signals, when testing high loss materials the scattered signals could be of the same order of magnitude as the reflected or transmitted signals and so could cause significant error in the results. To minimize this problem, the hydrophone and its rigging were thoroughly cleaned with dilute detergent before measurements were started. Additionally, during the lifting of the large lid to insert or remove a test panel, a water filled container was used to keep the hydrophone immersed throughout, with the container hung on wire from the slotted beams used to support the hydrophone mounting poles.

TABLE I. Test objects used to exercise the method.

Designation	Composition
Test object 1	Twin steel sheets, thicknesses of 1.46 and 0.90 mm, separation of 100 mm, 800 mm wide, 900 mm tall.
Test object 2	Twin PMMA sheets, thicknesses of 5.98 and 3.66 mm, separation of 150 mm, 800 mm wide, 900 mm tall.
Test panel 1	Expancel filled polyurethane, thickness of 50 mm, backed with 5 mm aluminium plate, 800 mm wide, 900 mm tall, lowest RL peak at 7 kHz at 26 °C.
Test panel 2	Expancel filled polyurethane, dual layer, thickness of 125 mm, backed with 5 mm steel, 800 mm wide, 900 mm tall, lowest RL peak at 2.0 kHz at 8 °C.

To undertake measurements as a function of hydrostatic pressure, the vessel was first stabilized at the required measurement temperature. Before measurements, a preliminary cycle of pressurization up to the maximum test pressure was performed in order to pressure cycle the acoustic filter (required for the reasons described in Sec. II). With the test panel in place, the hydrostatic pressure was then varied through the required pressure sequence, and measurements made using the receiving hydrophone. The lid was then lifted and the test panel removed, before a repeat set of reference measurements was taken (without the panel present) at each of the required pressures. Care was taken to reproduce the same pressure cycle for each measurement set as closely as possible. For measurements at different temperatures, the water temperature in the vessel was heated or cooled overnight and thoroughly mixed using the vessel circulation pump. During this period, the test panel was soaked at the correct temperature in a separate soak tank, which also has the facility for temperature control.

IV. TEST SAMPLES

The methods described were used to determine the reflection loss and transmission loss of four test samples. These are summarized in Table I. Each test object was 800 mm wide, the width being limited by the large port diameter. Although it is possible to accommodate test objects that are 1400 mm tall within the vessel, each of the test objects used here was limited to 900 mm tall.

Uncertainties in material parameters mean that viscoelastic materials do not make ideal reference samples for validating the performance of a measurement system. For this reason two test objects were conceived, using engineering materials and a simple design, which mimic the acoustic properties of real test materials but have well defined properties at low frequencies. The two test objects were each formed from two parallel thin rectangular plates separated by threaded studding located in each of the four corners of the plates. The plates were not enclosed so that water filled the

gap between them. Test object 1 consisted of two thin steel plates of thicknesses 1.46 and 0.90 mm with a separation of 100 mm. Test object 2 consisted of two thin plates made of polymethylmethacrylate (PMMA). These were of thicknesses 5.98 and 3.66 mm, with a separation of 150 mm. At low kilohertz frequencies, these test objects exhibited a high reflection loss, but since they had little absorption, their transmission loss at these frequencies was negligible (almost all the incident sound passing through the object). For example, at 10 kHz the steel and PMMA plates had transmission loss values of 0.5 and 0.2 dB, respectively. Therefore they were only used to test the measurement of reflection loss. Since the test objects were of very simple construction and were made from materials that had fairly accurately known properties, they could be described by a simple model that takes account of the material properties and the reflection from each interface (including multiple reflections) (Brekhovskikh, 1960). This enabled predictions to be made of their reflection loss as a function of frequency with reasonable accuracy for both the absolute amplitude and the frequency response.

In order to provide test samples that exhibited a measurable transmission loss as well as reflection loss, two test panels were constructed from elastomeric absorbing materials consisting of polyurethane layers containing gas-filled microspheres. The absorbing layers were mounted on metal backing plates for support, a commonly used method of mounting such panels (Rudgers and Solvold, 1984). In addition to substantial transmission loss, the reflection losses of these panels exhibit peaks at low kilohertz frequencies due to thickness mode resonances within the panels (which depending on the design can be either half-wave or quarter-wave resonances). Their performances also varied with hydrostatic pressure, once again providing a good test for the measurement method.

A theoretical model describing the response of such absorbing test panels has already been developed; it uses an effective media theory based on a multiple scattering formulation (Watermann and Truell, 1961; Varadan *et al.* 1985), together with the dynamic mechanical properties of the polymer matrix, to calculate the effective properties of a foam or microsphere filled material (Baird *et al.*, 1999). These properties are then used in a transmission matrix formulation (Brekhovskikh, 1960), allowing panels to be designed with the desired features in transmission and reflection. Assumptions and approximations made within the model, for example, regarding some of the material properties, result in potential inaccuracies in the predictions of the overall loss values, but the results produced are an extremely good indicator of performance trends. Test panel 1 was made from a single 50 mm polyurethane layer containing gas-filled microspheres (Expancel®) backed with a 4 mm aluminium plate. This panel was designed to exhibit a peak in reflection loss at approximately 7 kHz (at 26 °C). Test panel 2 was 125 mm thick and made from a twin-layer acoustic absorber also containing gas-filled microspheres backed with a 5 mm steel plate. This panel was designed to exhibit a peak in reflection loss at approximately 2 kHz (at 8 °C). These test panels were designed and fabricated by QinetiQ Ltd., UK.

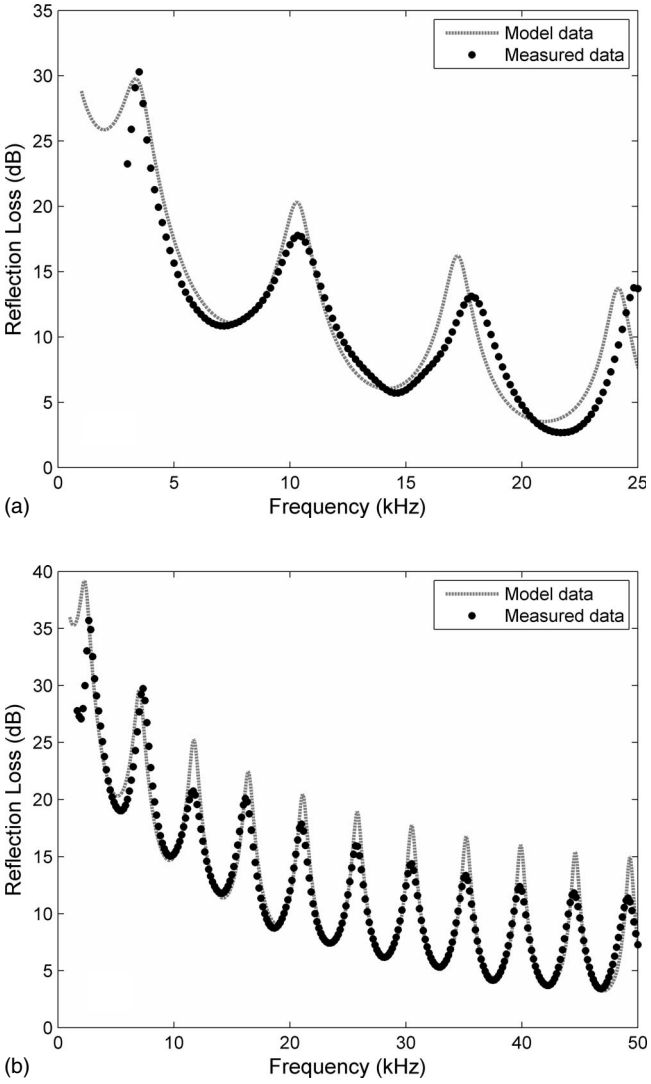
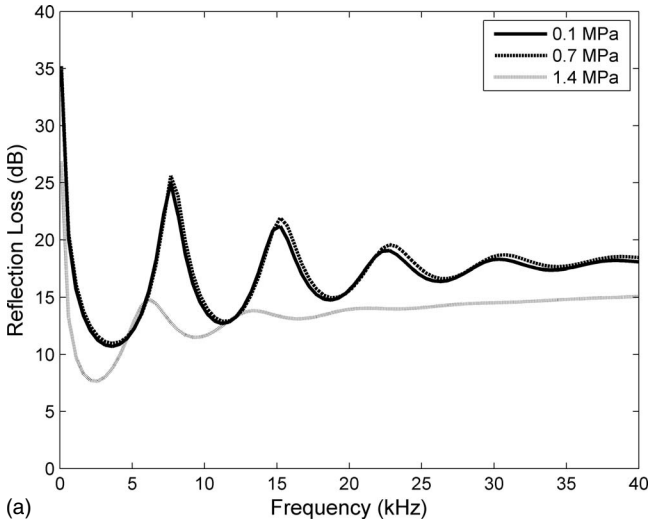


FIG. 7. Results obtained for reflection loss measurements for test objects 1 and 2 at ambient pressure and 8 °C. (a) Results for test object 1 (steel plates separated by 100 mm) using f_e=10 kHz and (b) results for test object 2 (PMMA plates separated by 150 mm) using f_e=30 kHz.

V. RESULTS

Sample results obtained for Test objects 1 and 2 at ambient pressure (0.14 MPa) and 8 °C are shown in Fig. 7. The measured results were obtained from pulses obtained using 10 and 30 kHz envelope frequencies and were corrected for spreading losses. As can be seen, excellent agreement is obtained with the theoretical model, in particular for test object 2 where the frequencies of the peaks in the response match those of the model well. In general, good agreement is achieved for the absolute levels of the minima in the response, but the measured peaks are often lower than the predictions. It should be noted that the test objects may not necessarily conform to the idealized description used in the model with resulting implications for the level of agreement. For example, small variations in sound speed or plate separation give rise to differences in the frequencies at which the maxima and minima occur. In addition, the levels of the peaks are sensitive to the absorption in the panel, and the model does not include the effect of damping due to flexing



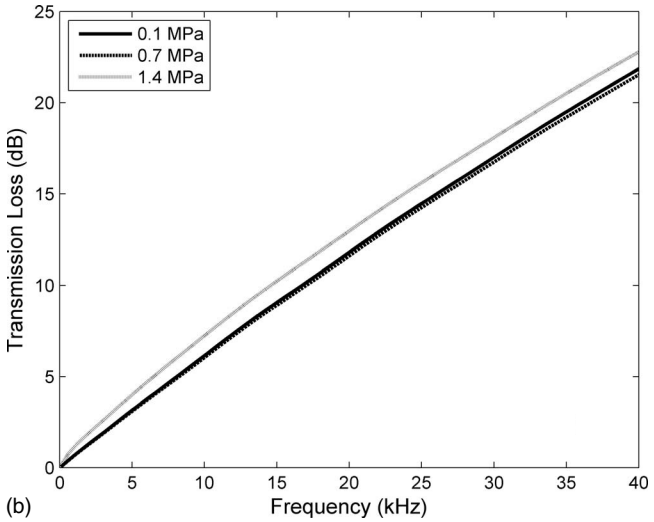
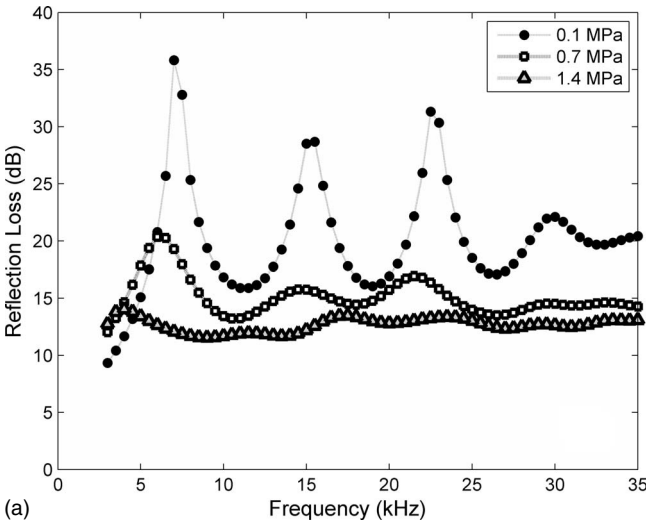


FIG. 8. Model predictions for test panel 1 at three values of hydrostatic pressure. (a) Reflection loss values calculated for 26 °C and (b) transmission loss values for 8 °C.



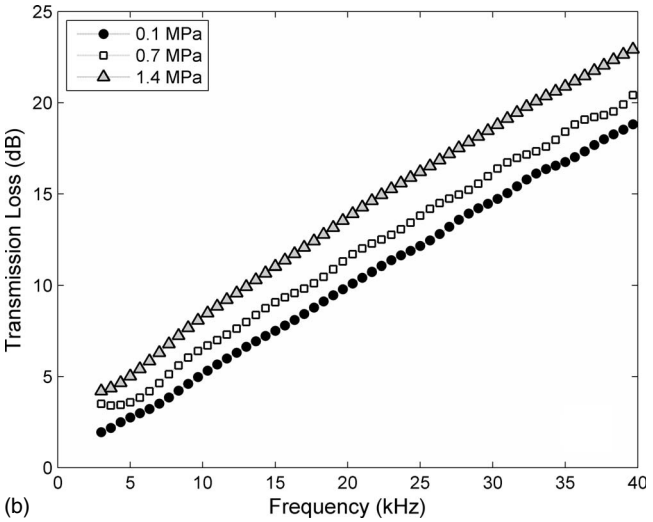
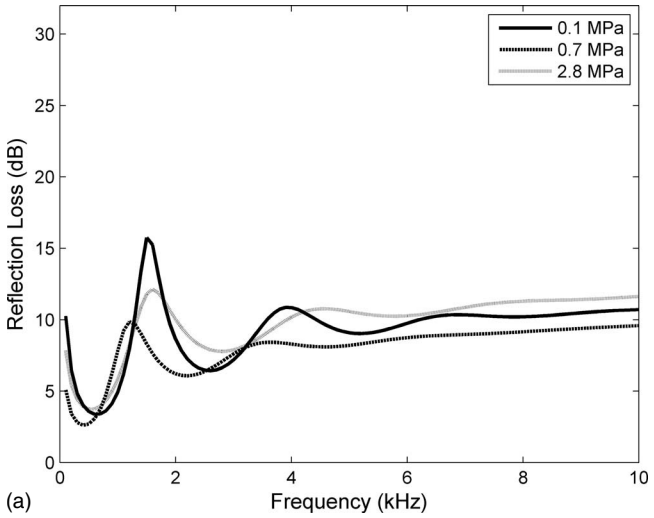


FIG. 9. Measured results for test panel 1 at three values of hydrostatic pressure using a 20 kHz pulse envelope. (a) Reflection loss was measured at 26 °C, whereas (b) transmission loss was measured at 8 °C.

or movement against the threaded spacers used to maintain the plate separation. With these thin panels, there were few problems from diffracted signals, but reflections from sidewalls and mounts posed greater problems and care was required to avoid including such signals in the measurement window.

Figure 8 shows the predictions from the model for test panel 1 at three values of hydrostatic pressure: 0.1, 0.7, and 1.4 MPa. Reflection loss values have been calculated for 26 °C, whereas the transmission loss values are for 8 °C. Figure 9 shows the corresponding measured results obtained using a 20 kHz modulated envelope at the same values of hydrostatic pressure and temperature. These results were taken for increasing hydrostatic pressure. As can be seen, the results for both reflection loss and transmission loss agree relatively well in terms of the overall loss levels and the general trends. The frequencies of the peaks in the reflection loss agree very well with the predictions, although the levels of the measured peaks are higher than predicted. One observation from the measurements is that there is a systematic change in loss value (reduction in reflection loss, increase in transmission loss) in going from ambient pressure to a pressure of 0.7 MPa, whereas the predicted values show no change. This may be due to the actual behavior of the gas-filled microspheres under hydrostatic pressure being slightly different from the predicted behavior. The current model does not take into account the distribution of sizes and assumes that the microspheres remain spherical during their deformation with pressure, nor does it account for hysteresis of the microspheres.

Figure 10 shows the predictions from the model for test panel 2 for 8 °C and for three values of hydrostatic pressure over a lower-frequency range. Reflection loss values have been calculated for 0.1, 0.7, and 2.8 MPa, whereas transmission loss values have been calculated for 0.1, 1.4, and 2.8 MPa. Figure 11 shows the corresponding measurements at the same values of hydrostatic pressure and temperature. Pulses with 5 and 3 kHz envelope frequencies were used for the measurement of reflection loss and transmission loss, respectively. Again the general trends in the predictions are observed in the results, with the peaks in the measured reflection loss observed at the predicted frequencies. However,



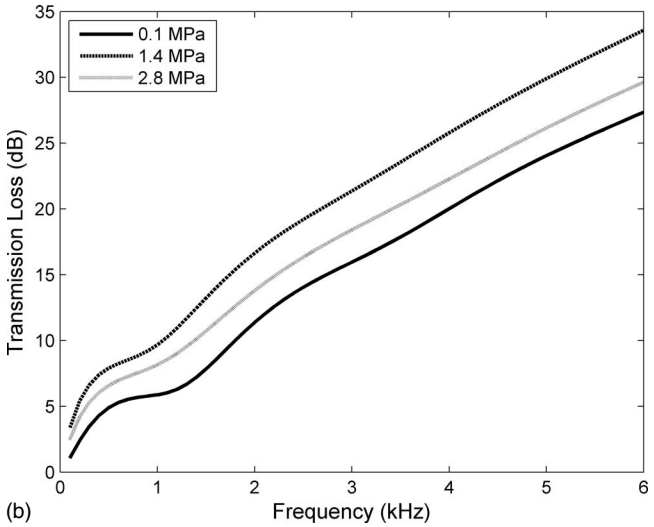
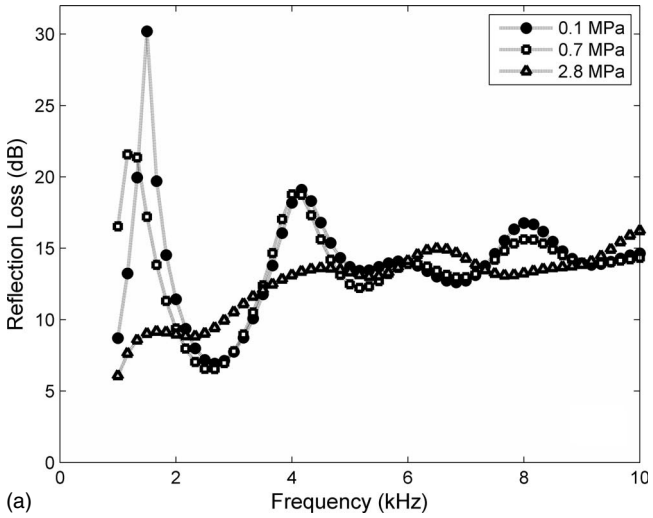


FIG. 10. Model predictions for (a) reflection loss and (b) transmission loss of test panel 2 at 8 °C for three values of hydrostatic pressure.



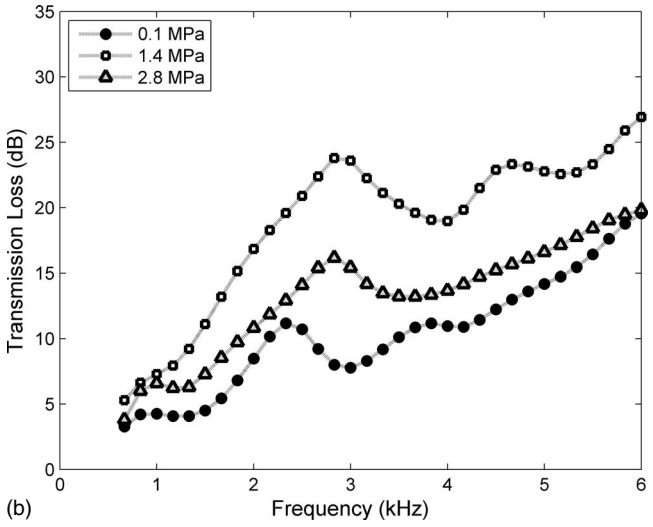


FIG. 11. Measured results for test panel 2 at 8 °C and three values of hydrostatic pressure. (a) Reflection loss was measured with a pulse generated using f_e=5 kHz, whereas (b) transmission loss was measured with a pulse generated using f_e=3 kHz.

the overall measured loss levels do not generally agree so well with the predictions, although this may well be due to inaccuracies in the material properties, of both the matrix and the microspheres, entered as model parameters. Interestingly, the overall level of measured reflection loss exceeds the predicted values. However, the low-frequency peak is not well resolved.

Panel 2 has a much higher transmission loss performance than panel 1; at 5 kHz and 1.4 MPa, a transmission loss of 30 dB is predicted compared with 5 dB. This makes it more difficult to measure accurately since the transmitted signal is now smaller and of the same order of magnitude as the diffracted signals from the panel edge (and the stray reflected signals from vessel walls and mounts). To fully capture the low-frequency pulses and to obtain satisfactory frequency domain information at frequencies of around 1 kHz requires a long time window (typically about 1.5 ms for the measurements shown in Fig. 11). A long window also enables any internal multipath reflections to fully contribute to the transmitted signal. However, it is highly probable that diffracted signals have contributed to the measurements made here, leading to a reduction in the measured transmission loss level and the unexpected fluctuations with frequency.

VI. DISCUSSION

There are a number of potential sources of error in the panel measurements described here. Some of these errors originate from the specific measurement procedure adopted. For example, the use of a broadband pulse gives accurate results only where there is sufficient energy in the spectrum to provide sufficient signal-to-noise ratio. As a rule of thumb, data were used over a frequency range where the spectral amplitude was greater than 5% of the spectral maximum (see Fig. 2). To cover the frequency range from 1 to 50 kHz requires the use of pulses generated using a range of envelope frequencies. For the results shown in Sec. V, appropriate pulse modulations were chosen to illustrate features in the response of specific panels. However, in general the data obtained for different envelope frequencies will overlap. To

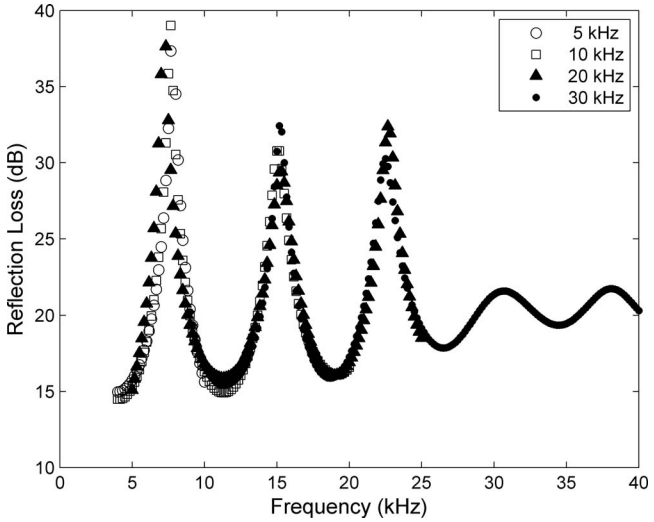


FIG. 12. Results for test panel 1 at 26 °C and ambient pressure (0.1 MPa) for pulses generated using envelope frequencies (f_e) of 5, 10, 20, and 30 kHz showing agreement between frequency ranges.

illustrate the agreement between frequency subranges, Fig. 12 shows the results of measurements for test panel 1 at 26 °C and 0.1 MPa for pulses with envelope frequencies of 5, 10, 20, and 30 kHz.

Due to the requirement to undertake separate measurement runs with and without the panel in place, the performance of the acoustic filter can introduce errors into the measurements. The procedure adopted to minimize the effect of the hysteresis in the filter performance was described in Sec. II. It is possible that there may be some residual effect contributing to the measurement error, but this is likely to be relatively small. The requirement for two measurement runs also places stringent requirements on the positioning of the hydrophone for reflection loss measurements where wave form subtraction is required (see Sec. III B). Any small residual uncanceled signals will interfere with the reflected signal and will give rise to erroneous fluctuations in the calculated loss. However, since any lack of cancellation will be apparent during the wave form processing, it should be evident that this problem has arisen, and this may inform any consideration of the overall uncertainty.

The hydrophone performance may also influence the accuracy of the results obtained. For reflection loss measurements, the reflected signal approaches the hydrophone from the opposite direction to the incident signal. However, at the relatively low kilohertz frequencies used here, the small hydrophones used are sufficiently omnidirectional for any difference in sensitivity to be small ($<$0.5 dB). If the hydrophone sensitivity varies with hydrostatic pressure, this should not be an issue unless the hydrophone exhibits hysteresis (not observed for the hydrophones used here). The procedure described in Sec. III B minimized problems from air bubbles by wetting the hydrophone and keeping it immersed throughout the measurements.

Notwithstanding the above, there are some sources of error which are perhaps more fundamental in nature, since they originate from the nonideal nature of experimental configuration. Clearly, the ideal configuration of ensonifying an infinite panel with a perfect plane wave is not realistic. As has been described earlier, the finite size of the panel leads to diffracted signals contaminating the measured transmitted and reflected signals. Reducing the effect by use of a larger panel is not feasible since the size of the panel is limited by the dimensions of the test vessel (the use of the APV being motivated by the desire to test panel performance as a function of temperature and depth). Although the use of the parametric array source reduces the signal level diffracted from the panel edges, some diffracted signals are still present, and the reduction is not significant at frequencies as low as 3 kHz.

Another factor that can potentially result in a systematic error is the nonplanar nature of the field, which is incident upon the test panel in the vessel. In fact, with the separation distances used for the measurements reported in Sec. V, the test panel may be in the near field of the source array. The lack of plane-wave ensonification has been shown to influence transmission loss measurements for elastic panels at higher frequencies (Humphrey, 1985, 1986) and has been predicted, but not observed, for measurements in the current frequency range (Piquette, 1988a, 1988b). To study this experimentally, a scan could be performed, and the incident wave field decomposed into its angular spectrum using spatial Fourier techniques. The effect of the plane wave spectrum on measurements is the subject of current research.

A factor that can significantly influence the calculated results is the positioning and length of the window used to isolate the desired signals from unwanted signals. Choice of the window length is a tradeoff between a number of factors. Too long a window will allow contribution to the measured signal from diffracted waves from panel edges and waves reflected from tank boundaries and rigging. With too short a window, any contributions from internal multipath reflections within the panel will not be captured, and the lowest frequencies may not be well represented. This is a particular issue for low-velocity materials. To test the influence of window length, examples of the measured wave form data for test panels 1 and 2 were processed using time windows of differing lengths. The results of this are shown in Fig. 13.

The plots in Fig. 13 show the mean of the results obtained with differing window lengths, with the error bars denoting the random uncertainties calculated for a confidence level of 95% (this is calculated from the standard deviation of the results, divided by the square root of the number of window lengths used, then multiplied by the appropriate Student's t-factor). The error bars provide a measure of the sensitivity of the results to window length. It should be noted that the variation observed is due to all the factors described above, and it is not possible to distinguish between the different influences merely by varying the window length.

In Fig. 13, greater variation is observed near the peaks of the reflection loss data for test panel 1. Some of this is due to the limited frequency resolution available (a slight movement of a sharp peak can cause relatively large apparent change in loss value). However, this does give an indication that the loss values at the maxima are more difficult to measure accurately. Away from the peaks and at higher frequen-

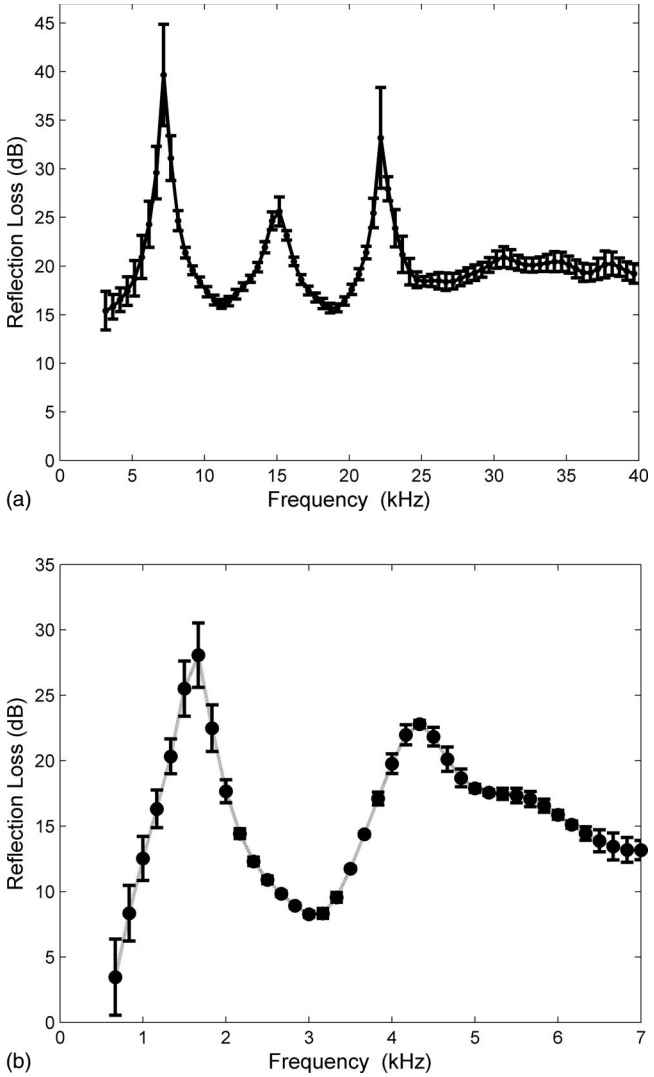


FIG. 13. Results for reflection loss of test panel 1 (left) at a pressure of 0.35 MPa and a temperature of 26 °C (f_e=20 kHz) and for test panel 2 at a pressure of 0.1 MPa and a temperature of 8 °C (f_e=3 kHz). For panel 1, the measurement window was varied from 0.55 to 0.77 ms in steps of 12 μs; for panel 2, it was varied from 1.3 to 1.54 ms in steps of 12 μs. The mean values of the calculated results are shown with the error bars calculated for a confidence level of 95%.

cies, the typical uncertainty values are between 1.0 and 1.5 dB. For the transmission loss (not shown), the error bar values are fairly constant with an increase at the lowest frequencies. For test panel 2, the error bar values increase for frequencies less than 2 kHz for both reflection and transmission losses.

The plots in Fig. 13 show that for the majority of the frequency range, the results do not depend too strongly on window position and length. However, it should be noted that since the two test panels used had reasonably high inherent absorption within the material, there were not many multipath internal reflections to contribute to the transmitted or reflected signals. For a panel that is less absorbing, the extra multipaths would make it more difficult to measure accurately with the experimental arrangement used here.

Another general observation from the measurements was that the test panels themselves could exhibit some hysteresis. This means that when comparing different measurements on the same panel, it was important that the recent pressure history was also as near identical as possible. For example, two sets of results may not be comparable if one set was obtained on the first pressure cycle of the day whereas another set was obtained on the second or third cycle, or if the cycles were to different maximum pressures. As with the acoustic filter, an overnight period at ambient pressure was sufficient for the panel to recover fully.

VII. CONCLUSIONS

A technique has been described for determining the acoustic performance of panel materials under simulated ocean conditions in a laboratory test facility using a parametric array as a source of sound. The measurement technique has been established within a laboratory test vessel capable of simulating ocean depths down to 700 m, with water temperature control in the range of 2–35 °C. With this technique, the reflection loss and transmission loss of the materials comprising the test panel may be determined at frequencies from a few kilohertz to 50 kHz.

The parametric array enables wideband measurements to be undertaken with short-duration pulses and reduces the effects of diffraction from the panel edges. An acoustic filter was used to truncate the array in order to provide a source-free measurement region and to simplify the measurement process. Establishing a parametric array in the confined space of the vessel posed a number of engineering challenges, and a specific experimental procedure had to be adopted to avoid problems with hysteresis effects in the acoustic filter.

To illustrate and validate the technique, measurements have been made on a number of test objects and panels. Experimental results have been presented for the measurement of echo reduction and transmission loss of two absorbing test panels in the frequency range from a few kilohertz to 30 kHz for hydrostatic pressures of up to 2.8 MPa and for temperatures of 8 and 26 °C.

The presence of diffracted waves from the panel edges (and some reflected waves from the tank boundaries) limits the accuracy of the technique at low frequencies. The benefit of using the parametric array is not so pronounced at frequencies below 10 kHz, and it may be of benefit to use a directional sensor as a receiver to provide extra discrimination against diffracted signals (Martin *et al.* 2007).

ACKNOWLEDGMENTS

The authors acknowledge the support of the National Measurement System Policy Unit of the UK's Department of Innovation, Universities and Skills and the UK Ministry of Defence for funding the work described here. The authors would also like to thank John House (QinetiQ) for advice and assistance, and Dr. Richard Bryant (QinetiQ), and Alison Daniel (QinetiQ) for fabrication of the test panels.

Audoly, C. (**1992**). "Global characterisation of acoustic panels at normal incidence for underwater applications," Acustica **76**, 129–136.
Audoly, C., and Giangreco, C. (**1990**). "Improvement of measurement of the transmission coefficient of panels at normal incidence using surface receivers," J. Acoust. **3**, 369–379.

Baird, A. M., Kerr, F. H., and Townend, D. J. (**1999**). "Wave propagation in a viscoelastic medium containing fluid-filled microspheres," J. Acoust. Soc. Am. **105**, 1527–1538.
Beamiss, G. A., Hayman, G., and Robinson, S. P. (**2001**). "The provision of standards for underwater acoustics at simulated ocean conditions by use of the, National Physical Laboratory, UK, Acoustic Pressure Vessel," NPL Report No. CMAM76.
Berktay, H. O. (**1965**). "Possible exploitation of non-linear acoustics in underwater transmitting applications," J. Sound Vib. **2**, 435–461.
Bobber, R. J. (**1988**). *Underwater Electroacoustic Measurements*, 2nd ed. (Peninsula, Los Altos, CA).
Brekhovskikh, L. M. (**1960**). *Waves in Layered Media* (Academic, New York).
Cao, H., Humphrey, V. F., and Berktay, H. O. (**1995**). "Sound pulse scattering from an edge of thick elastic plates: Experimental and numerical investigations," Acta Acust. (Beijing) **14**, 317–329 (in Chinese).
Capps, R. N., Weber, F. J., and Thompson, C. M. (**1981**). "Handbook of sonar transducer passive materials," Naval Research Laboratory, Memorandum Report No. 4311.
Darner, C. L. (**1954**). "An anechoic tank for underwater sound measurements under high hydrostatic pressures," J. Acoust. Soc. Am. **26**, 221–222.
Esward, T. J., Humphrey, V. F. Evans, L. C., Beamiss, G. A., and Hayman, G. (**2002**). "Acoustic characterisation of panel materials in simulated ocean conditions using the, NPL acoustic pressure vessel," Proceedings of 6th ECUA2002, pp. 665–670.
Ferry, J. D. (**1980**). *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley, New York).
Hagelberg, M. P., and Corsaro, R. H. (**1985**). "A small pressurised vessel for measuring the acoustic properties of materials," J. Acoust. Soc. Am. **77**, 1222–1228.
Hamilton, M. F. (**1998**). *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York), Chap. 6, pp. 223–261.
Humphrey, V. F. (**1985**). "The measurement of acoustic properties of limited size panels by use of a parametric source," J. Sound Vib. **98**, 67–81.
Humphrey, V. F. (**1986**). "The influence of the plane wave spectrum of a source on measurements of the transmission coefficient of a panel," J. Sound Vib. **108**, 261–271.
Humphrey, V. F. (**1988**). "Applications of parametric acoustic arrays in laboratory scale experiments," Proceeding of the Institute of Acoustics (UK). **10**, 37–55.
Humphrey, V. F. (**2002**). "Parametric arrays: Laboratory applications in underwater acoustics," Proceedings of the Sixth European Conference on Underwater Acoustics, Gdansk, Poland, 24-27 June 2002, pp. 3–8.
Humphrey, V. F., and Berktay, H. O. (**1985**). "The transmission coefficient of a panel measured with a parametric source," J. Sound Vib. **101**, 85–106.
Humphrey, V. F., Carroll, N. L., Smith, J. D., Beamiss, G. A., Hayman, G., Esward T. J., and Robinson, S. P. (**2003**). "Acoustic characterisation of panel materials under simulated ocean conditions," Proc. I.O.A. **25**, 11.1–11.8.
Humphrey, V. F., and Hsu, C. H. (**1980**). "Non-linearity of cylindrical hydrophones used for the measurement of parametric arrays," Proceedings of the Institute of Acoustics Conference on Transducers for Sonar Applications, University of Birmingham, UK, December 1980, pp. 5.1–5.10.
Hurrell, A. (**2002**). "Finite difference modelling of acoustic propagation and its applications in underwater acoustics," Ph.D. thesis, University of Bath, UK.
Martin, M. J. Hugin, C. T., Robinson, S. P., Beamiss, G. A., Hayman, G., Smith, J. D., and Humphrey, V. F. (**2007**). "The low frequency characterisation of underwater material properties in a pressure vessel via single, dual and multiple hydrophone techniques," Proceedings of the Second International Conference and Exhibition on "Underwater Acoustic Measurements: Technologies and Results," Crete, 15-29 June 2007, pp. 413–420.
Mikeska, E. E., and Behrens, J. A. (**1976**). "Evaluation of transducer window materials," J. Acoust. Soc. Am. **59**, 1294–1298.
Moffet, M. B., and Henriquez, T. A. (**1982**). "Hydrophone non-linearity measurements," J. Acoust. Soc. Am. **72**, 1–6.
Piquette, J. C. (**1986**). "An analytical technique for reducing the influence of edge diffraction in reflection measurements made on thin acoustical panels," J. Acoust. Soc. Am. **80**, 19–27.
Piquette, J. C. (**1987**). "An extrapolation procedure for transient reflection measurements made on thick acoustical panels composed of lossy dispersive materials," J. Acoust. Soc. Am. **81**, 1246–1258.
Piquette, J. C. (**1988a**). "Spherical-wave scattering by a finite-thickness solid plate of infinite lateral extent, with some implications for panel measurements," J. Acoust. Soc. Am. **83**, 1284–1294.
Piquette, J. C. (**1988b**). "Interactions of a spherical wave with a bilaminar plate composed of homogeneous and isotropic solid layers," J. Acoust. Soc. Am. **84**, 1526–1535.
Piquette, J. C. (**1991**). "Technique for detecting the presence of finite sample-size effects in transmitted-wave measurements made on multiplayer underwater acoustic panels," J. Acoust. Soc. Am. **90**, 2831–2842.
Piquette, J. C. (**1994**). "Direct measurement of edge diffraction from soft underwater acoustic panels," J. Acoust. Soc. Am. **95**, 3090–3099.
Piquette, J. C. (**1996**). "Some new techniques for panel measurements," J. Acoust. Soc. Am. **100**, 3227–3236.
Piquette, J. C., and Paolero, A. E. (**2003**). "Phase change measurement, and speed of sound and attenuation determination, from underwater acoustic panel tests," J. Acoust. Soc. Am. **113**, 1518–1525.
Preston, R. C., and Robinson, S. P. (**1998**). "UK underwater acoustical measurement standards," Proceedings of the Fourth European Conference on Underwater Acoustics (ECUA 1998), Vol. **1**, pp. 133–138.
Rudgers, A. J., and Solvold, C. A. (**1984**). "Apparatus-independent acoustical material characteristics obtained from panel test measurements," J. Acoust. Soc. Am. **76**, 926–934.
Thibieroz, T., and Giangreco, C. (**1991**). "Impulse acoustical measurements of the transmission coefficients of immersed panels," J. Acoust. **4**, 215–236.
Trivett, D. H., and Robinson, A. Z. (**1981**). "Modified Prony method approach to echo reduction measurements," J. Acoust. Soc. Am. **70**, 1166–1175.
Varadan, V. K., Ma, Y., and Varadan, V. V. (**1985**). "A multiple scattering theory for elastic wave propagation in discrete random media," J. Acoust. Soc. Am. **77**, 375–385.
Waterman, P. C., and Truell, R. (**1961**). "Multiple scattering of waves," J. Math. Phys. **2**, 513–537.
Westervelt, P. J. (**1963**). "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537.

Analysis of time delay effects on a linear bubble chain system[a)]

Andrew Ooi[b)] and Aneta Nikolovska[c)]
Department of Mechanical Engineering, The University of Melbourne, Parkville Melbourne, Victoria 3010, Australia

Richard Manasseh[d)]
Fluid Dynamics Group, CSIRO Manufacturing Science and Engineering, P.O. Box 56 (Graham Road), Highett, Melbourne, Victoria 3190, Australia

(Received 30 July 2007; revised 16 May 2008; accepted 19 May 2008)

A chain of vertically rising discrete air bubbles represents a transition phenomenon from individual to continuum behavior in a bubbly liquid. Previous studies have reported that there is a preference for acoustic energy to propagate along the bubble chain and that this behavior could be explained by a coupled-oscillator model. However, it has recently been demonstrated that quantitative results from the coupled-oscillator model do not match experimental data. In this paper, it is shown how adding time delays to the coupled-oscillator model can produce results that are in better agreement with experimental data. In addition, the effects of time delays on the natural frequencies and damping of individual eigenmodes of the vertical bubble chain are also investigated. It was found that adding time delays can dramatically change the damping of the different modes of the system while having less dramatic impact on the natural frequencies of the individual eigenmodes. Counterintuitively, it is found that the effects of time delays appear to be more important when the bubbles are closer together than when they are farther apart. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945156]

PACS number(s): 43.30.Nb, 43.20.Fn, 43.20.Px, 43.20.Bi [RCG] Pages: 815–826

I. INTRODUCTION

Many researchers have directed efforts towards the development of a model to describe the volumetric oscillations of gas bubbles in liquid. The acoustic behavior of an isolated bubble has been extensively researched (see Ref. 1 and references therein). There have also been many publications on the dynamics of the acoustic field in the vicinity of bubble pairs (see Refs. 2–9). More recently, Ida[10] has extended this line of investigation and studied the natural frequencies of a system consisting of three bubbles of different sizes. When there are many bubbles in the medium, Commander and Prosperetti,[11] Duraiswami *et al.*,[12] and Nicholas *et al.*[13] have taken the continuum approach and have assumed a spatially homogeneous medium with the void fraction being the important parameter. These continuum theories have been used by Phelps *et al.*,[14] Duraiswami *et al.*,[12] and Terrill and Melville[15] as the basis of several instruments for oceanographic and industrial applications. Most of these systems measure bubble-size distributions, relying on an active principle. Sound is sent into the water and the attenuation or reflection of the resulting signal is interpreted to infer the bubble-size distribution. On the other hand, there are passive systems, which rely on bubbles emitting sound at their natural frequencies. It is not difficult to obtain experimental pressure signals emitted (passively) by a bubbly flow and such measurements have been made for many years (see Refs. 15–18). However, most analyses have been carried out in the frequency domain, with attention focusing on how to convert the spectra into bubble-size distributions. Above all, most analyses assume the bubbles are isotropically and homogeneously distributed in the fluid, which is, in fact, a rarity, either in nature or industry.

The acoustic signature of a finite number of discrete air bubbles represents an important yet little-researched area in multiphase flow. The scattering of acoustic waves from gas bubbles in liquids plays a crucial role in determining the acoustic attenuation and dissipation in the surrounding medium. Such systems could provide insight into sound propagation in the anisotropic and inhomogeneously distributed bubbly systems that are the norm in practice. The current work is aimed at the analysis of the coupled-oscillator mathematical model when applied to a vertical chain of bubbles (for example, see Fig. 1) in which the individual bubbles resonate in the monopole (isotropic) mode (see Ref. 19). Even though the propagation of acoustic energy generated by each individual bubble might be isotropic, it has been found that the distribution of acoustic energy close to a bubble chain is not isotropic, due to the interaction of the acoustic waves with other bubbles. Manasseh *et al.*[20] and Nikolovska *et al.*[21] have applied a coupled-oscillator formalism to describe the collective scattering due to multiple gas bubbles in a chain. The sound was initiated naturally on the detachment of each bubble from a nozzle at the base of the chain and the effects of other bubbles were modeled through a set of coupled ordinary differential equations.

[a)]Submitted in August 2007 to the Journal of the Acoustical Society of America.
[b)]Electronic mail: a.ooi@unimelb.edu.au
[c)]Electronic mail: aneta@uni-bremen.de
[d)]Electronic mail: richard.manasseh@csiro.au

FIG. 1. (Color online) Photo of a typical bubble chain.

In this paper, an analysis is carried out using the coupled-oscillator model proposed by Feuillade[6] and used by many others (Feuillade,[19] Tolstoy,[3] and Doinikov *et al.*[22]). In particular, the effects of time delays on the natural frequencies and damping of the bubble chain system will be highlighted. A time delay arises from the finite speed of sound propagation in the liquid or, in other words, from the finite compressibility of the liquid. As a result, the acoustic pressure field influencing the oscillations of all the other bubbles is a time-retarded field.[22] The present work is essentially an extension of the study carried out by Doinikov *et al.*,[22] which highlighted the effects of time delays on a bubble chain with different numbers of bubbles. In Ref. 22, the effects of adding more bubbles in the chain were investigated. The distance between the bubbles was kept constant so when more bubbles were added into the system, the bubble chain just became longer. This is very different to what happens in laboratory experiments, where it is more common for the length of the bubble chain to be kept constant because the water tank used is likely to be of a constant height. The number of bubbles in the chain can be increased by increasing the airflow rate in the nozzle and as a result, the distance between the bubbles in the chain will also become smaller. Hence, the distance between the bubbles is an important parameter in the problem and the effects of time delays as a function of bubble separation are investigated in the present paper.

II. GOVERNING EQUATIONS AND ANALYSIS

In the following analysis, the "self-consistent" approach (see previous work by Tolstoy,[3] Feuillade,[19] and Feuillade[6]) will be used. This is one method of overcoming (or rather side-stepping) the problem of infinite rereflexions in coupled multibubble systems. The total pressure field incident on any bubble is taken as the sum of the pressure fields radiated by all the other bubbles in the system, plus any external forcing. However, acoustic energy emitted by a bubble, once incident on a second bubble, is reradiated back to the first, and so on, creating an infinite series, analogous to a pair of mirrors facing each other. With many bubbles, this is analytically problematic. This problem arises in many scattering phenomena in physics, not only in acoustics.[3] The key to the self-consistent approach is that the pressure fields from the other bubbles are defined to have already experienced the infinity of other interactions, which affect them. One consequence is that the dependent variables [which will be defined as $x_n(t)$ below] lose their precise physical meaning as the radial perturbation of the nth bubble. The merits of using the self-consistent approach as opposed to the alternative multiple scattering approach have been discussed by Feuillade[6] and Twersky.[23]

In order to simplify the analysis, it will also be assumed that the equilibrium radii of all bubbles in the chain are the same and that the distance between the neighboring bubbles in the chain is constant. In the experiments, the bubble rise velocity was higher further up the chain[20] owing to the initial acceleration of the bubbles from their detachment point at the base of the chain. Hence, the spacing between the bubbles was larger higher up the bubble chain.

As the effort in this paper is to extend the work presented by Doinikov *et al.*,[22] the same notation as in that paper will be used. Using the assumptions above and assuming a compressible liquid (i.e., a finite speed of sound propagation), small, radial free oscillations of N coupled bubbles are described by the following equations:

$$\ddot{x}_n(t) + \omega_0 \delta \dot{x}_n(t) + \omega_0^2 x_n(t) = -\sum_{m=1, m\neq n}^{N} \frac{R_0}{d_{nm}} \ddot{x}_m(t - d_{nm}/c), \tag{1}$$

where, in the absence of radiative interactions, $x_n(t)$ is the (small) change in radius of the nth bubble. Because we use the self-consistent approach, x_n is defined to include the effects of radiative interactions in the system and hence can no longer be precisely compared with the behavior of the nth bubble. However, for the present paper, that is not a problem since comparisons with experiment are only being made with quantities such as pressure and propagation speed, not on the

behavior of individual bubbles. R_0 is the equilibrium radius, d_{nm} is the distance between the centers of the nth and mth bubbles, and c is the speed of sound in bubble-free water. δ is the total damping coefficient of an isolated bubble consisting of radiation and thermal and viscous dampings (see Refs. 1 and 24). We assume adiabatic conditions and the resonant (Minnaert) angular frequency is calculated as

$$\omega_0 = \frac{1}{R_0}\sqrt{\frac{3\gamma p_0}{\rho}},$$

where p_0 is the static pressure, $\gamma = 1.4$ is the ratio of specific heats, and ρ is the density of the liquid. The coupling term on the right hand side of Eq. (1) is the time delay term and it reflects the finite time it takes for a disturbance to reach a nearby bubble. For an incompressible fluid, $c = \infty$, hence Eq. (1) can be simplified to

$$\ddot{x}_n(t) + \omega_0 \delta \dot{x}_n(t) + \omega_0^2 x_n(t) = -\sum_{m=1, m\neq n}^{N} \frac{R_0}{d_{nm}} \ddot{x}_m(t). \tag{2}$$

It must be noted that Eqs. (1) and (2) assume that the acoustic field that affects a bubble is predominantly from a monopolar source (see Ref. 6). This suggests that Eqs. (1) and (2) should only be used in situations where $d_{nm} \gg R_0$ (see Ref. 22).

The solution to Eqs. (1) and (2) can be expressed in the form

$$\mathbf{x}(t) = \mathbf{A}e^{\lambda t}, \tag{3}$$

where $\mathbf{x}(t)$ is a vector whose individual components are $x_n(t)$ and $\mathbf{A}$ is the eigenvector corresponding to the eigenvalue λ. Substituting Eq. (3) into Eq. (1) gives

$$[\lambda^2 \mathbf{I} + \lambda \mathbf{C} + \mathbf{K}]\mathbf{A} = -\lambda^2 \mathbf{R}(\lambda)\mathbf{A}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix,

$$\mathbf{C} = \omega_0 \delta \mathbf{I},$$

$$\mathbf{K} = \omega_0^2 \mathbf{I},$$

and

$$\mathbf{R} = \begin{bmatrix} 0 & (R_0/d_{12})e^{-\lambda\tau_{12}} & (R_0/d_{13})e^{-\lambda\tau_{13}} & \cdots & \cdots & (R_0/d_{1N})e^{-\lambda\tau_{1N}} \\ (R_0/d_{21})e^{-\lambda\tau_{21}} & 0 & (R_0/d_{23})e^{-\lambda\tau_{23}} & \cdots & \cdots & (R_0/d_{2N})e^{-\lambda\tau_{2N}} \\ (R_0/d_{31})e^{-\lambda\tau_{31}} & \cdots & 0 & (R_0/d_{34})e^{-\lambda\tau_{34}} & \cdots & (R_0/d_{3N})e^{-\lambda\tau_{3N}} \\ (R_0/d_{41})e^{-\lambda\tau_{41}} & \cdots & \cdots & 0 & \cdots & (R_0/d_{4N})e^{-\lambda\tau_{4N}} \\ \vdots & \cdots & \cdots & \cdots & 0 & \vdots \\ (R_0/d_{N1})e^{-\lambda\tau_{N1}} & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}.$$

$\mathbf{R}(\lambda)$ is a symmetric matrix with zeros along the main diagonal. The off-diagonal terms are exponential functions of $\lambda\tau_{nm}$, where $\tau_{nm} = d_{nm}/c$ is the time delay. It is more conventional to write Eq. (4) as

$$\mathbf{D}(\lambda, \tau_{nm}) = [\lambda^2(\mathbf{I} + \mathbf{R}(\lambda)) + \lambda\mathbf{C} + \mathbf{K}]\mathbf{A} = \mathbf{0}. \tag{5}$$

Equation (5) represents a nonlinear eigenvalue problem. Given values of τ_{nm}, one needs to find values of λ (eigenvalues) such that the determinant of $\mathbf{D}$ is zero. For a $N \times N$ system, there are $2N$ eigenvalues,

$$\lambda = \xi \pm i\omega, \tag{6}$$

where ω is the natural frequency, ξ is the damping for the particular eigenvector (mode), and i is the imaginary unit. The eigenvalues occur in complex conjugate pairs.

For a bubble in isolation, it can easily be shown that

$$\lambda = -\frac{\omega_0 \delta}{2} \pm i\omega_0 \sqrt{\left(1 - \frac{\delta^2}{4}\right)}.$$

thus,

$$\xi_0 = \frac{\omega_0 \delta}{2}$$

and

$$\omega_{01} = \omega_0 \sqrt{\left(1 - \frac{\delta^2}{4}\right)}.$$

Note that for the millimeter-sized bubbles considered in this paper, δ is small so $\omega_{01} \approx \omega_0$. In the presentation of results below, all data will be normalized with the parameters from the isolated bubble case, i.e.,

$$\xi^* = \frac{\xi}{\xi_0} \quad \text{and} \quad \omega^* = f^* = \frac{\omega}{\omega_{01}}. \tag{7}$$

If time delays are neglected (i.e., if the liquid is assumed to be incompressible), then $\tau_{nm} = 0$ and the matrix $\mathbf{I} + \mathbf{R}$ in Eq. (5) no longer consists of any exponential functions. Hence, we have a conventional quadratic eigenvalue problem. This type of problem has many applications and has been studied extensively by many researchers (see Ref. 25 and references therein). The eigenvalues and eigenvectors for the quadratic eigenvalue problem can be obtained using standard numeri-

cal routines that are widely available (e.g., the *polyeig* function in MATLAB 7.0). If time delays are taken into consideration, then obtaining the eigenvectors and eigenvalues of Eq. (5) is more complicated. A method of finding the eigenvalues for time delay systems has been outlined by Hu *et al.*[26] In that paper, they introduced a numerical method to obtain eigenvectors and eigenvalues for a system where there is time delay in x and $\dot{x}$ in the governing equation. Furthermore, Hu *et al.*[26] only considered systems where there is only one (constant) time delay for the whole system. Thus, the system considered here is different to the problems considered by Hu *et al.*[26] in two respects. Firstly, the time delay is in the $\ddot{x}$ term [see Eq. (1)], and secondly, the time delays, τ_{nm}, are not constants but are dependent on the distance between bubbles n and m. Hence, the method suggested by Hu *et al.*[26] needs to be modified and extended for the system of equations that is of interest here.

First, consider the case where time delays are neglected, i.e., $\tau_{nm}=0$. We require a solution for $\mathbf{A}_r$ and λ_r that satisfies

$$\mathbf{D}(\lambda_r,0)\mathbf{A}_r = 0. \tag{8}$$

This represents a conventional quadratic eigenvalue problem. When one takes into account time delays, then there exists an eigenvalue, λ_t, and eigenvector, $\mathbf{A}_t$, near λ_r and $\mathbf{A}_r$ such that the following condition is true:

$$\mathbf{D}(\lambda_t,\tau_{nm})\mathbf{A}_t = 0, \tag{9}$$

i.e., λ_t and $\mathbf{A}_t$ are the eigenvalues and eigenvectors of the time delay system [Eq. (9)]. We will write

$$\lambda_t = \lambda_r + \Delta\lambda_r \quad \text{and} \quad \mathbf{A}_t = \mathbf{A}_r + \Delta\mathbf{A}_r. \tag{10}$$

Substituting Eq. (10) into Eq. (9) gives

$$\mathbf{D}(\lambda_r + \Delta\lambda_r, \tau_{nm})(\mathbf{A}_r + \Delta\mathbf{A}_r) = 0. \tag{11}$$

If we perform a Taylor series expansion about λ_r and ignore terms of the order of $(\Delta\lambda_r)^2$ and $\Delta\lambda_r\Delta A_r$, then one may write

$$\mathbf{D}(\lambda_r,\tau_{nm})(\mathbf{A}_r + \Delta\mathbf{A}_r) = \Delta\lambda_r\mathbf{E}(\lambda_r,\tau_{nm})(\mathbf{A}_r), \tag{12}$$

where

$$\mathbf{E}(\lambda_r,\tau_{nm}) = -(2\lambda_r\mathbf{I} + \mathbf{R}(-\lambda_r^2\tau_{nm} + 2\lambda_r) + \mathbf{C}). \tag{13}$$

Following Hu *et al.*,[26] we define a vector $\mathbf{P}_r$ such that

$$\mathbf{P}_r = \frac{1}{\Delta\lambda_r}(\mathbf{A}_r + \Delta\mathbf{A}_r). \tag{14}$$

Equation (12) can now be written as

$$\mathbf{D}(\lambda_r,\tau_{nm})\mathbf{P}_r = \mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r. \tag{15}$$

In order to solve Eq. (15), consider the ratio

$$\frac{\mathbf{P}_r^*\mathbf{D}(\lambda_r,\tau_{nm})\mathbf{P}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{P}_r} = \frac{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{P}_r}, \tag{16}$$

where $\mathbf{P}_r^*$ is the complex conjugate of $\mathbf{P}_r$. However, $\mathbf{P}_r$ and $\mathbf{A}_r$ are related to Eq. (14), so Eq. (16) can be written as

$$\begin{aligned}\frac{\mathbf{P}_r^*\mathbf{D}(\lambda_r,\tau_{nm})\mathbf{P}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{P}_r} &= \frac{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})(\mathbf{A}_r+\Delta\mathbf{A}_r)/\Delta\lambda_r} \\ &= \Delta\lambda_r\frac{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})(\mathbf{A}_r+\Delta\mathbf{A}_r)} \\ &= \Delta\lambda_r\frac{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r + \mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\Delta\mathbf{A}_r} \\ &= \Delta\lambda_r\left(1 - \frac{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\Delta\mathbf{A}_r}{P_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{A}_r} + \cdots\right).\end{aligned} \tag{17}$$

Hence, $\Delta\lambda_r$ can be approximated as

$$\Delta\lambda_r \approx \frac{\mathbf{P}_r^*\mathbf{D}(\lambda_r,\tau_{nm})\mathbf{P}_r}{\mathbf{P}_r^*\mathbf{E}(\lambda_r,\tau_{nm})\mathbf{P}_r}. \tag{18}$$

So the eigenvalues, λ_t, and eigenvectors, $\mathbf{A}_t$, for the time delay problem can be calculated as follows.

(1) Obtain the eigenvalues, λ_r, and eigenvectors, $\mathbf{A}_r$, assuming that time delays $\tau_{nm}=0$. This can be done using standard techniques described in Ref. 25.
(2) λ_r and $\mathbf{A}_r$ are good initial guesses for the eigenvalues and eigenvectors of the system with time delay. We let $\lambda_t'=\lambda_r$ and $\mathbf{A}_t'=\mathbf{A}_r$, where λ_t' and $\mathbf{A}_t'$ are the initial guesses for λ_t and $\mathbf{A}_t$.
(3) Construct the matrix $\mathbf{E}(\lambda_t', \tau_{nm})$ [see Eq. (13)] from λ_t'.
(4) Solve Eq. (15) to obtain P_r.
(5) Use Eq. (18) to obtain an estimate for $\Delta\lambda_r$. The new value for λ_t' can now be calculated to be $\lambda_r+\Delta\lambda_r$.
(6) Use Eq. (14) to calculate a new estimate for the eigenvector $\mathbf{A}_t'=\mathbf{A}_r+\Delta\mathbf{A}_r$.
(7) Go back to step 3 with the new values of λ_t' and $\mathbf{A}_t'$ until the values of λ_t' and $\mathbf{A}_t'$ do not change anymore. Once the solution is converged, $\lambda_t=\lambda_t'$ and $\mathbf{A}_t=\mathbf{A}_t'$.

From our numerical calculations, it was found that, in general, for small values of D/R_0, where D is the spacing between the centers of two adjacent bubbles, the eigenvalues of the time delay system occur in distinct complex conjugate pairs for small values of N. For larger values of N, the eigenvalues move closer together. Sometimes, in the numerical iterations, it is possible to converge to the same eigenvalue even though different starting guesses were used. To overcome this problem, the eigenvectors and eigenvalues are first obtained with small values of D (only in the time delay term). Then, these eigenvalues and eigenvectors are used as guesses for the system with the desired value of D.

Once the eigenvalues and eigenvectors of the time delay system are obtained, the solution in the physical domain can be constructed by a linear combination of $\mathbf{A}_r$ and corresponding λ_r,

$$\mathbf{x}(t) = \sum_{n=1}^{N} \beta_n\mathbf{A}_{n,t}e^{\lambda_{n,t}t}, \tag{19}$$

where β_n are constants to be determined from the initial conditions.[19] At time $t=0$, it will be assumed that

$$\mathbf{x}(t=0) = \begin{Bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}. \tag{20}$$

It will also be assumed that

$$\frac{d\mathbf{x}}{dt}(t=0) = \mathbf{0}.$$

This is only for the sake of convenience because it is not known what values to use at $t=0$. The correct values to use for $d\mathbf{x}/dt(t=0)$ and $\mathbf{x}(t=0)$ require extensive studies of the bubble wall oscillations as the bubble is formed and initially perturbed. Neither experimental nor numerical data on the magnitudes of these naturally initiated oscillations are easy to obtain; this is a subject of active research (see, for example, Refs. 27–31) that is beyond the scope of the current manuscript.

III. RESULTS AND DISCUSSION

The results will be divided into two sections. Section III A investigates the effects of adding time delays to the mathematical model. The bubbles that were considered had a radius of 2 mm, since millimeter-sized bubbles are usually excited only to the small amplitudes for which the linear theory presented above is valid. Results from simulations carried out using the model both with and without time delays are presented in Sec. III B. The theoretical data are compared with experimental data in order to ascertain if time delays are needed in order to better represent the physics of the problem.

A. Analysis of the eigenvalues

As a reference, it would be instructive to first compare the effects of time delay for the case when there are just two bubbles in the chain. For this case, the higher order $n=2$ mode corresponds to the case when the bubbles oscillate 180° out of phase with each other. The $n=1$ mode corresponds to the situation when the two bubbles oscillate in phase with each other. Figure 2 shows the damping and frequency plots if time delays are not taken into account. The distance, D, between the two bubbles is normalized by $\lambda = c/f_0$, which is the wavelength associated with the natural frequency $f_0 = \omega_0/(2\pi)$ of an isolated bubble. The calculated mode damping, ξ, and frequency, f, are normalized according to Eq. (7).

As can be seen, the highest frequency ($n=2$) mode has the highest values of damping as is found in many natural oscillatory systems. An extension of this work to investigate how the natural frequencies of a bubble system change in the presence of a wall have been reported by Payne *et al.*[32] It is also interesting to note that there is no crossing over of the frequency and the damping, i.e., for all values of D, the $n=2$ mode always has a higher frequency and damping than the $n=1$ mode. When the time delay is taken into consideration, the results are shown in Fig. 3. There are clear differences with the information shown in Fig. 2. Most obviously, the curves for the $n=1$ and the $n=2$ modes cross over at distinct values of D/λ. The $n=1$ and $n=2$ curves for ξ^* cross over at $D/\lambda=(k+1)/2$ where $k=0,1,2,\ldots$, and the curves for f^* cross over when $D/\lambda=(2k+1)/4$ where $k=0,1,2,\ldots$. When the bubbles are close together [i.e., when $D/\lambda < (1/4)$], the damping for the $n=1$ mode is greater than the damping of the $n=2$ mode but the natural frequency for the $n=2$ mode is greater than the $n=1$ mode. This is in stark contrast to the results when the time delay is not taken into account. When $(1/4) < D/\lambda < (2/4)$, there is a crossing over of the natural frequencies to create a situation where the $n=1$ mode has a higher frequency than the $n=2$ mode and the damping of the $n=1$ mode is still higher than the $n=2$ mode.

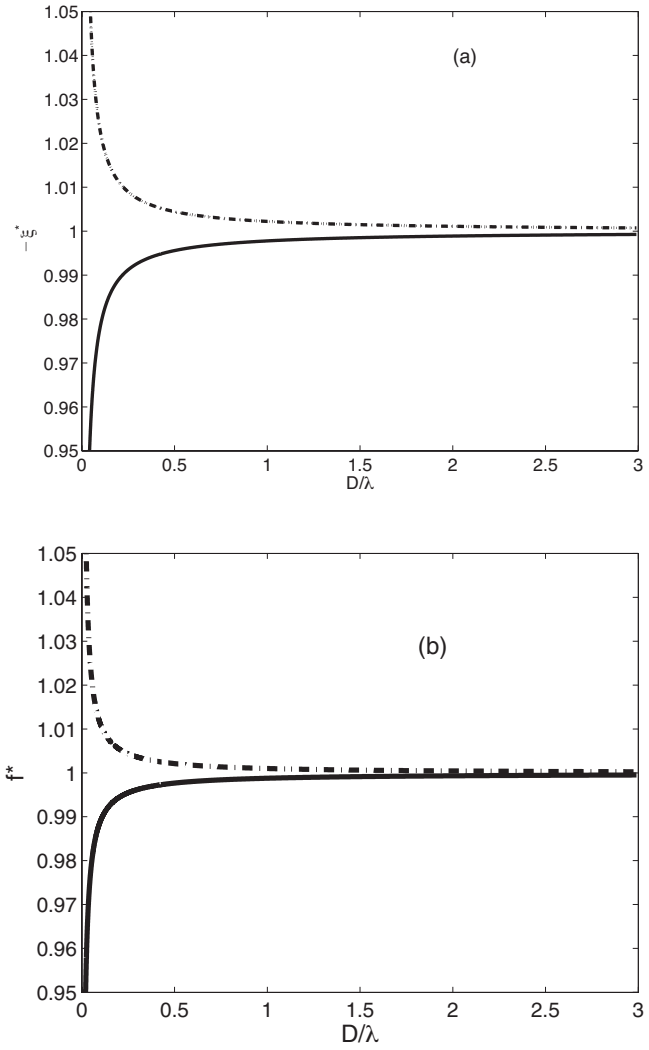


FIG. 2. Plot of $-\xi^*$ (a) and f^* (b) for the different modes predicted by the model without time delay. (—) $n=1$; (–·–) $n=2$. There are only two bubbles in the system and $R_0=2$ mm.

To understand the effects of time delay on the damping of a two-bubble system, consider the following coupled set of delay differential equations:

$$\ddot{x}_1(t) + \omega_0 \delta \dot{x}_1(t) + \omega_0^2 x_1(t) = -\frac{R_0}{D}\ddot{x}_2(t-\tau), \tag{21}$$

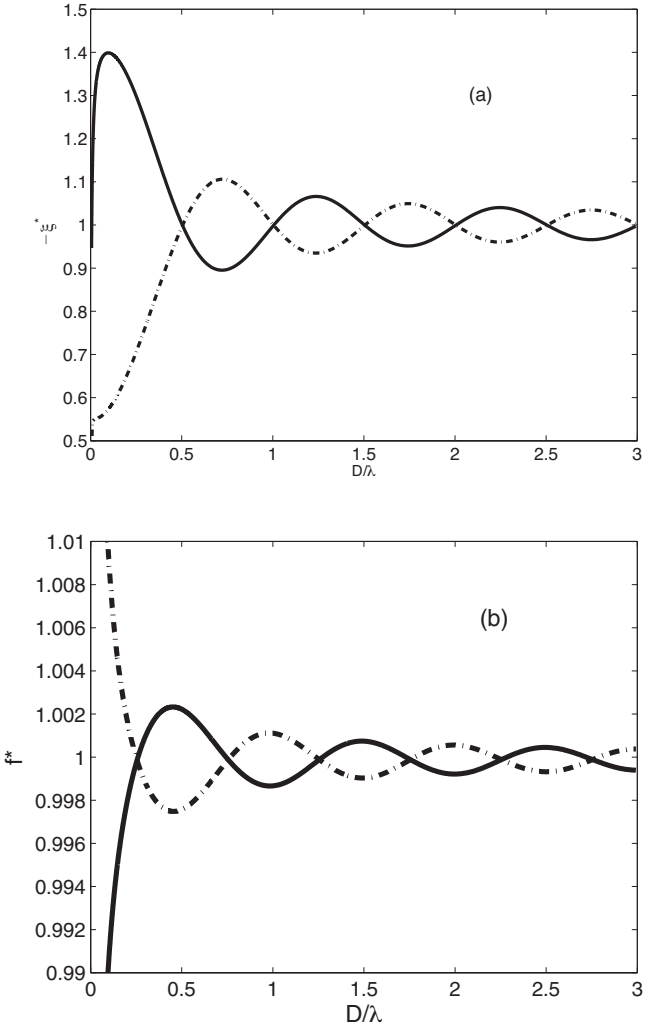


FIG. 3. Plot of $-\xi^*$ (a) and f^* (b) for the different modes predicted by the model with time delay. (—) $n=1$; (–·–) $n=2$. There are only two bubbles in the system and $R_0=2$ mm. Note the different y-axis scales.

$$\ddot{x}_2(t) + \omega_0\delta\dot{x}_2(t) + \omega_0^2 x_2(t) = -\frac{R_0}{D}\ddot{x}_1(t-\tau). \tag{22}$$

Both bubbles are of the same equilibrium radius R_0 separated by a distance D. The time delay is given by $\tau=D/c$. If we assume that τ is small, then using a Taylor series, the time delay term in Eq. (21) can be approximated as

$$\ddot{x}_2(t-\tau) \approx \ddot{x}_2(t) - \tau\dddot{x}_2(t). \tag{23}$$

Note that this approximation is only made for the purpose of the analysis in this section. Equation (23) is not used to obtain the results in Fig. 3. It is possible to obtain an expression for $\dddot{x}_2(t)$ by differentiating Eq. (22) as follows:

$$\dddot{x}_2(t) = -\omega_0\delta\ddot{x}_2(t) - \omega_0^2\dot{x}_2(t) - \frac{R_0}{D}\dddot{x}_1(t-\tau).$$

Substituting Eq. (23) into Eq. (21) gives

$$\ddot{x}_1(t) + \frac{R_0}{D}(1+\omega_0\delta\tau)\ddot{x}_2 + \omega_0\delta\dot{x}_1(t) + \frac{R_0}{D}\omega_0^2\tau\dot{x}_2(t) + \omega_0^2 x_1(t) + \left(\frac{R_0}{D}\right)^2\tau\dddot{x}_1(t-\tau) = 0. \tag{24}$$

Performing a similar exercise on the time delay term on the right hand side of Eq. (22) gives

$$\frac{R_0}{D}(1+\omega_0\delta\tau)\ddot{x}_1(t) + \ddot{x}_2 + \frac{R_0}{D}\omega_0^2\tau\dot{x}_1(t) + \omega_0\delta\dot{x}_2(t) + \omega_0^2 x_2(t) + \left(\frac{R_0}{D}\right)^2\tau\dddot{x}_2(t-\tau) = 0. \tag{25}$$

The coefficient of the last term of Eqs. (24) and (25), $(R_0/D)^2\tau$, is usually small so it is possible to ignore the (third derivative) time delay term on the right hand side of Eqs. (24) and (25). This will give us a coupled set of ordinary differential equations, which can be written in matrix form as

$$\begin{bmatrix} 1 & \frac{R_0}{D}(1+\omega_0\delta\tau) \\ \frac{R_0}{D}(1+\omega_0\delta\tau) & 1 \end{bmatrix}\begin{pmatrix}\ddot{x}_1(t)\\ \ddot{x}_2(t)\end{pmatrix} + \omega_0\delta\begin{bmatrix} 1 & \frac{R_0\omega_0}{D\delta}\tau \\ \frac{R_0\omega_0}{D\delta}\tau & 1 \end{bmatrix}\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{pmatrix} + \begin{bmatrix}\omega_0^2 & 0\\ 0 & \omega_0^2\end{bmatrix}\begin{pmatrix}x_1(t)\\ x_2(t)\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}. \tag{26}$$

Equation (26) will result in a quadratic eigenvalue problem, which can be studied using conventional techniques (see Ref. 25). Since the damping is predominantly influenced by the matrix of the first derivative term, it is clear that the effects of time delay are to increase the overall damping of the system. This influence can be quantified by the parameter $R_0\omega_0\tau/D\delta$. Equation (26) also suggests that when the damping is small (i.e., when $\omega_0\delta \ll 1$), the effect of the time delay is more significant on the overall system's damping than on its frequency. Moreover, time delay apparently introduces an additional form of damping [see Eq. (26)] that is independent of the individual-bubble damping δ. A similar but more thorough exposition of this analysis can be found in the paper by Doinikov *et al.*[22]

It is possible to think of the physical effects of coupling with time delay as introducing an additional source of radiation "damping," which could either increase or decrease the net system damping. When time delay introduces a phase lag, the energy transferred to another bubble could be either more or less rapidly dissipated, depending on the location of the other bubble with respect to the phase of the traveling wave. The cyclical variation of the damping is also found by Feuillade[19] and a very similar explanation is given in that paper.

Kapodistrias and Dahl[33] carried out experimental measurements of the backscattering of sound from a two-bubble

system in water. Bubbles with a radius of 585 μm were used in their experiments and they were excited at frequencies between 80 and 140 kHz. The distances between the bubbles were varied between 1.2 and 70 mm and they reported that "for $D/\lambda \approx 1/2$, the backscattered radiation is maximized, while for $D/\lambda < 1/2$ the backscattered radiation is reduced considerably." This observation could be explained by considering the information in Fig. 3. While this figure was created for bubbles of 2 mm radii, a very similar graph can be generated for bubbles with radii 585 μm. If we assume that the symmetric mode is the dominant mode in the experiments, then Fig. 3 shows that when $D/\lambda \approx 1/2$ the damping is quite small compared to when $D/\lambda < 1/2$. This would lead to higher values of backscattered radiation for $D/\lambda \approx 1/2$ when compared to the data for $D/\lambda < 1/2$. Kapodistrias and Dahl[33] used a multiple scattering approach to explain this variation in scattered acoustic energy as a function of the ratio of the spacing of two bubbles to the sound wavelength. Here, we have shown that the self-consistent coupled-oscillator model with time delays can predict a similar dependence on the bubble spacing to wavelength ratio in a multiple bubble system.

Natural frequencies and damping plots for the situation for ten bubbles ($N=10$) are shown in Figs. 4 (no time delay) and 5 (with time delays). Similar to the two-bubble ($N=2$) case, Fig. 4 shows that the curves for damping and natural frequencies for the model without time delays do not cross over. This indicates that the natural frequencies and damping corresponding to the $n=1$ mode will always be smaller than the damping and natural frequencies for the higher modes. This situation seems to be independent of the separation D of the bubbles in the chain. When time delays are taken into account, Fig. 5 shows that similarly to the $N=2$ case, the curves for the ξ^* cross over at $D/\lambda=(k+1)/2$ and the curves for f^* cross over when $D/\lambda=(2k+1)/4$ where $k=0,1,2,\ldots$. Similar to the two-bubble case, the results here for the time delay model can be explained when the incident wave emanated from the neighboring bubble is assumed to be a traveling wave. The ξ^* plots of the time delay model for the $N=2$ and $N=10$ cases [Figs. 3(a) and 5(a)] show that the maximum and minimum damping values occur because the phase of the incident wave is changed, i.e., depending on its phase, the incident wave can suppress or increase the oscillation mode of the bubble.

One way of analyzing the effects of time delay on the system is to plot the ratios of the natural frequencies,

$$\alpha = f_{\text{time delay}}/f_{\text{no time delay}}, \tag{27}$$

and damping

$$\beta = \xi_{\text{time delay}}/\xi_{\text{no time delay}}, \tag{28}$$

for both models at particular values of D/λ. Plots of α for $N=2$ and $N=10$ are shown in Fig. 6 and similar plots for β are shown in Fig. 7. It is clear that adding time delays do not have much effect on the natural frequencies of the system. The maximum variation for the $N=2$ chain is only about 0.5% and the $N=10$ chain has a maximum variation of less than 3%. Thus, there are only very small variations in the natural frequencies even when there are more bubbles in the chain. It is also clear that, in general, adding time delays increases the natural frequency of the lowest ($n=1$) mode but decreases the frequency of the highest ($n=N$) mode. In terms of damping, Fig. 7 shows that time delays have a dramatic effect on the damping of the individual modes. For $N=2$, there is almost a 40% increase in damping for the $n=1$ mode for small values of D/λ. The $n=2$ mode damping decreases by about 45% when time delays are taken into account. When N is increased to 10, Fig. 7(b) shows that the effects of time delays are even larger. The reason why time delay has a greater effect on system damping than on the natural frequencies is explained with Eq. (26) and the corresponding discussion on page 16.

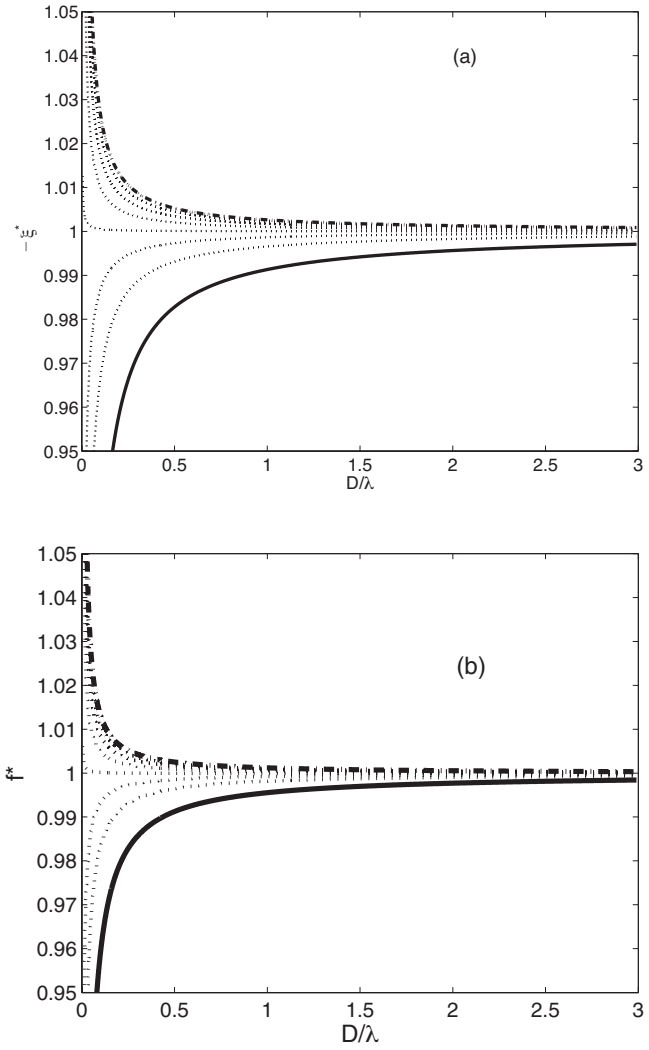


FIG. 4. Plot of $-\xi^*$ (a) and f^* (b) for the different modes predicted by the model without time delay. (—) $n=1$; (– –) $n=10$; (···) all intermediate modes. The analyzed system consists of an equally spaced bubble chain consisting of ten bubbles with equilibrium radii, $R_0=2$ mm.

B. Comparison with experimental data

In order to assess the importance of time delays, predicted data from the theoretical models will be compared with experimental data from Nikolovska *et al.*,[21] who carried out a high-spatial resolution experimental investigation on the evolution of the acoustic energy along the bubble chain. The experimental principles can be found in Ref. 20 and full

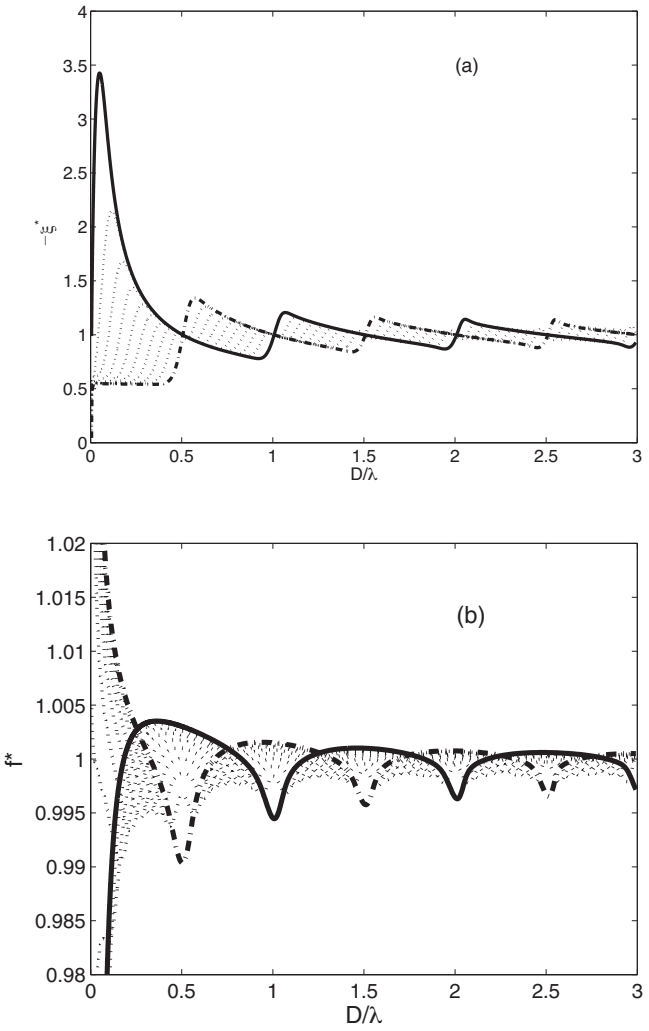


FIG. 5. Plot of $-\xi^*$ (a) and f^* (b) for the different modes predicted by the model with time delay. (—) $n=1$; (–·–) $n=10$; (···) all intermediate modes. The analyzed system consists of an equally spaced bubble chain consisting of ten bubbles with equilibrium radii, $R_0=2$ mm. Note the different y-axis scales.

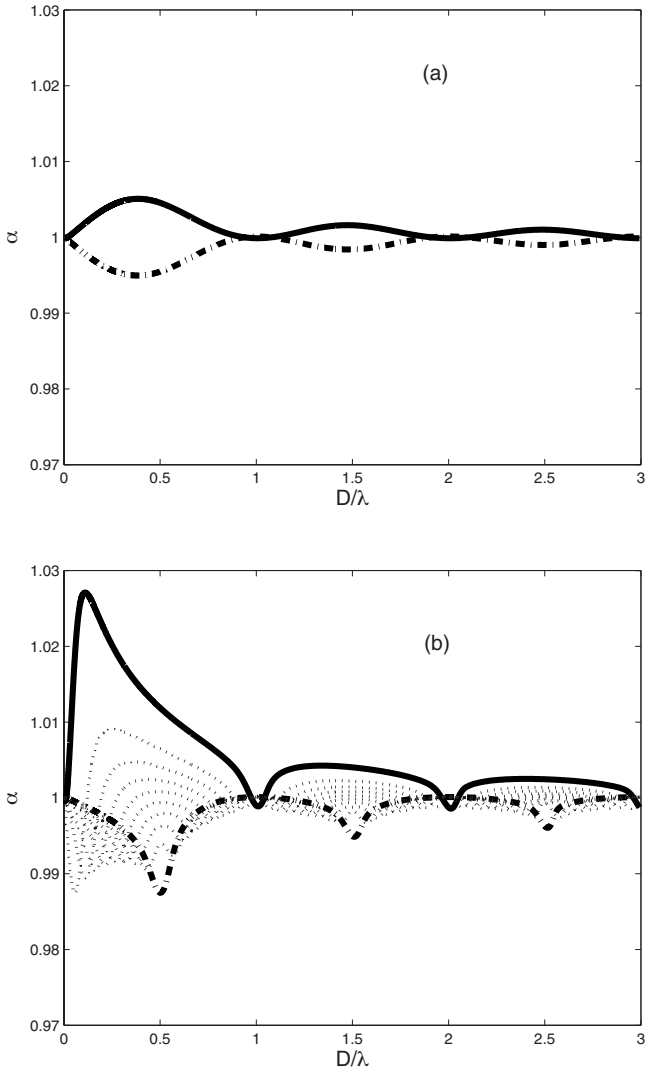
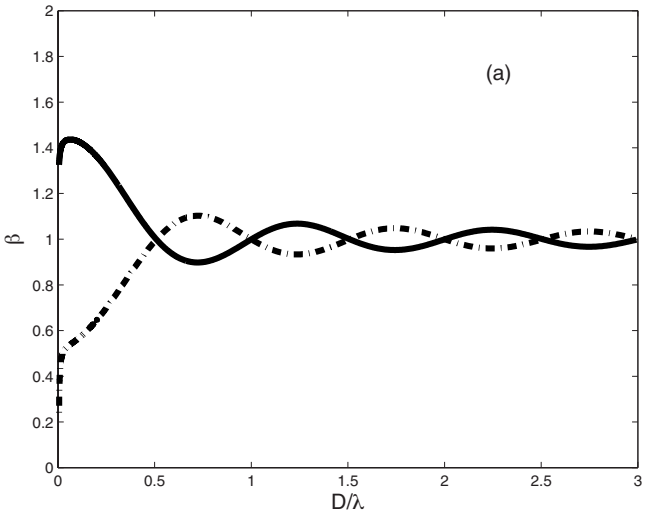


FIG. 6. Ratio of the natural frequencies, α, for $N=2$ (a) and $N=10$ (b). (—) $n=1$; (–·–) $n=N$; (···) all intermediate modes. The analyzed system consists of an equally spaced bubble chain consisting of bubbles with equilibrium radius, $R_0=2$ mm.

details of the high-resolution experiments can be found in Ref. 21. Air bubbles were introduced into a tank with a 1 mm radius nozzle. Depending on the bubble production rate (BPR), the bubbles generated from this nozzle had radii between 1 and 2.35 mm (see Table I). Photographic images similar to that shown in Fig. 1 were used to determine the size of the bubbles. The comparison with theoretical predictions will be made along a vertical line, which is 6 cm from the nozzle. Hydrophones were used to record data at 30 kHz. The total period of data acquisition, T, was $1024/30\,000 = 0.0341$ s ≈ 34 ms. The distance between bubbles in the chain ranged from 36 mm at the lowest BPR to 6.5 mm at the highest BPR. In this system, most of the acoustic energy is generated when the bottom bubble detaches from the nozzle. In an earlier study, Manasseh *et al.*[20] showed that there is an anisotropic distribution of acoustic energy in the vicinity of the bubble chain. Experimental data show that the rms value of pressure dies off much faster as we move away from the chain than along the bubble chain, indicating that the transfer of acoustic energy is more efficient along the bubble chain.

Instantaneous snapshots from the experimental data of the acoustic pressure profile are shown in Fig. 8 at time instances $t/T_0=7.2$, 7.6, 7.9, and 8.3, where T_0 is the Minnaert period of a single isolated bubble. It is clear that there is a preference for the propagation of acoustic energy along the bubble chain. In order to compute the propagation speed along the bubble chain, V_p, a numerical algorithm was developed to detect and track the local maximum of the pressure profile (peak pressure) measured from the experimental data. The location of the local pressure maxima found by the algorithm is indicated by ∘ in Fig. 8. By following these ∘ symbols, it is possible to calculate V_p. When the peak of the acoustic pressure wave leaves the domain, the algorithm would detect and follow another peak from the bottom of the bubble chain.

The results are shown in Fig. 9. The vertical axis is the location of the peak pressure along the bubble chain and the horizontal axis shows the corresponding value of V_p normalized by the speed of sound in water, c. Only data from the 20 Hz BPR case are shown. Data from all other cases show



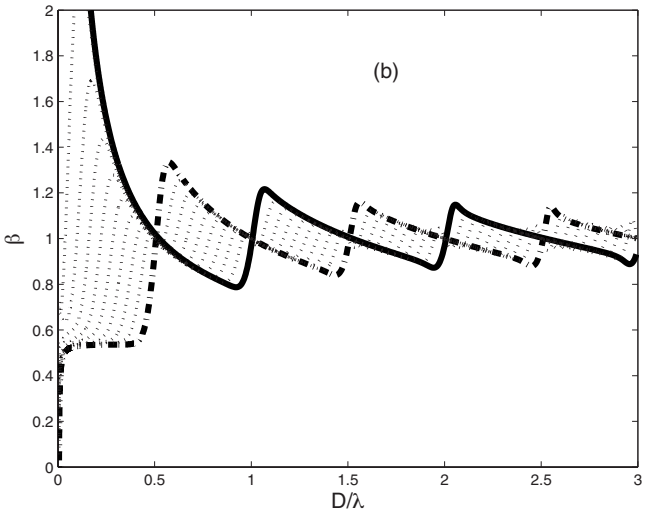


FIG. 7. Ratio of the damping, β, for $N=2$ (a) and $N=10$ (b). (—) $n=1$; (–·–) $n=N$; (···) all intermediate modes. The analyzed system consists of an equally spaced bubble chain consisting of bubbles with equilibrium radius, $R_0=2$ mm.

similar trends. Postprocessing of the experimental data shows that V_p is smaller at the bottom of the bubble chain and increases slightly further up the bubble chain [see Fig. 9(a)]. In general, V_p calculated from the experimental data is usually smaller than c. However, there are a small number of instances where V_p calculated from the experimental data exceeds c, which is due to the noise in the experimental data. Data from predictions using the coupled-oscillator model without time delays are shown in Fig. 9(b). There are large variations in V_p as the acoustic pressure moves up the bubble chain. When the pressure peak is at the bottom of the chain, predicted values of V_p are usually smaller than c. As the pressure peak moves up the chain, predicted values of V_p increase to nearly $3c$. This is unrealistic and can be explained by recalling that Eq. (2) assumes that any disturbance from a neighboring bubble immediately affects (travels at an infinite speed to) a nearby bubble. Thus, it is unsurprising that coupling these equations without time delays will produce an estimate of V_p, which is much larger than c. This anomaly is overcome by incorporating time delays into the coupled-oscillator model, as shown in Fig. 9(c). The predicted values of V_p are much more reasonable, closer to the values of V_p calculated from experimental data [compare Fig. 9(a) with Fig. 9(c)]. The coupled-oscillator model with time delays also predicts a smaller variation of V_p as the pressure peak moves up the bubble chain, again consistent with V_p calculated using experimental data.

TABLE I. Parameters from experimental studies. Data are from the 1 mm nozzle.

BPR (Hz)	Frequency (kHz)	Radius (mm)	D (mm)	N
10	1.97	1.18	36	40
12	1.90	1.35	32.8	48
14	1.88	1.41	30.0	56
18	1.8	1.50	26.6	72
20	1.79	1.52	19.7	80
22	1.68	1.55	17.0	88
24	1.52	1.64	16.0	96
26	1.21	1.70	14.0	104
29	1.15	1.88	10.3	116
31	1.06	2.00	9.2	124
34	0.91	2.10	7.5	136
38	0.84	2.35	6.5	152

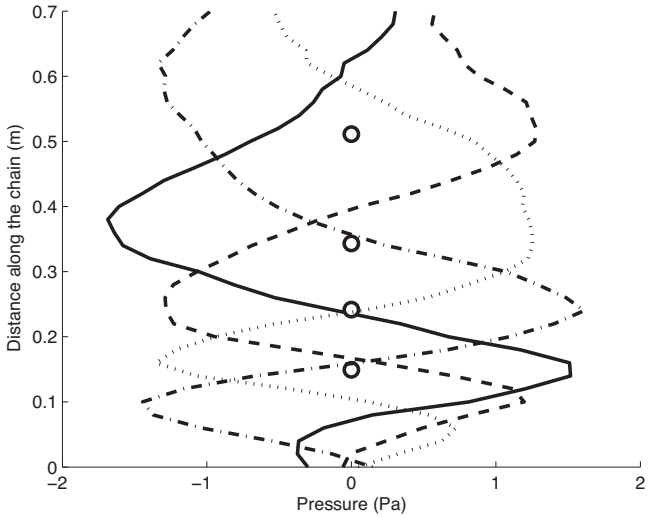


FIG. 8. Vertical profiles of instantaneous pressure at $t/T_0=7.2$ (—), 7.6 (–·–), 7.9 (···), 8.3 (----). The ○ in the figure marks the positions of the numerically predicted local pressure maxima.

From the experimental and numerical data, an observation that it was not possible to follow the pressure peak throughout the domain for $t/T_0<5$ and $t/T>1/4$ where $T=1024/30\,000=0.034$ s=34 ms is the total time period of data acquisition in the experiments was made. Usually, $t/T>1/4$ would correspond to approximately $t/T_0>20$. For $t/T_0<5$, the peak of the the acoustic pressure profile flattens out as it reaches the top of the domain, which makes the maximum difficult to detect numerically. On the other hand, for $t/T>1/4$, the signal decays to essentially zero, thus any local peak in the data would be almost undetectable. So it is only possible to calculate the average values of V_p for $T_0/5<t<(1/4)T$ and the results are shown in Fig. 10. At a smaller BPR, there are large discrepancies between the

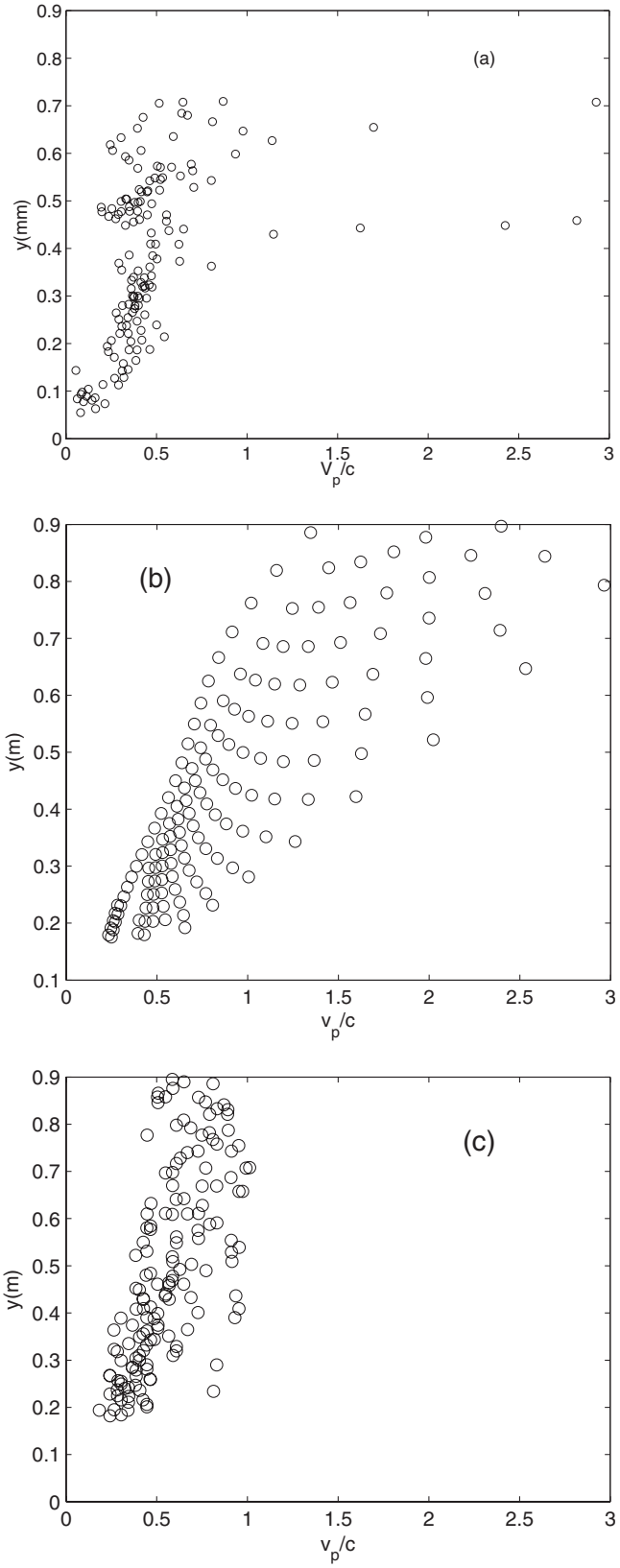


FIG. 9. Comparing the predicted phase velocity V_p/c of the different models with experimental data at BPR=20 Hz. (a) is the experimental data, (b) model without time delays, and (c) model with time delays.

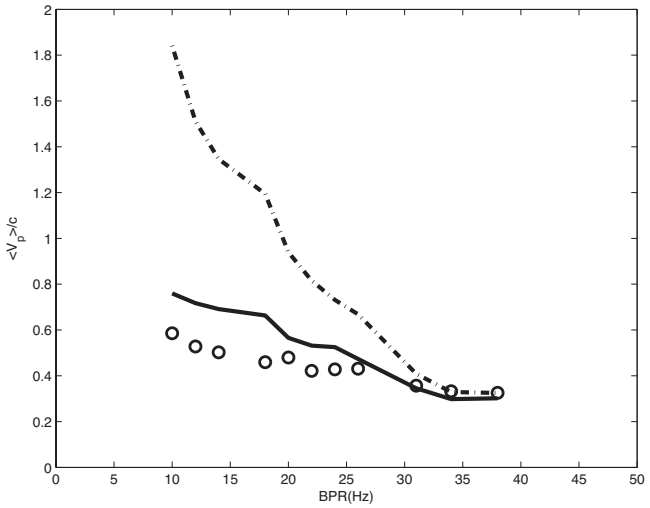


FIG. 10. Comparing the average phase velocity $\langle V_p \rangle$, normalized by the speed of sound in water c for experimental data at different bubble production rates (BPRs). (○) Experimental data; (–·–) numerical model with no time delays; (—) numerical model with time delays.

model without time delay and the model with time delay. As noted above, there is better agreement between the experimental data and the mathematical model when time delays are taken into account. As we increase the BPR, there are more bubbles in the chain and the distance between bubbles in the chain becomes smaller. Hence, the effects of time delays are reduced. This is one possible reason why Fig. 10 shows that there is good agreement with experimental data for both models at a large BPR.

The eigenvalue results (Fig. 7) had showed that time delays had an increasing effect on the damping for larger values of N. However, Fig. 9 shows time delays had a decreasing effect on the propagation speed as N (BPR) increased. This is probably because different physical phenomena are responsible for overall system damping and propagation speed. As N increases, the energy transfer to all other bubbles (affecting overall dissipation) is enhanced because the coupling becomes stronger as bubbles become closer. This increases the relevance of time delays, which provide a mechanism for altered dissipation as explained on page 18. However, as N increases, the lag in information transfer to neighboring bubbles (affecting propagation speed) is decreased. This reduces the relevance of time delays.

From Fig. 10, the time delay model produces a better comparison on propagation speeds of the non-time-delay model, particularly when bubbles are far apart (low BPR). In contrast, the comparison between experimental and numerical data for the distribution of rms pressure, P_{rms} (Fig. 11), does not suggest a clear superiority of one model over another. In calculating P_{rms} for Fig. 11, the averaging is done over $0<t<T$. Only data from four BPRs are shown but data from all other BPRs show the same trends. Farther from the bubble generation point (i.e., at larger y), the time delay model generates pressures that are closer to the experimental data than the non-time-delay model. Pressures in the time delay model are lower than in the non-time-delay model, mainly due to the higher damping of the modes when time delays are taken into account [compare data in Figs. 4(a) and 5(a)]. As the acoustic wave travels up the bubble chain, the incorporation of time delays produces an enhanced damping effect on the wave, which reduces the predicted values of

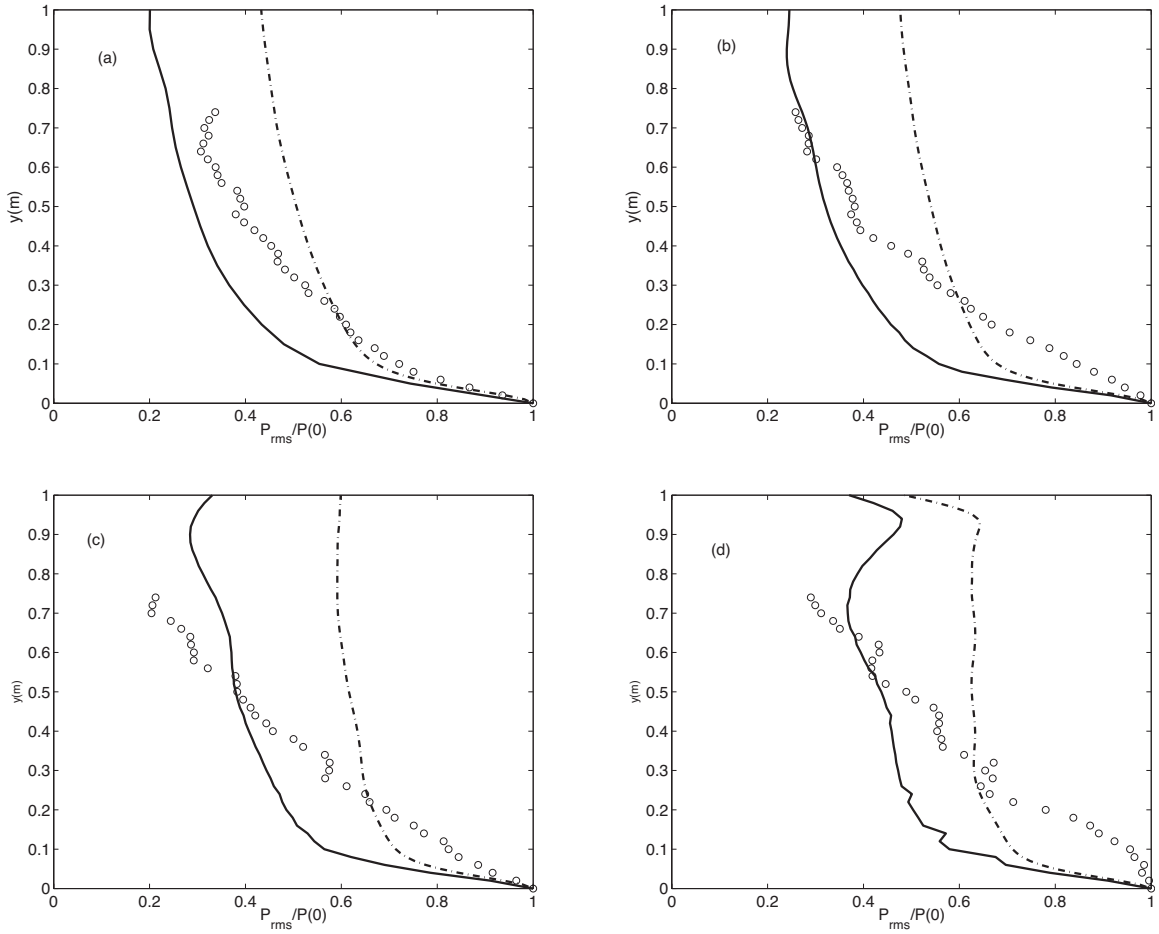


FIG. 11. Comparison of the rms pressure for different bubble production rates (BPRs). (a) BPR=22 Hz, (b) BPR=26 Hz, (c) BPR=31 Hz, and (d) BPR =38 Hz. (○) Experimental data; (—) predictions using the model with time delays; (–·–) predictions using numerical model without time delay.

P_{rms} further up the bubble chain. However, the trends, in general, do not match the experimental data, particularly at a higher BPR.

The discrepancies between the experimental data and theoretical models could be due to a number of factors. Clearly, the shape of the bubbles is not constant (Fig. 1), and this is known to "detune" the resonant frequency ω_0.[34] The spacing between the bubbles is also not a constant but varies as the bubble transitions from its initial vertical trajectory to a spiral trajectory.[35,36] Perhaps the most significant issue is that both theoretical models assume that the ratio R_0/D is very small (see discussion on page 6). In the experiments, R_0/D ranges from about 0.09 to approximately 0.36. These values are not very small, and the approximation worsens for a high BPR, consistent with the worsening of the comparison with experiment at a high BPR.

IV. CONCLUSIONS

An investigation on the effects of time delays in the coupled-oscillator model has been carried out. The models have been applied to the case of a bubble chain and results from the analysis show that the main effects of time delays are to increase the amount damping of the lowest mode and decrease the amount of damping in the highest mode. If there are no time delays, the lowest frequency mode has the smallest amount of damping. When time delays are taken into account, the situation is reversed with the highest damping occurring at the lowest frequency. This study also shows that time delays do not have a significant effect on the natural frequencies of each mode. However, the addition of time delays can have a major impact on the damping of each individual eigenmode.

Previous investigations have shown that there is a preference for acoustic energy to propagate along a bubble chain. It has also been reported that the average speed of propogation of acoustic energy is much smaller than the speed of sound in water. Calculations conducted without taking time delay into consideration show propagation speeds much faster than the speed of sound in water. Time delays reduce the speed of propagation and predict propagation speeds much closer to experimental data.

ACKNOWLEDGMENT

We would like to thank. A. Doinikov, from Byelorussian State University, Belarus, for introducing us to the idea of time delay and many fruitful discussions.

[1]T. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[2]E. A. Zabolotskaya, "Interaction of gas bubbles in a sound field," Sov. Phys. Acoust. **30**, 365–368 (1984).
[3]I. Tolstoy, "Superresonant systems of scatterers," J. Acoust. Soc. Am. **80**, 282–294 (1986).
[4]H. Ogũz and A. Prosperetti, "A generalization of the impulse and virial

theorems with an application to bubble oscillations," J. Fluid Mech. **218**, 143–162 (1990).
[5]A. A. Doinikov and S. T. Zavtrak, "On the mutual interaction of two gas bubbles in a sound field," Phys. Fluids **7**, 1923–1930 (1995).
[6]C. Feuillade, "Acoustically coupled gas bubbles in fluids: Time-domain phenomena," J. Acoust. Soc. Am. **109**, 2606–2615 (2001).
[7]P.-Y. Hsaio, M. Devaud, and J.-C. Bacri, "Acoustic coupling between two air bubbles in water," Eur. Phys. J. D **4**, 5–10 (2001).
[8]M. Ida, "A characteristic frequency of two mutually interacting gas bubbles in an acoustic field," Phys. Lett. A **297**, 210–217 (2002).
[9]M. Ida, "Number of transition frequencies of a system containing an arbitrary number of gas bubbles," J. Phys. Soc. Jpn. **71**, 1214–1217 (2002).
[10]M. Ida, "Avoided crossings in three coupled oscillators as a model system of acoustic bubbles," Phys. Rev. E **72**, 036306 (2005).
[11]K. W. Commander and A. Prosperetti, "Linear pressure waves in bubbly liquids: Comparison between theory and experiments," J. Acoust. Soc. Am. **85**, 732–746 (1989).
[12]R. Duraiswami, S. Prabhukumar, and G. L. Chahine, "Bubble counting using an inverse scattering method," J. Acoust. Soc. Am. **104**, 2699–2717 (1998).
[13]M. Nicholas, R. A. Roy, L. A. Crum, H. N. Ogũz, and A. Prosperetti, "Sound emission by a laboratory bubble cloud," J. Acoust. Soc. Am. **95**, 3171–3182 (1994).
[14]A. D. Phelps, D. G. Ramble, and T. G. Leighton, "The use of a combination frequency technique to measure the surf zone bubble population," J. Acoust. Soc. Am. **101**, 1981–1989 (1996).
[15]E. J. Terrill and K. W. Melville, "A broadband acoustic technique for measuring bubble size distributions: laboratory and shallow water measurements," J. Atmos. Ocean. Technol. **17**, 220–239 (2000).
[16]A. B. Pandit, J. J. Varley, R. B. Thorpe, and J. F. Davidson, "Measurement of bubble size distribution: An acoustic technique," Chem. Eng. Sci. **47**, 1079–1089 (1992).
[17]J. W. R. Boyd and J. Varley, "The uses of passive measurement of acoustic emissions from chemical engineering processes," Chem. Eng. Sci. **56**, 1749–1767 (2001).
[18]R. Manasseh, R. F. LaFontaine, J. Davy, I. C. Shepherd, and Y. Zhu, "Passive acoustic bubble sizing in sparged systems," Exp. Fluids **30**, 672–682 (2001).
[19]C. Feuillade, "Scattering from collective modes of air bubbles in water and the physical mechanism of superresonances," J. Acoust. Soc. Am. **117**, 1178–1190 (1995).
[20]R. Manasseh, A. Nikolovska, A. Ooi, and S. Yoshida, "Anisotropy in the sound field generated by a bubble chain," J. Sound Vib. **278**, 807–823 (2004).
[21]A. Nikolovska, R. Manasseh, and A. Ooi, "On the propagation of acoustic energy in the vicinity of a bubble chain," J. Sound Vib. **306**, 507–523 (2007).
[22]A. Doinikov, R. Manasseh, and A. Ooi, "Time delays in coupled multiple bubble systems," J. Acoust. Soc. Am. **117**, 47–50 (2005).
[23]V. Twersky, "Multiple scattering of waves and optical phenomena," J. Opt. Soc. Am. **52**, 145–171 (1962).
[24]C. Clay and H. Medwin, *Acoustical Oceanography* (Wiley, New York, 1977).
[25]F. Tisseur and K. Meerbergen, "The quadratic eigenvalue problem," SIAM Rev. **43**, 235–286 (2001).
[26]H. Hu, E. Dowell, and L. Virgin, "Stability estimation of high dimensional vibrating systems under state delay feedback control," J. Sound Vib. **214**, 497–511 (1998).
[27]M. Longuet-Higgins, B. Kerman, and K. Lunde, "The release of air bubbles from an underwater nozzle," J. Fluid Mech. **230**, 365–390 (1991).
[28]A. Prosperetti and H. Ogũz, "The impact of drops on liquid surfaces and the underwater noise of rain," Annu. Rev. Fluid Mech. **25**, 577–602 (1993).
[29]H. C. Pumphrey and P. A. Elmore, "The entrainment of bubbles by drop impacts," J. Fluid Mech. **220**, 539–567 (1990).
[30]Y. Y. Hu and B. C. Khoo, "An interface interaction method for compressible multifluids," J. Comput. Phys. **198**, 35–64 (2004).
[31]A. Bui and R. Manasseh, in A CFD Study of the Bubble Deformation During Detachment, Fifth International Conference on CFD in the Process Industries, Melbourne, Australia, 13–15 December (2006).
[32]E. Payne, S. Illesinghe, A. Ooi, and R. Manasseh, "On the resonances of bubbles attached to a rigid boundary," J. Acoust. Soc. Am. **118**, 2841–2849 (2005).
[33]G. Kapodistrias and P. Dahl, "Effects of interaction between two bubble scatterers," J. Acoust. Soc. Am. **107**, 3006–3017 (2000).
[34]M. Strasberg, "The pulsation frequency of nonspherical gas bubbles in liquid," J. Acoust. Soc. Am. **25**, 536–537 (1953).
[35]R. Clift, J. R. Grace, and M. E. Weber, *Bubbles, Drops and Particles* (Academic, London, 1978).
[36]S. Yoshida, R. Manasseh, and N. Kajio, in The Structure of Bubble Trajectories Under Continuous Sparging Conditions, Third International Conference on Multiphase Flow, Lyon, France (1998), Vol. **426**, pp. 1–8.

Tank measurements of scattering from a resin-filled fiberglass spherical shell with internal flaws

Alessandra Tesei,[a] Piero Guerrini, and Mario Zampolli
NURC, Viale San Bartolomeo 400, 19126 La Spezia, Italy

(Received 9 October 2007; revised 8 May 2008; accepted 23 May 2008)

This paper presents results of acoustic inversion and structural health monitoring achieved by means of low to midfrequency elastic scattering analysis of simple, curved objects, insonified in a water tank. Acoustic elastic scattering measurements were conducted between 15 and 100 kHz on a 60-mm-radius fiberglass spherical shell, filled with a low-shear-speed epoxy resin. Preliminary measurements were conducted also on the void shell before filling, and on a solid sphere of the same material as the filler. These data were used to estimate the constituent material parameters via acoustic inversion. The objects were measured in the backscatter direction, suspended at midwater, and insonified by a broadband directional transducer. From the inspection of the response of the solid-filled shell it was possible to detect and characterize significant inhomogeneities of the interior (air pockets), the presence of which were later confirmed by x-ray CT scan and ultrasound measurements. Elastic wave analysis and a model–data comparison study support the physical interpretation of the measurements. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945701]

PACS number(s): 43.30.Ky, 43.20.Fn, 43.20.Jr, 43.35.Zc [DF] Pages: 827–840

I. INTRODUCTION

Literature on nondestructive testing (NDT) for structural health monitoring (SHM) proposes several methods[1–5] based on the acoustic excitation and measurement of leaky surface-guided Lamb waves supported by thin plates made of metals and, more recently, with composite multilayered materials and bonded joints (see also a recent review article by Su *et al.*[6]). Acoustic inversion based on elastic scattering measurements either in air[7,8] or in water[9–11] is generally applied to thin plates of various materials, including composite laminates and plastics. Kinra *et al.*[12] presented a method for simultaneous health monitoring and inversion of the properties of the individual thin layers comprising a multilayered medium.

In this work, low- to midfrequency elastic scattering measurement and analysis are applied to water-loaded, finite objects with circular cross section (namely, solid spheres and spherical shells, either void or totally solid filled). Low- to midfrequency sound has been experimentally shown to penetrate into the metallic casing of an elastic object, and, hence, to provide information on its interior structure and content.[13,14] Here the purpose is to investigate whether elastic scattering can be significant (with respect to diffraction), and, consequently, exploitable for applications such as SHM and inversion, when the specimen to analyze is a finite object made with a dissipative, low-shear-speed material (i.e., a plasticlike material, possibly coated by a shell of composite, such as fiberglass). Unlike many recent works in NDT mentioned earlier, the goal here is the detection of internal flaws and the structural analysis of the material filling a composite shell, and not the SHM of the composite itself, which, in the selected bandwidth, is supposed to appear as a homogeneous medium.

As far as the inversion is concerned, the proposed technique takes advantage of the shape of the specimen. If compared to methodologies designed to invert the material properties of plates,[9–11,15] the insonification of an object with circular cross section allows the simultaneous excitation of all the elastic waves supported in a certain bandwidth through a single monostatic measurement, without the need for a bistatic measurement system and/or of the insonification at a fan of possible incident angles, as is generally done in the case of a flat interface. The resulting acoustic response is rich in information, but also more complicated to analyze, and, hence, requires model-based inversion.

Inversion is based on the analysis of the measured elastic echo structure, and on the application of an analytical model of scattering by layered elastic spheres. The approach can be easily extended to circular cylinders. The identification of each wave echo and the estimation of their times of arrival are exploited to feed the model with an initial guess of possible values for the bulk speeds of the material to invert. The estimate of the material bulk speeds derives from the minimization of the error between the measured and predicted scattering responses of the specimen. Similar acoustic inversion approaches, based on elastic scattering measurements of water-loaded shells with circular cross section, either void or liquid filled, can be found in the literature.[13,14] In these references, the material parameters were estimated by exploiting analytic relations with either the dispersion curves of the supported elastic waves or the periodicity of their time echoes. However, the identification of the elastic wave resonance modes in the spectral form function of the object can be nontrivial in the case of interference among closely spaced resonance modes, especially when the materials are

[a]Electronic mail: tesei@nurc.nato.int

dissipative, and, hence, many resonance modes are too weak to be detected. Moreover, the inversion methods based on the detection and identification in the time domain of a number of subsequent echoes for each wave are not applicable to dissipative media, for which only one echo per wave type is generally detectable. In these cases, a model-based approach exploiting the times of arrival of a few echoes is more widely applicable. However, being less accurate than resonance mode identification or estimation of echo periodicity, this wave analysis cannot be used to directly invert the parameters, but only to limit the search space.

Following a parametric study[16] conducted on thin-walled, totally filled spherical shells, with the purpose of investigating the effect of different fillers, acoustic measurements were performed on a fiberglass spherical shell filled with a low-shear-speed epoxy resin. Preliminary measurements were conducted also on the void shell before filling and on a solid sphere of the same material as the interior. The objects were selected to be simple enough to be treated by currently available modeling techniques, but realistic enough to give a first insight into the physics of the elastic waves present in such material combinations. The shell material is fiberglass consisting of layers of randomly distributed (hence approximately isotropic) "MAT" fiber. Acoustic data are in the range of 15–100 kHz, roughly corresponding to a *ka* range of 4–26 (with *k* being the water wave number and *a* the object radius), which is suitable for the excitation of a number of elastic waves. At these frequencies, a layer of randomly distributed fiber is assumed to be approximately isotropic and homogeneous, and sound is assumed to propagate in a multilayered composite shell as if it consisted of an homogeneous material. The same hypothesis can be made for the epoxy-resin filler, where air microbubbles can be trapped during manufacturing. Consequently, an analytical modeling tool is expected to be suitable to realistically predict the scattering response of a resin-filled fiberglass shell in a *ka* range of 4–26.[16] In this bandwidth, the elastic scattering component is expected to be significant despite the wave damping due to material dissipation.

The preliminary measurements of the solid resin sphere and of the void fiberglass shell were conducted to achieve, through acoustic inversion, an estimate of the constituent material parameters, namely of shear and compressional sound speeds and their attenuations.

Given the parameter estimates of the constituent materials, the measurement of the resin-filled fiberglass shell aimed to verify whether the elastic waves predicted by analytical models were detectable and to check its general structural health. It was also addressed to investigate whether perfect contact at the filler–shell interface (as assumed in past simulation studies[16]) was achieved during manufacturing. This analysis revealed inhomogeneities of the interior (in particular an extended air pocket having a quasisymmetric annular shape), which were later confirmed by x-ray CT scan and high-frequency ultrasound spot measurements. The air pocket was presumably caused by a local detachment of the resin filler from the shell during the solidification process. Estimation of the geometry of the main air pocket was useful for obtaining a more accurate model of the object. The AXISCAT modeling tool recently developed at NURC[17] was used to generate the backscattering response of the filled spherical shell, including the air pocket.

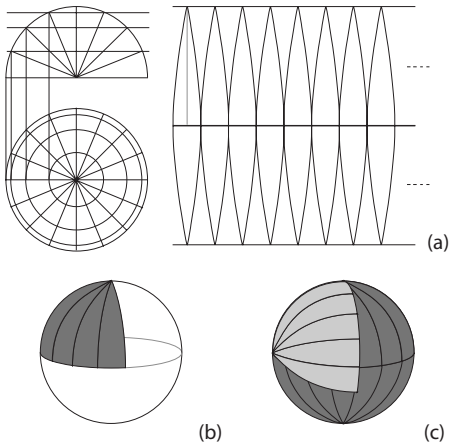


FIG. 1. Schematics of the main steps of the manufacturing process of the fiberglass shell. (a) Two-dimensional projection of a spherical surface according to the Monge method; (b) building of the first layer of the shell by positioning the spindle-shaped fiber patches (dark gray) over a glass mother shell (white); and (c) positioning of the next layer of patches (light gray) according to an orthogonal orientation with respect to the previous one.

The paper is organized as follows. Section II provides details on the manufacturing of the spherical objects, on the tank facility, and on the experiments conducted. The theoretical background for elastic wave analysis is briefly described in Sec. III, and the acoustic data are discussed in Sec. IV, along with the inversion results. A detailed analysis of the resin-filled fiberglass shell is reported in Sec. V, including additional NDT measurements and results of the model–data comparison. Conclusions are presented in Sec. VI.

II. DESCRIPTION OF THE SPHERICAL OBJECTS AND OF THE MEASUREMENT SETUP

The manufacturing constraints on the shell were: (i) to have approximately uniform thickness (within 0.5 mm over a wall thickness of 2.5 mm) and uniform radius curvature (within 1.5 mm over a nominal outer radius of 62.5 mm), and (ii) to have a material as homogeneous and isotropic as possible along the surface and through the wall thickness. To achieve this, a MAT random fiber (namely, MAT 300) was used, cut in a set of spindle-shaped patches according to the planar representation of a hemispherical surface obtained through the Monge method (or method of double orthogonal projections), as shown in Fig. 1(a). Figures 1(b) and 1(c) describe how each spherical layer of fiber was built by mosaicking the set of spindles over a thin glass shell in order to reconstruct a spherical surface by totally covering it without overlapping. At each layer, the orientation of the patches was changed [see Fig. 1(c)], in order to minimize the effect of joints at the sides of the patches themselves. After the solidification of the first layer, the mother glass shell was broken and extracted from the interior through a small hole. The sphericity was checked for layer after layer while the matrix was still liquid; at the end of the process, the sphericity was ensured by polishing. The percentage of glass with respect to resin was kept around 50%.

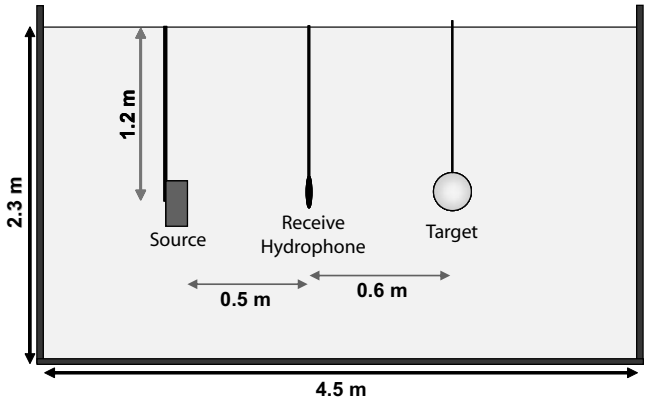


FIG. 2. Geometry of the tank measurement setup (not to scale).

The epoxy resin used as a solid filler was originally liquid; in order to maintain an isothermal chemical reaction during the curing process, it was cast layer by layer in about five steps, given the sphere size, each layer needing a predefined time to solidify. In order to build an uncoated solid sphere the resin was cast into a thin-walled glass spherical shell, which was broken and removed at the end of the process. A final smoothing ensured the sphericity also around the shell filling hole. The same step-by-step casting procedure was adopted to fill the fiberglass shell through a small hole (about 5 mm diameter) in the shell itself, which was closed at the end with some resin.

The measurements were conducted in the NURC tank, which is 4.5 m × 3 m × 2.3 m, has steel walls and bottom, and is filled with fresh water. For suspension, the objects were placed in a thin nylon wire net, fixed to the object by several resin drops (having a maximum diameter of about 10 mm and an average maximum thickness of about 0.5 mm). Whereas the filled objects were suspended through a thin nylon wire, as shown in Fig. 2, the hollow, buoyant shell was connected through a nylon wire to a lead weight placed on the floor of the tank. The source (Reson TC2138) was located at midwater at one end of the tank. Although its sensitivity is roughly flat between 40 and 100 kHz, the high signal-to-noise ratio made it possible to obtain useful data from about 15 to 100 kHz. The receiver was an omnidirectional hydrophone with a roughly flat response between 1 and 300 kHz. As shown in Fig. 2, it was located on the transmission axis between transmitter and target in such a way as to minimize the surface and bottom interference. This constraint made it necessary to locate the object at about 1.1 m from the transducer and at about 0.6 m from the hydrophone, which are distances theoretically insufficient to guarantee a plane wave insonification at frequencies lower than 50 kHz (roughly corresponding to $ka=13$), and a far field measurement of the object scattering at frequencies lower than 57 kHz ($ka \approx 15$). In the modeling phase of this study, the incident field is approximated by a plane wave for simplicity, while the actual distance between object and receiver is taken into account. Model–data comparison results follow, which show that the plane wave approximation is applicable. The directionality of the source (having 30° of beamwidth null to null at 50 kHz and sidelobes −20 dB down) allowed the complete illumination of a target without strong interferences with the tank boundaries and the water surface. The residual reverberation was mitigated by subtracting a coherent average of 20 pings of scattering from the tank boundaries. Data were coherently averaged over 20 pings and equalized in the spectral domain by using the direct measurement of the transmitted pulse on the same hydrophone. The target strength data were Hamming windowed before applying an inverse Fourier transform in order to get a smooth time response.

III. ELASTIC SCATTERING THEORY FOR LOW-SHEAR-SPEED MATERIAL SPHERICAL SHELLS

The theoretical basis for the interpretation of the scattering measurements derives from wave analysis applied to simulated far-field backscattering by fluid-loaded spheres insonified by a plane wave in the free field. The most general case considered is scattering by a solid-filled spherical shell made of possibly lossy materials. The analytical scattering model uses a standard expansion[18] for the scattered field P_{sc} into a partial wave series at range r and angle θ,

$$P_{sc}(r,\theta) = \sum_{n=0}^{\infty} i^n (2n+1) a_n h_n^{(1)}(kr) P_n(\cos(\theta - \theta_{inc})), \quad (1)$$

where h_n are the outgoing spherical Hankel functions, P_n are the Legendre polynomials, θ_{inc} is the angle of the incident plane wave, and a_n are the coefficients to be determined. These coefficients, and the coefficients of the field expansion in the solid layers of the sphere, are determined by applying the appropriate continuity conditions at the interfaces. In particular, the model assumes perfect contact at the filler–shell interface. It is convenient to determine the coefficients a_n by Cramer's rule, from which

$$a_n = B_n / D_n, \quad (2)$$

where B_n and D_n are two determinants, the elements of which are reported for example by Veksler[18] for a solid sphere, by Kargl *et al.*[19] for a void spherical shell, and can be computed for a solid-filled shell in a way similar to what is reported in these references. The backscattered field is obtained by imposing $(\theta - \theta_{inc}) = \pi$ in Eq. (1).

Wave analysis consists in the identification of the elastic waves supported by the scatterer, and in the study of their properties in terms of phase (and group) speed dispersion curves, and attenuation. This is achieved by studying the modal coefficients a_n as the wave number k, and, hence, the frequency f, varies. For a given n, the lth maximum of a_n locates the nth-order resonance frequency of the lth elastic wave, f_n^l. The corresponding wave phase speed is computed by

$$c_{ph}^l = 2\pi f_n^l a/(n+1/2), \quad (3)$$

and its group speed can be approximated by[7]

$$c_g^l \approx 2\pi (f_n^l - f_{n-1}^l) a. \quad (4)$$

The attenuation of the wave modes as frequency varies is studied through the Sommerfeld–Watson transformation.[20] A numerical method, developed for solid spheres by Williams *et al.*[21] and for void spherical shells by Kargl *et al.*[19] and

TABLE I. Material parameters.

	c_p (m/s)	c_s (m/s)	α_p (dB/λ)	α_s (dB/λ)	ρ (kg/m^3)
Water	1490		0		1000
Fiberglass	3500	1560	0.3	0.7	1530
Epoxy resin	3000	1550	0.8	1.8	1845
Steel	5950	3240	0	0	7700

working under free field conditions, has been extended here to lossy materials and applied to give an indication of the damping of the supported elastic waves due to both radiation and material absorption. The method, for a fixed ka value, solves

$$D_{\nu^l}(ka) = 0, \tag{5}$$

where D_{ν^l} is the denominator of Eq. (2), after replacing the real modal order n^l with its complex version

$$\nu^l = n^l + i\beta^l, \tag{6}$$

according to the Sommerfeld–Watson method. The complex modal order ν is the unknown of the equation. Its imaginary part β^l represents the damping of the elastic wave when it propagates along a circular arcpath corresponding to a unit central angle. It includes radiation damping and material absorption. Because it is based on an extension of the geometric diffraction theory to elastic phenomena,[20] this approach is valid for mid to high-ka values; in this study it is applied for a minimum ka of 5.

Hefner and Marston[22] proposed an alternative approximated formula for the damping factor of the elastic waves supported by solid spheres made of lossy materials:

$$\beta^l \approx \beta^l_{\text{lossless}} + \gamma^l\left(\frac{ka}{c^l_{\text{ph}}/c^{\text{ext}}} - 1/2\right), \tag{7}$$

where $\beta^l_{\text{lossless}}$ is the radiation damping factor that derives from the solution of Eq. (5) with real wave numbers. The symbol γ^l represents the imaginary part of the complex wave number of the lth wave supported by a water-loaded elastic half-space having the same properties as the dissipative material of the sphere. This is an alternative to solving Eq. (5) with complex bulk speeds, and is particularly interesting when the quality factor of a wave becomes too low due to material dissipation to allow the exact method to converge. Here the exact method is applied, except where otherwise indicated.

The analysis is performed over a ka range ($ka \leqslant 150$) much broader than what could be measured in the tank, with the purpose of making the theoretical study more complete, by including a large set of elastic waves and giving an idea of their asymptotic behavior, and, hence, of their general nature. The wave analysis is first applied to the components of the solid-filled shell of interest, namely (i) a void spherical shell made of fiberglass and (ii) a solid sphere made of the same epoxy resin as the filler. This helps in understanding the basic physical phenomena involved. The values of the material parameters used for the simulations are listed in Table I. The sound speed c^{ext} of the loading water was measured in the tank during the experiments. The values chosen for the material parameters are those estimated from the acoustic inversion of the fiberglass (for the external casing) and of the epoxy resin (for the solid filler), as presented in Sec. IV. The speed values are similar, but the attenuations are much bigger in the resin than in the fiberglass. One can notice that the values of the shear speed are supersonic, but very close to the sound speed in water. In a previous parametric study,[16] the shear speed turned out to be the parameter that mostly influences the classification of a material in terms of its elastic behavior. In particular, it is customary[23] to classify a material as plasticlike if its shear speed is subsonic with respect to loading water and its density is close to the water density and as metal-like if its shear speed is supersonic and its density much higher than the water density. In fact, it would be preferable to classify as plasticlike a material having the Rayleigh speed subsonic (even if its shear speed is supersonic), since it has been shown[16] to support the Scholte–Stoneley wave as a plastic material does.[24] For material classification purposes the Rayleigh speed considered is the speed of the Rayleigh wave supported by a free half-space. For plastic materials, both compressional and shear waves are generally highly damped.

The materials used in this experimental work fall within the category of plasticlike materials, since their shear speed is slightly supersonic, but their Rayleigh speed is subsonic. Furthermore, both the compressional and shear waves are significantly damped, and the density is close to water density, compared to the density of metal-like materials.

Figure 3(a) shows the wave analysis of the void shell with thickness $d=0.04a$ in terms of phase speed dispersion curves (upper plot) and damping factor β (lower plot) of the waves supported by the structure. The sphere geometry is the nominal geometry of the fiberglass shell measured in the tank, i.e., outer radius equal to 62.5 mm, and wall thickness 2.5 mm. For comparison, Fig. 3(b) shows similar results for a void steel shell of the same size and thickness (see Table I for the steel properties). The differences between the dispersion curves for the two materials are evident. In plastic shells, the A_{0-} Lamb-type wave is not supported, and the so-called coincidence frequency[25,26] does not exist. The other Lamb-type waves exist in both cases, but with very different asymptotic behaviors. As they are shear in nature, in plastic the S_0 and all the A Lamb-type waves tend relatively rapidly to the shear speed, which is generally slightly supersonic or subsonic. In metal, they are known to tend to the Rayleigh speed value in the mid to high-ka range and, at very high ka, to tend asymptotically to the bulk shear speed. For this reason, for example, in plastic the speed of the A_{0+} wave varies rapidly at low ka, then exponentially decays toward the material's low shear speed and becomes almost nondispersive; in metal, beyond the coincidence frequency, it increases with frequency and is highly dispersive. In the fiberglass shell, the only wave that significantly contributes to scattering, and for which the damping factor can be computed, is the S_0 wave. The other waves predicted are so highly attenuated that they cannot be detected at all over the bandwidth. The damping curves are shown for $\beta < 3.5$, since for β values even just slightly larger than 1, the resonance modes cannot be detected in the target response. In steel, the S_0 wave, being

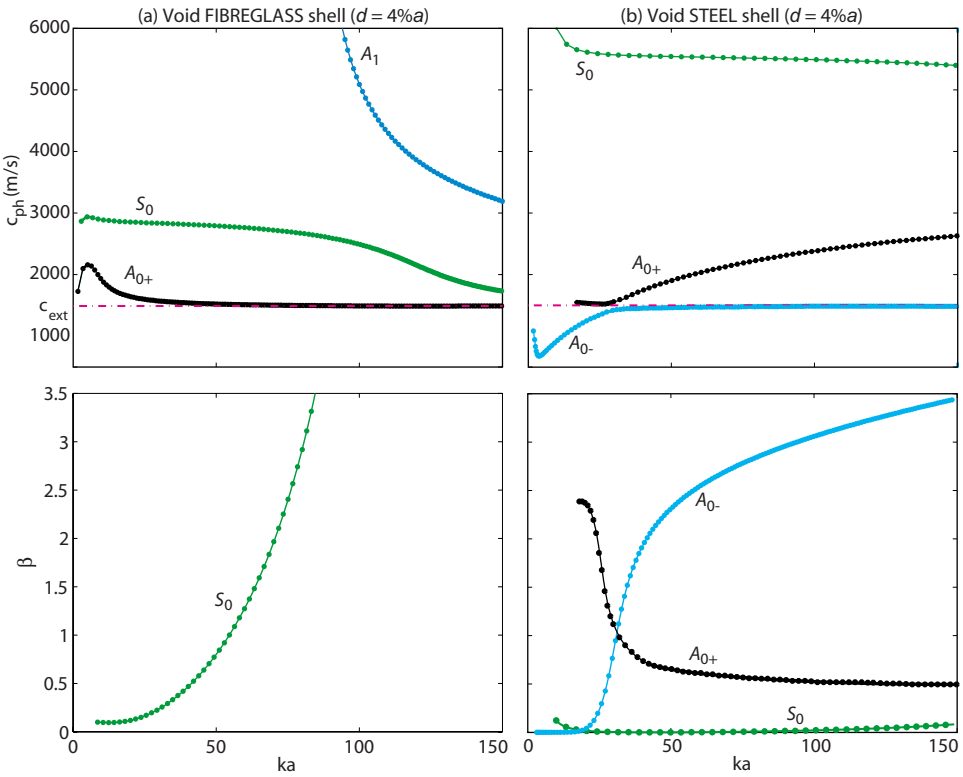


FIG. 3. (Color online) Wave analysis applied to a void, thin-walled shell made of fiberglass (a) and of steel (b). The upper plots show the phase speeds of the supported waves, as predicted by analytical modal analysis; the dash-dotted line indicates the sound speed in the external fluid. The lower plots show the damping factor β, as estimated by the numerical method originally introduced by Kargl *et al.* (Ref. 19). The numerical method is not able to estimate the damping factor of some of the waves due to their high attenuation. Their excitation is predicted by the theoretical scattering model, but they cannot be observed. The dot markers localize the resonance modes of each wave.

almost nondissipative in nature, has a very low damping factor, which remains almost flat over the whole bandwidth; in fiberglass, its damping rapidly increases due to intrinsic material dissipation. As is well known,[25] the A_{0+} and A_{0-} waves in steel exchange their nature around the coincidence frequency, which is located here around $ka=30$, depending on the shell-wall thickness and the material parameters.[27]

The second basic case is a water-loaded solid sphere, either made of epoxy resin as measured in the tank [Fig. 4(a)] or, for reference, made of steel [Fig. 4(b)]. The phase speed dispersion curves and the attenuation factor are predicted for the waves supported by solid spheres of outer radius 60 mm, which is the nominal outer radius of the measured solid sphere. The dashed lines connecting modes of the dispersion curves of some Whispering–Gallery waves indicate that the modes in between are missing due to their extremely high attenuation. The subsonic Scholte–Stoneley wave is excited only in the plasticlike sphere, and is expected to dominate the response at $ka<30$, beyond which its damping starts to increase. Its dispersion curve asymptotically tends to the Scholte–Stoneley speed computed for a water-loaded half-space and shows that the wave is nondispersive for $ka>10$. The Whispering–Gallery waves of longitudinal type asymptotically tend to the material compressional speed; the Whispering–Gallery waves of transverse type and the Rayleigh wave tend to the shear speed. Although the Rayleigh speed in a free half-space made with the same resin is predicted to be subsonic ($c_R \approx 1400$ m/s), the Rayleigh wave in the water-loaded resin sphere is slightly supersonic. In the steel sphere, the compressional speed is so high that the longitudinal waves are not included in the dispersion plot of Fig. 4(b). Their modes interfere so strongly with the transversal Whispering–Gallery wave modes that they are very difficult to identify and analyze. For this reason, the damping factor of the first-order longitudinal Whispering–Gallery wave is predicted by the approximated method proposed by Hefner and Marston.[22] The Whispering–Gallery waves and the Scholte–Stoneley wave are almost nondissipative in nature. Their strong damping in the resin sphere is mostly caused by material attenuation: At mid to high-ka values, the slope of the damping curves changes and tends to the damping they would respectively have if they traveled in a water-loaded half-space of the same resin as the sphere.[22] Studies on the Rayleigh wave at solid–fluid interfaces[23] showed that it is leaky in metals, as confirmed by the trend of its damping factor for steel, but it is unleaky in plastics. This cannot be confirmed by the analysis in this paper, since, for the epoxy-resin sphere, its damping factor cannot be evaluated.

The effect of a fiberglass, thin-walled shell coating a resin solid sphere is weak,[16] if the contact at the interface is perfect (i.e., continuity is satisfied for all the components of stress and displacement). This is due to the thin walls and the very similar properties of the two materials. In this case the shell is practically transparent, especially at low to mid-ka. For this reason a complete wave analysis is not reported for the resin-filled fiberglass shell. The dominating waves scattered by the solid-filled shell are the same as predicted by a solid sphere of the same size made of the filler material.[16] The filler-borne Whispering–Gallery waves of longitudinal and transverse types are unaffected by the shell over the entire bandwidth.

In a coated sphere, the Scholte–Stoneley and the Rayleigh surface waves travel at the interface between two solids, and not, as for the solid sphere, at the interface between the solid and water. However, with the combination of the two materials investigated and a shell sufficiently thin, this

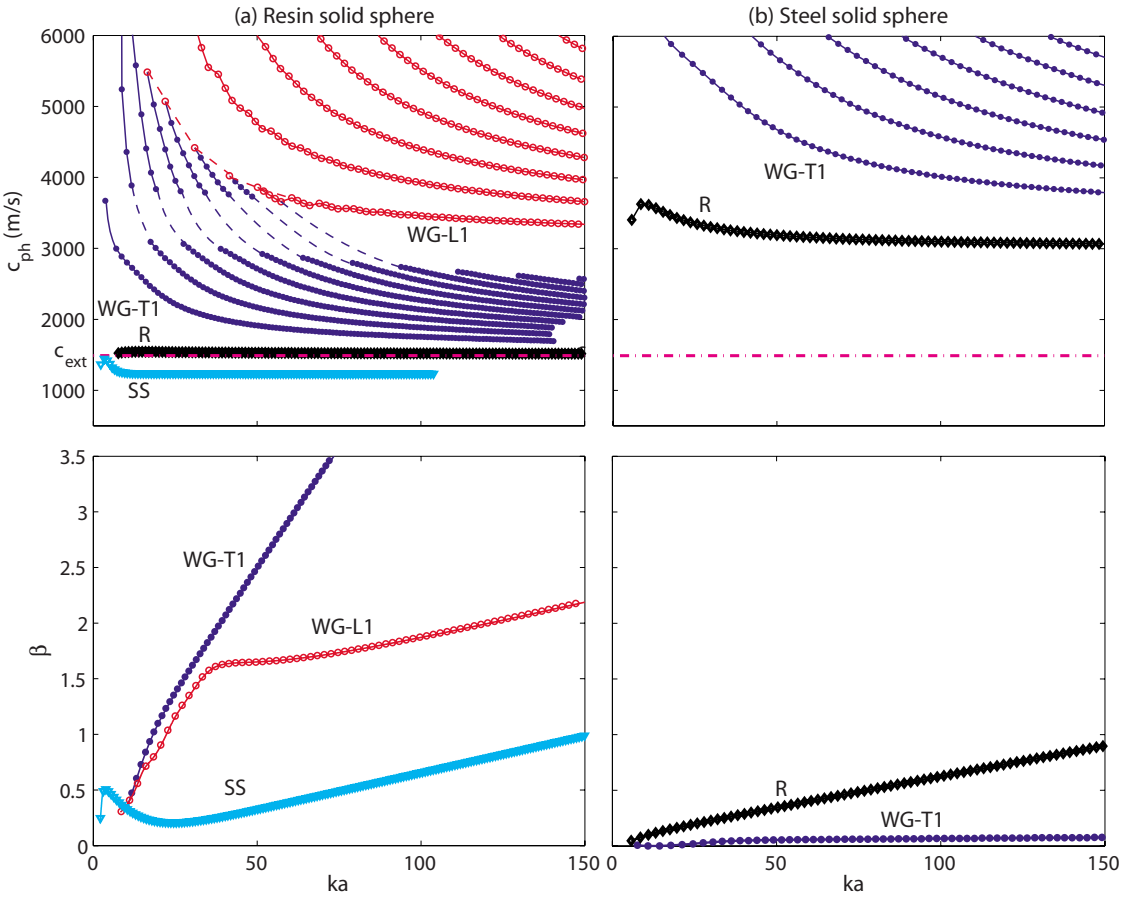


FIG. 4. (Color online) Wave analysis applied to a solid sphere made of an epoxy resin (a) and with steel (b). The upper plots show the phase speed dispersion curves of the supported waves, as predicted by analytical modal analysis. The dash-dotted line indicates the sound speed in the loading fluid. The lower plot shows the waves' damping factor. All the curves with closed circles in the upper plots refer to Whispering–Gallery waves of transverse type (labeled as WG-T), all the curves with open circles refer to the Whispering–Gallery waves of longitudinal type (WG-L). Only the first-order waves of the two types are labeled as they are the only ones contributing to backscattering. For the same reason, the damping factor is computed only for these two waves. The label R stands for Rayleigh wave and the label SS for Scholte–Stoneley.

has shown not to affect their properties. In particular, Fig. 5 shows the phase and group speed dispersion curves predicted for the Scholte–Stoneley and Rayleigh waves in the case of the fiberglass-coated and uncoated solid resin sphere studied in Fig. 4. It should be noticed that in this case the Scholte–Stoneley wave is expected to provide a strong contribution to the object response, while the Rayleigh wave is expected not to be seen. For other material combinations, theoretical and experimental studies[28] demonstrated that both waves can simultaneously propagate, and that they can simultaneously be observed. Under conditions of perfect bonding, the damping factor of the Scholte–Stoneley wave is negligibly affected by coating, and, hence, it is not reported here. If the contact is not perfect, the Scholte–Stoneley wave was found to undergo strong attenuation and dispersion, depending on the type of contact. The bonding types most commonly studied in the literature are the pure transverse slip,[29] and those boundary conditions characterized by discontinuity of either tangential[30] or radial, or both displacements.[31]

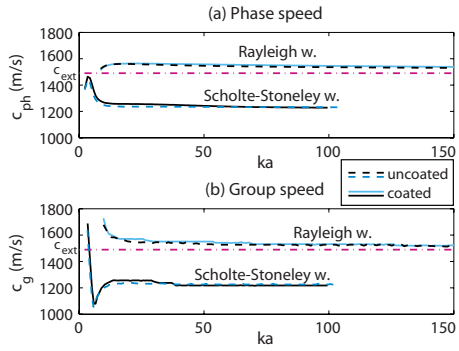


FIG. 5. (Color online) Phase (a) and group (b) speed dispersion curves of the Scholte–Stoneley and Rayleigh waves predicted for a solid resin sphere either uncoated or coated by a thin fiberglass shell (d=4%a). The dash-dotted line indicates the sound speed in the loading fluid. Perfect contact is assumed between the shell and the filler in the coated sphere. Due to the similar properties between the two materials and the thin walls of the casing the two waves are predicted to be negligibly affected by the presence of the shell.

Acoustic inversion procedure. The block diagram in Fig. 6 describes the inversion procedure applied to estimate the target material parameters either for a void shell or for a solid sphere. As suggested by previous investigations[16] and by the theoretical considerations presented earlier, this approach could be applied also to a completely filled, thin-walled shell in order to estimate the properties of the filler material. The approach can be easily extended to specimen of cylindrical shape having length significantly bigger than the radius.

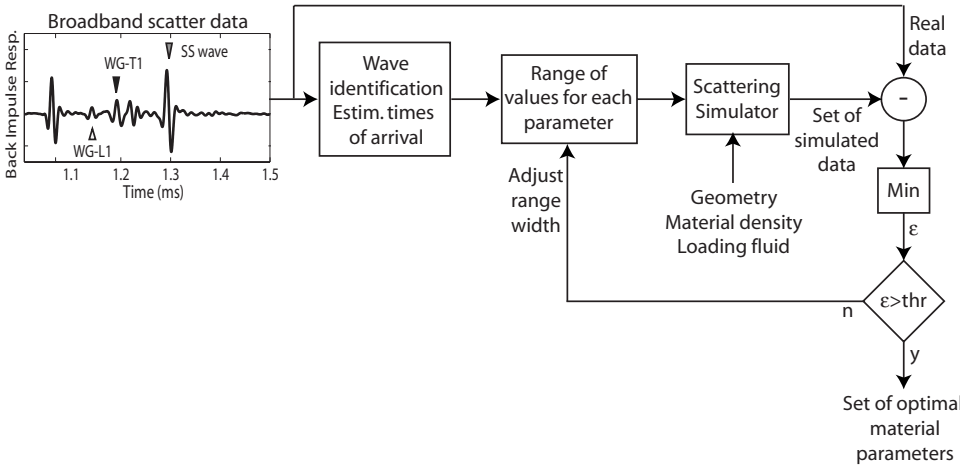


FIG. 6. Block diagram of the procedure applied to backscatter target response to invert a set of unknown material parameters.

The core of the approach is an analytical modeling tool for scattering by a free-field, fluid-loaded sphere based on Eq. (1). Hence, the materials are assumed to be homogeneous and isotropic, and the hypothesis of plane wave insonification is applied. Measurement geometry (i.e., sphere radius and thickness, and receiver position), loading fluid, and material density should be *a priori* known. Precision in the knowledge of the sphere's thickness and radius curvature should be on the order of about five-hundredths of a wavelength. This tolerance is generally achieved in standard manufacturing processes which are based upon the addition of successive layers of fiber, starting from a mother shape. Such processes are customarily implemented by specialized companies, such as those operating in the construction of high-end sail and motor yachts. The simulator must also be fed with a value for each of the unknown parameters. A range of possible values for each parameter is forecast based on the analysis of the measured scattered data, from which the time of arrival of each wave echo can be estimated. The delay Δt_m^l of the mth echo of the wave l with respect to the time of arrival of the front echo is related to the group and phase speed of the wave (c_g^l and c_{ph}^l respectively) as follows:[19]

$$\Delta t_m^l = 2a\left[\frac{1-\cos(\theta_c^l)}{c^{ext}} + \frac{\pi - \theta_c^l + m\pi}{c_g^l}\right], \tag{8}$$

where $m=0,1,2,\ldots$ is the number of complete circumnavigations, and $\theta_c^l = \arcsin(c^{ext}/c_{ph}^l)$ is the wave coupling angle. The values of c_g^l and c_{ph}^l (hence of θ_c^l) generally vary with respect to frequency; the frequency value used in Eq. (8) is the middle frequency of the band. In turn, the phase and the group speed of a wave varies with the material compressional and/or shear speeds, depending on the nature of the wave analyzed (as discussed for example by Viktorov[32]). The forecast is more precise, and, hence, the range of possible values more limited, when nondispersive waves are supported, as their phase and group speeds do not depend on frequency. Knowing the kind of material (namely either metal/stone or plasticlike) helps to predict what kinds of waves will be excited, and hence helps in the wave identification process, which is an essential preliminary phase of the inversion method. When several elastic waves are excited (such as in a solid sphere or in a very broadband measurement), the parameter estimation becomes more precise, due to redundancy of information. The wave attenuations are very difficult to predict from data analysis, and, hence, their possible ranges of values are kept wide.

Given the range of the search domain for each unknown, a simulated response for each combination of parameters is computed. The functional given by the difference between measured and simulated impulse responses in a certain bandwidth is minimized to obtain a set of optimal parameters. Until the minimum achieved exceeds a fixed threshold, the simulation/minimization process is iterated by feeding the simulator with a wider range of values for each unknown parameter. In order to keep the problem size relatively small, the customary procedure is to invert the compressional and shear speeds of the material, keeping their respective attenuations fixed to reasonable numbers. Once the two speeds are estimated, a new inversion procedure is applied to estimate their attenuations. The two problems can be separated because the attenuation values affect the amplitudes of the elastic wave echoes, but only negligibly their times of arrival, which, instead, are directly related to the compressional and shear speeds of the material.

IV. ACOUSTIC TANK MEASUREMENTS

Preliminary tests of homogeneity of the materials at the frequencies of interest were conducted by measuring the backscattering at different aspects of insonification, evaluated with respect to arbitrary zero references, as sketched in Fig. 7. First, backscattered responses by a void fiberglass spherical shell (outer radius $a=62.5$ mm, thickness $d=2.5$ mm) and by a solid sphere of epoxy resin (radius $a=60$ mm) were measured with the target suspended in midwater. In the investigated bandwidth, the shell appeared homogeneous (Fig. 8 compares measurements at aspects 45° and 180°). At the two ends of the investigated bandwidth (i.e., for $ka<7$ and $ka>20$), discrepancies are possibly due to the lower SNR level at the extremes of the source bandwidth. In the considered bandwidth, only the S_0 Lamb-type wave can be detected, as anticipated in Sec. III. The exponential decay of the amplitudes of its echoes is evident in the time responses.

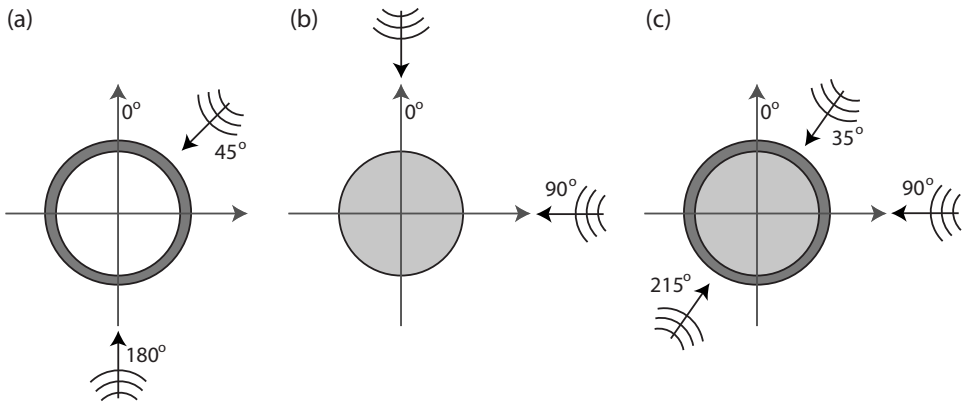


FIG. 7. Geometry of the measured spheres and respective insonification angles: (a) Void fiberglass shell; (b) solid epoxy-resin sphere; and (c) resin-filled fiberglass shell. Geometry is not to scale.

Figure 9 shows the data comparison of the solid resin sphere insonified at two different aspects. Although the resin was cast in a set of steps during the sphere manufacturing, as described in Sec. II, from the data agreement between different aspects of insonification one can deduce that the sphere is homogeneous in the bandwidth. The orientation of the interfaces between adjacent layers of the filling resin with respect to the direction of insonification is unknown. Wave identification can be applied only in the time domain, since the target strength is dominated by the strong interference between the specular and the Scholte–Stoneley echoes, which hides all the resonance modes. The impulse response is dominated by the first echo of the Scholte–Stoneley wave, which is expected to be the strongest elastic wave supported at low- to mid-*ka*. The first echoes of the first-order Whispering–Gallery waves of longitudinal and transverse types, respectively, can be identified. The latter one is very dispersive in this bandwidth, as predicted by its dispersion curve in Fig. 4.

The resin-filled shell was measured under the same geometry at different aspects of insonification. The comparison of time responses between aspect 35° and 90° is shown in Fig. 10(a), and between aspects 90° and 215° in Fig. 10(b). The main elastic wave echoes are identified. The response at 35° shows phase reversal of the front echo. Small differences in amplitude and phase are related to the various Whispering–Gallery wave echoes. As in the solid sphere data, the transversal Whispering–Gallery echo is very dispersive. The most significant changes among the different measurements appear in the amplitude, phase, and dispersion of the first Scholte–Stoneley wave echo. In particular, the Scholte–Stoneley wave echo is bigger and less dispersed at 35° of insonification, i.e., when the front echo is reversed. Data at aspect 215° is perfectly in phase with the ones at aspect 90°, but the Scholte–Stoneley echo has almost disappeared and new significant echoes can be seen around 1.35 ms, possibly reflected/diffracted by local, sparse inhomogeneities. The phase reversal of the front echo in the data at aspect 35° suggests the presence of a considerably extended air bubble or air layer immediately behind the part of the shell hit by the incident pulse. In the data at aspect 90° and 215°, the filler appears in contact with the shell in the illuminated part of the sphere (front), whereas the lower level of the Scholte–Stoneley wave echo may indicate, as expected by theory,[29–31] that under these geometries the air pocket is in the rear part, where this wave travels before its first back reradiation. Hence, the sphere is evidently neither three-dimensional (3D) symmetric nor homogeneous, due to the presence of one or more extended air pockets, which are

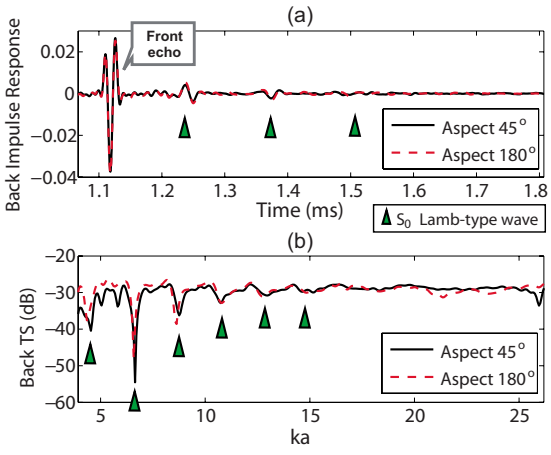


FIG. 8. (Color online) Comparison of backscattered response by the void fiberglass shell at two different aspects: (a) Time impulse response and (b) target strength. Elastic wave identification is superimposed.

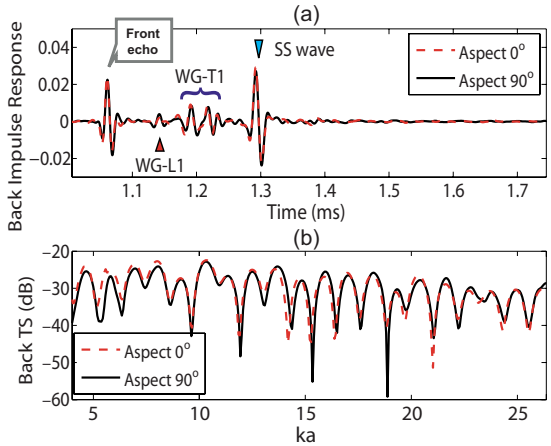


FIG. 9. (Color online) Comparison of backscattered response by the solid resin sphere at two different aspects: (a) Impulse response and (b) target strength. Elastic wave analysis is applied to the impulse response: WG-L/T1 represent longitudinal/transverse Whispering–Gallery waves of the first order; SS represents Scholte–Stoneley wave. The strong interference between the specular echo and the first Scholte–Stoneley wave echo does not allow the identification of wave resonance modes in the frequency domain.

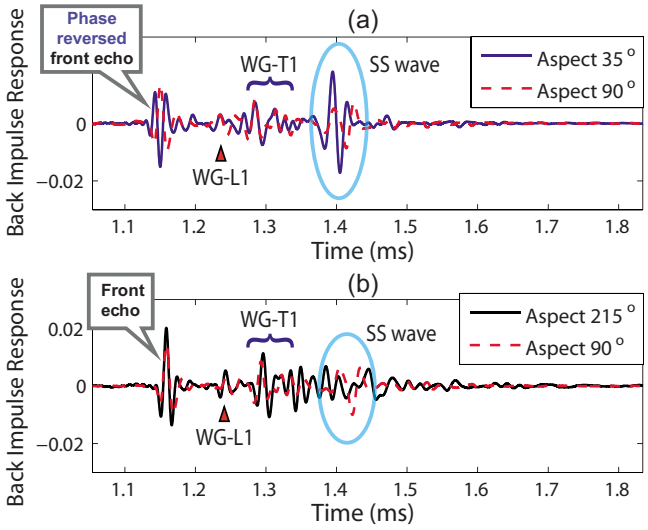


FIG. 10. (Color online) Backscattered impulse response by the resin-filled fiberglass spherical shell at different insonification aspects: Comparison between angles 35° and 90° (a), and between angles 90° and 215° (b). Wave analysis is superimposed.

expected to be located at the interface between solid and shell, as they strongly affect only the Scholte–Stoneley surface wave.

Acoustic inversion of material properties based on elastic scattering features. Acoustic inversion is applied to the void shell and the solid sphere in order to estimate the elastoacoustic parameters of their materials in the studied frequencies. The inversion methodology used is described in Sec. III. The material densities are estimated by measuring the weight and by estimating the volume of the objects. Figure 11 shows the result of model–data comparison for the void shell in the time and frequency domains after acoustic inversion of the material parameters. The inversion results are indicated in Table I. The estimation error on the speeds is

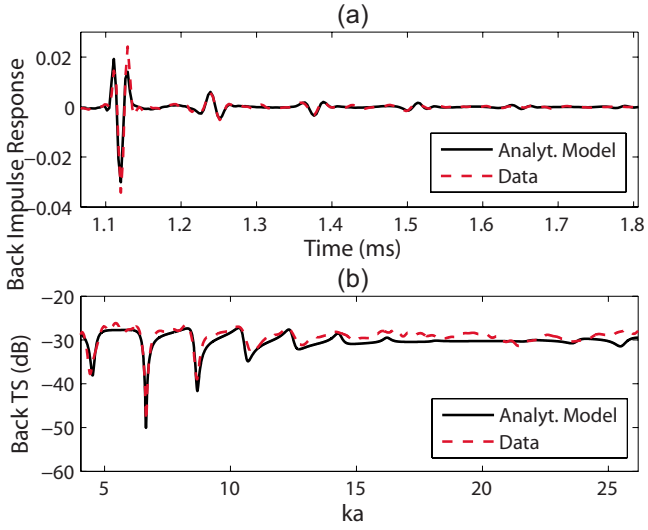


FIG. 11. (Color online) Model–data comparison of backscattering by the void shell. The analytical model is applied. Data were measured at aspect 180° (see Fig. 8). The plots show the model–data comparison of (a) the impulse response and (b) the target strength. The model is fed with the material parameters obtained from the inversion process and listed in Table I.

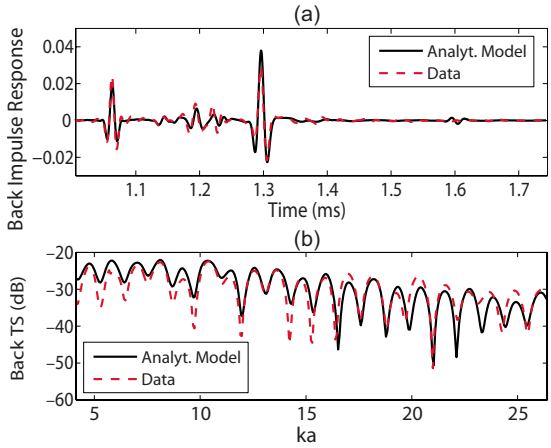


FIG. 12. (Color online) Model–data comparison of backscattering by the resin solid sphere (analytical model is applied): (a) Impulse response and (b) target strength. The model is fed with the material parameters obtained from inversion and reported in Table I.

of the order of ± 30 m/s, on the attenuations around ± 0.2 dB/λ. The disagreement beyond $ka=15$ is possibly due to increasing relevance of the fiber structure details as the frequency increases, which may perturb the propagation of the S_0 wave, being shell-borne and surface guided in nature. The relatively high uncertainty of the estimate derives from the excitation of only one elastic wave in the measured bandwidth and from the fact that, in this band, its phase speed derives from a combination of the bulk compressional and shear speeds of the material.[26]

The model–data comparison achieved after acoustic inversion of the solid sphere data (Fig. 12) shows a generally good agreement. The inversion results are indicated in Table I. The estimation error of the two speeds is of the order of ± 20 m/s, lower than in the shell case since here more elastic waves are excited, providing redundant information, hence more robust estimation. The very good agreement achieved between model and data suggests that the model approximation of assuming plane wave insonification is acceptable also for $ka<13$. It also seems to suggest that the material is nondispersive in the whole bandwidth of interest. The main discrepancy in the time-domain model–data comparison is in the level of the Scholte–Stoneley surface wave echo arriving at $t=1.3$ ms, and corresponding to a mismatch in the target strength level, mainly at low ka. This is probably due to partial diffraction (and, hence, leakage) of the wave at the small protrusions of the suspension system.

The estimated values of the two bulk speeds are in very close agreement (within a few tens of m/s) with the results obtained from ultrasonic spot measurements. These measurements were conducted at 2 MHz (for compressional speed measurement) and at 5 MHz (for shear speed measurement) on a 10-mm-thick, flat disc specimen made of the same material. Nevertheless, this is not sufficient to conclude that the material is nondispersive over a bandwidth ranging from 100 kHz up to frequencies of the order of 1 MHz, since the frequency dependence of the phase velocity (and attenuation) of ultrasonic waves in an epoxy resin was experimentally shown to depend on the curing process.[33]

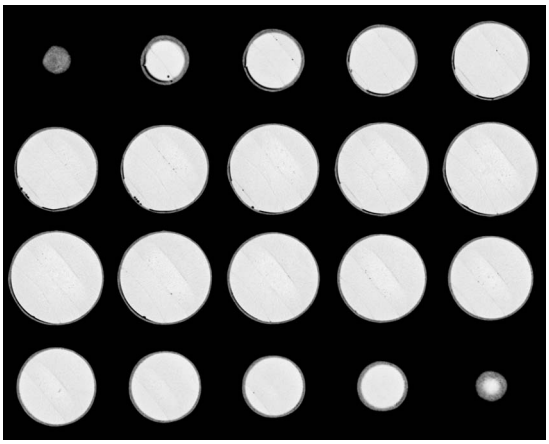


FIG. 13. Sequence of a subset of CT-scan vertical slices of the resin-filled fiberglass shell (horizontal resolution subsampled to 6.25 mm).

V. DETAILED CHARACTERIZATION OF THE SOLID-FILLED FIBERGLASS SPHERICAL SHELL

The detection of a considerably sized air layer at the solid–shell interface of the totally filled sphere induced the authors to conduct independent measurements aimed to localize and characterize it.

A. Additional NDT measurements and estimation of the air-pocket geometry

Confirmation of the presence of an extended thin air pocket between shell and filler comes from additional independent measurements. The x-ray CT scan of the sphere, performed with a GE Medical Systems multislice scanner at the Radiology Branch of the Carrara Hospital, has a horizontal resolution of 0.625 mm and reveals the presence of an extended interlayer of air below the shell, along with sparse air bubbles of different size, and various small patches of heterogeneous material (possibly corresponding to small bubble clouds or blobs of different density) within the filler. An example of data acquired is given by the sequence of a subset of CT scan images in Fig. 13. From the analysis of the slices, it is also possible to check the sphericity of the object and the uniformity of the shell wall thickness. In the first and last slices of the sequence, the small gray-scale texture of the images reveals the composite nature of the shell material; however, it also appears relatively homogeneous, since no joint between adjacent fiber patches can be distinguished. A detailed analysis of one of the most significant slices is provided in Fig. 14. The image clearly shows the section of two air pockets which are almost symmetric with respect to an ideal symmetry axis superimposed on the image. Their thickness is maximum (about 1.5 mm) close to pole P and then gradually decreases to zero closer to the equator. Visual inspection allows also the detection of the interfaces between the different layers of the casting process, all approximately perpendicular to the symmetry axis. The application of some basic image processing procedures allows the 3D reconstruction of the air pocket, as shown in Fig. 15, where the shell is rendered as semitransparent and the air bubbles and the air pocket are opaque. The roughly symmetric annular shape of the air interlayer is evident. From a slice-by-slice analysis

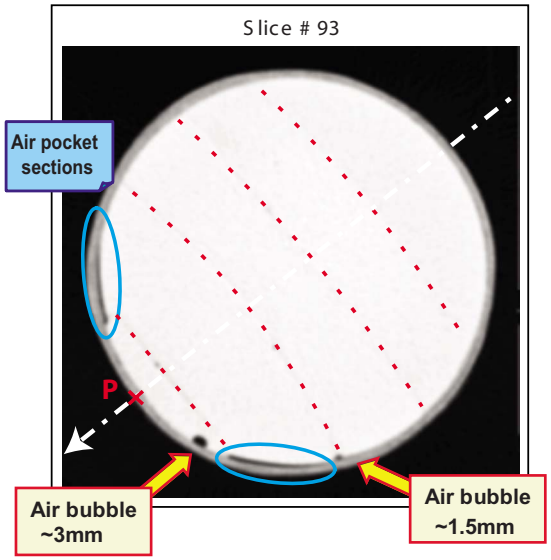


FIG. 14. (Color online) Analysis of one significant CT-scan vertical slice. The image background is black, the shell midgray, the filler light gray. The air bubbles/pockets appear black or dark gray in the x-ray image. The superimposed dotted lines localize the interfaces between adjacent cast layers of the filling resin. The dashed-dotted arrow shows a virtual axis of symmetry. The pole P, marked by a cross on the axis of symmetry, roughly corresponds to the filling hole. The main air-pocket cross sections, roughly symmetric with respect to the drawn axis, are surrounded by ellipses.

and from the 3D reconstruction it is possible to deduce that the air pocket was caused by the shrinkage of the inner material during the solidification of the fourth casting step. During the following, and last, casting step the resin could not fill the thin gap as some air remained trapped. The 3D reconstruction also shows that several big air bubbles remained trapped during the last casting step very close to the small hole through which the shell was filled.

In order to better measure the extent of the ring along the sphere surface, ultrasound scanning was conducted at 5 MHz with a Krautkramer transducer on the surface of the

FIG. 15. (Color online) Three-dimensional reconstruction of the resin-filled shell from x-ray CT scanning. The sphere shell is confined between two semitransparent surfaces, while the trapped air is enclosed within opaque surfaces.

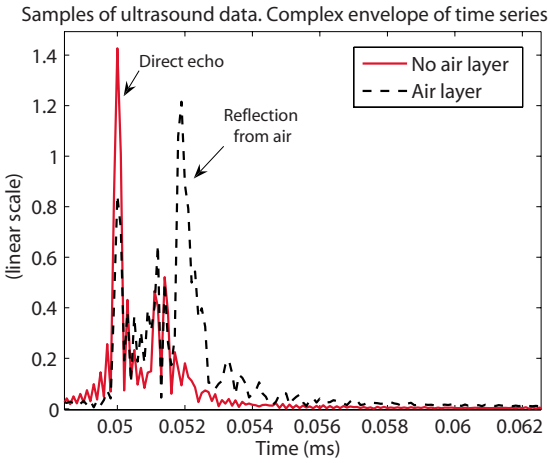


FIG. 16. (Color online) Ultrasound measurement of the resin-filled shell. Example of signals in presence and absence of an internal air bubble.

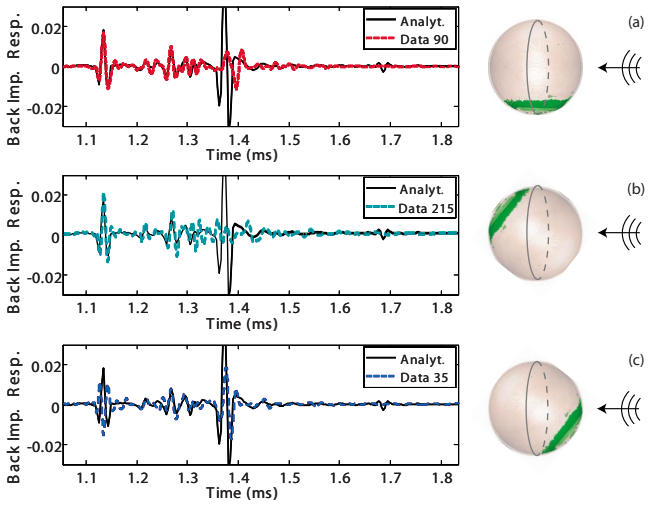


FIG. 18. (Color online) Model–data comparison of backscattering by the resin-filled fiberglass shell (impulse response). The analytical model is applied. On the left-hand side, the impulse response is shown as the insonification aspect varies; on the right-hand side, the cartoons show the position of the annular air pocket at the different aspects of insonification: response at (a) 35°, (b) 90°, and (c) 215°.

hemisphere where the CT scan localized the air pocket. Figure 16 compares the time envelope of the signal in the presence and absence of the air interlayer below the shell wall. A strong echo is reflected back from the internal interface of the shell wall only if there is air at the other side; otherwise the impedance between the shell and filler materials is too low to give a significant reflection at the interface. Measurements were performed on a grid of points having an approximate resolution of 5 mm [see Fig. 17(a)]. The 3D map of points where the internal air was detected is shown in Fig. 17(b). Detection is decided by thresholding the measured signal in the time window between 0.0512 and 0.053 ms. The cross labeled with P on the map indicates the pole of the sphere and corresponds to the pole on the symmetry axis drawn on the CT-scan image of Fig. 14.

B. Model–data comparison and discussion

Based on the results of the NDT measurements, it is possible to discuss the acoustic measurements of the fiberglass-coated resin sphere and their comparison to the simulation obtained by feeding an ideal analytical model (assuming perfect bonded contact at the filler–shell interface) with the material parameters estimated in Sec. IV (see Fig. 18). The parameters still appear to be valid, as confirmed by the generally satisfactory agreement in amplitude and phase of both the longitudinal and transverse Whispering–Gallery wave echoes. The amplitude and dispersion of the Scholte–

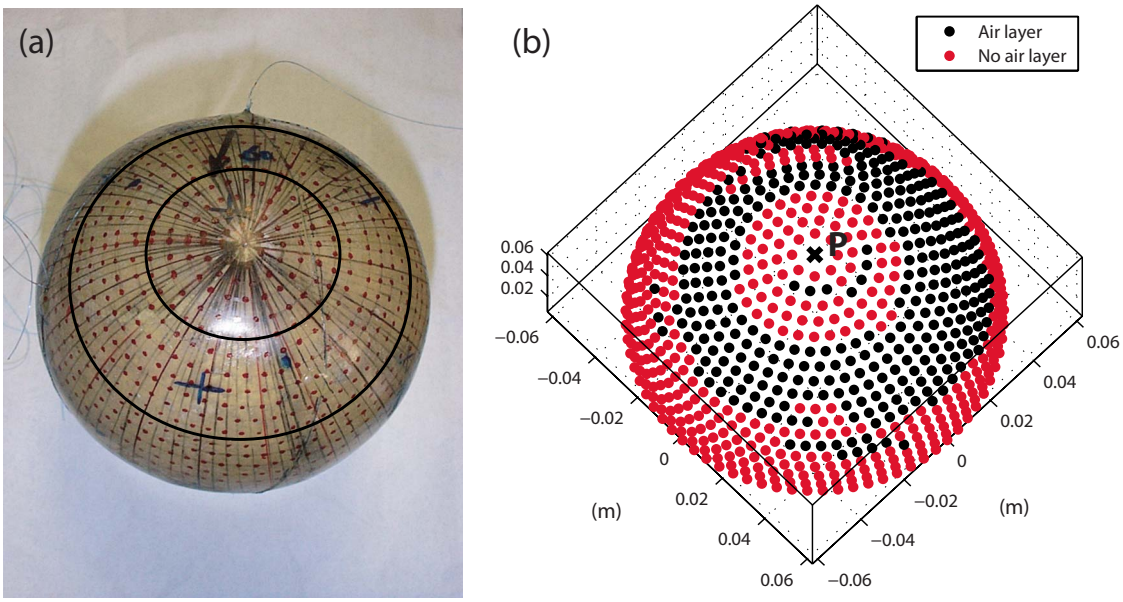


FIG. 17. (Color online) Ultrasound measurement of the resin-filled shell. (a) The hemisphere containing the internal flaw. The grid of dots indicates the points of ultrasound measurement. The two black circles delimit the area of the air pocket as detected by the ultrasonic measurements (b). At the top of (a) the thin wire used for suspension can be distinguished. (b) Three-dimensional mapping of the measured points on the sphere surface, where the detection of internal air is localized.

Stoneley wave echo strongly depends on the air-pocket location with respect to the travel path of the wave itself. As the wave is subsonic, its generating line is the circumference drawn on the sphere in the right column of Fig. 18. The first echo of the wave reradiates back from the sphere at the same circumferential line after the wave has traveled along the shell–filler interface around the nonilluminated hemisphere. Hence, the Scholte–Stoneley wave echo amplitude and shape deviate from the analytical prediction for the ideal geometry depending on the insonification angle, i.e., depending on the degree of interaction with the air pocket. When the air layer is completely included in the nonilluminated hemisphere and, hence, its interaction with the wave is total [Fig. 18(b)], the wave echo almost disappears; when half of the area of the air pocket interacts with the wave [Fig. 18(a)], the wave echo is significantly attenuated and dispersed. Finally, Fig. 18(c) shows a situation in which the interaction is very limited, and, hence, the phase of the Scholte–Stoneley wave is in good agreement with the ideal expectations, while its amplitude is lower than expected, but much higher than in the other measurements.

Due to the asymmetry of the sphere interior, attempts to apply analytical models with different hypotheses of boundary conditions at the filler–shell interface (such as pure transverse slip,[29] or discontinuity of one or both displacement components[31]) are unsuccessful in properly modeling the Scholte–Stoneley wave echo since they apply the same boundary condition at any point of the interface. Only a fully 3D model can provide high-fidelity solutions to this problem, but it would need an extremely precise knowledge of the size and distribution of all the main inhomogeneities (air pockets) and of the actual type of local contact at the interface, which is hard to measure or estimate, and, hence, is not a practical approach.

However, the roughly axisymmetric shape of this air pocket makes it possible to apply the modeling tool AXISCAT[17] with the purpose to refine the model–data comparison achievable by analytical tools. AXISCAT is a frequency-domain finite-element model for scattering by axially symmetric, fluid-loaded structures subject to a nonsymmetric forcing.

The geometry of the air pocket is modeled from the NDT measurements as described in the following (see Fig. 19). An equivalent, average width and thickness of the ring is estimated around the axis of symmetry of the sphere, and used to model an axisymmetric air pocket. The mesh is discretized with triangular cubic Lagrange elements of maximum edge length equal to 1.7 mm. Around the corners of the air pocket the mesh is refined with element clouds having edge lengths on the order of 0.1 mm. The result of the model–data comparison, obtained by feeding the model with the material parameters inverted in Sec. IV, is shown in Fig. 20, where the data at 90° of aspect are selected. In the time domain, the agreement for the Scholte–Stoneley wave echo has much improved with respect to the analytical model result in terms of both amplitude and initial phase, while the dispersion of the echo cannot be perfectly modeled. The remaining dispersion mismatch is particularly evident in the frequency domain, where the pattern of peaks/dips characterizing the target strength is very sensitive to the interaction between the specular and the Scholte–Stoneley echoes. This mismatch may be due either to a not sufficiently accurate modeling of the geometry of the air pocket, or to the total absence in the model of other significantly big air bubbles (e.g., those ones close to the sphere pole as shown in Fig. 15), but also from the presence of boundary conditions at the filler–shell interface locally or generally different from the perfect-contact condition assumed in the model. As mentioned earlier, precise knowledge of local contact conditions is hard to obtain, as they can vary point to point and from object to object, depending on a series of factors, including the external temperature during manufacturing, which could not be precisely controlled. The very good agreement in the

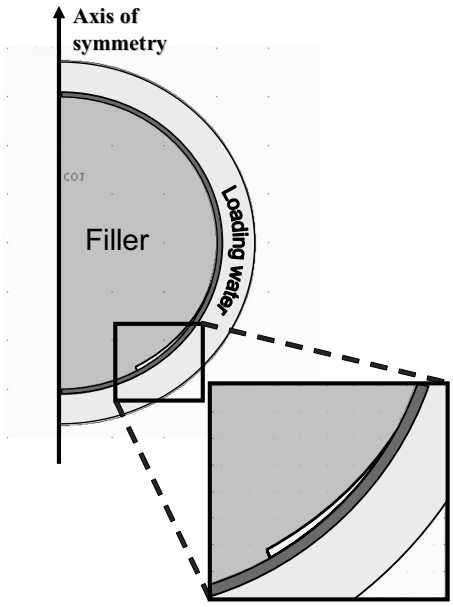


FIG. 19. Model of the resin-filled shell in AXISCAT: An axially symmetric three-dimensional object is obtained by rotation of the two-dimensional surface around the symmetry axis. The geometry of the air-pocket cross section is shown in the blow-up.

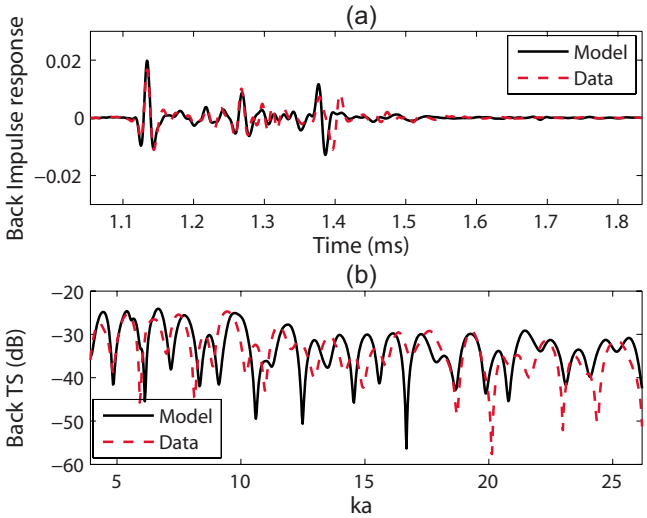


FIG. 20. (Color online) Model–data comparison of backscattering by the resin-filled fiberglass shell: (a) Time impulse response and (b) target strength. The simulation is obtained from AXISCAT. The data are acquired at the insonification angle of 90°.

times of arrival of the Whispering–Gallery waves between model and data suggests not only that the material parameters used are correct, but also that the filler material is acoustically nondispersive in the bandwidth, and, hence, that the heterogeneities noticed in the CT scan are acoustically transparent at these frequencies. As predicted by theory, the fiberglass shell appears practically transparent.

The results of comparison between the model and the data at the other insonification aspects are not reported here. At the nominal aspect angles of 35° and 215°, the agreement is similar, but worse than the result obtained at 90°. The uncertainty in the orientation of the air pocket with respect to the direction of insonification is higher. Furthermore, in these geometries, the role played by the asymmetric parts of the annular air pocket, which cannot be modeled by AXISCAT, appears to be more relevant.

VI. CONCLUSIONS

A series of monostatic acoustic scattering measurements were conducted in a tank on an epoxy-resin-filled fiberglass spherical shell and its components (filler sphere and void shell) in the frequency bandwidth 15–100 kHz, corresponding roughly to a ka range of 5–26. The data were analyzed in terms of elastic waves supported by the structure, and compared to suitable models. Acoustic inversion was applied to the void fiberglass shell and to the solid resin sphere to estimate the respective material properties. The main advantage of the proposed inversion approach consists in the possibility of inverting all the parameters of interest by means of a single, broadband measurement, due to the curved (in particular, circular) cross section of the target, and to the broadband nature of the measurement itself, which allows the simultaneous excitation of all the elastic waves supported within the bandwidth of interest. Furthermore, no *a priori* information is needed except for a precise (on the order of about five-hundredths of a wavelength) knowledge of the geometry (sphere radius and thickness).

Model–data comparisons show that the investigated fiberglass is nondispersive below $ka=17$, beyond which the decay of the S_0 Lamb-type resonance modes is stronger than expected, which may imply a further leakage due to diffraction by the fibers along the wave travel path. The absence of strong elastic waves prevented possible evaluations of the dispersiveness of the material beyond $ka=17$. The epoxy resin appears to be nondispersive over the whole bandwidth investigated.

The main NDT result obtained was that an extended air pocket could be detected in the resin-filled shell by comparing underwater acoustic measurements at various aspects. The flaw was independently measured by x-ray CT scan and ultrasound measurements of the object. This study also exposes some of the potential pitfalls associated with the manufacturing of simple objects made of composite/plastic materials. Furthermore, the results show how much the low-frequency response of an object can change due to material heterogeneities deriving from manufacturing flaws. Hence, this defect was useful to prove the potential of acoustic elastic scattering analysis for SHM applications in the case of highly lossy, multilayered materials.

ACKNOWLEDGMENTS

The authors are grateful to the NURC engineering staff involved in the tank experiment, in particular to A. Figoli, A. Sapienza, and L. Troiano, for their precious help, and to F. Jensen for the fruitful scientific discussions. Special thanks go to S. Jensen, M.D., and G. Tognini, M.D., both from Carrara Hospital, who kindly provided the CT scan data. G. Canepa helped in the 3D reconstruction of the sphere from the CT-scan images. J. Fawcett, from DRDC-Atlantic, provided the analytical modeling tool. Many thanks go to S. Kargl for providing the root finder for void shells of lossless materials. The spheres were manufactured by TechnoWave.

[1]Y. Bar-Cohen, A. K. Mal, and S.-S. Lih, "NDE of composite materials using ultrasonic oblique insonification," Mater. Eval. **51**, 1285–1296 (1993).
[2]D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).
[3]M. J. S. Lowe and O. Diligent, "Low-frequency reflection characteristics of the s_0 Lamb wave from a rectangular notch in a plate," J. Acoust. Soc. Am. **111**, 64–74 (2002).
[4]M. J. S. Lowe, P. Cawley, J.-K. Kao, and O. Diligent, "The low-frequency reflection characteristics of the fundamental antisymmetric Lamb wave a_0 from a rectangular notch in a plate," J. Acoust. Soc. Am. **112**, 2612–2622 (2002).
[5]J. Rajagopalan, K. Balasubramaniam, and C. V. Krishnamurthy, "A phase reconstruction algorithm for Lamb wave based structural health monitoring of anisotropic multilayered composite plates," J. Acoust. Soc. Am. **119**, 872–878 (2006).
[6]Z. Su, L. Ye, and Y. Lu, "Guided Lamb waves for identification of damage in composite structures: A review," J. Sound Vib. **295**, 753–780 (2006).
[7]W. Sachse and Y.-H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**, 4320–4327 (1978).
[8]T. W. Taylor and A. H. Nayfeh, "Damping characteristics of thick rectangular laminates," J. Acoust. Soc. Am. **100**, 1561–1570 (1996).
[9]H. J. McSkimin and J. P. Andreatch, "A water immersion technique for measuring attenuation and phase velocity of longitudinal waves in plastics," J. Acoust. Soc. Am. **49**, 713–722 (1971).
[10]M. R. Karim, A. K. Mal, and Y. Bar-Cohen, "Inversion of leaky Lamb wave data by simplex algorithm," J. Acoust. Soc. Am. **88**, 482–491 (1990).
[11]D. Fei, D. E. Chimenti, and S. V. Teles, "Material property estimation in thin plates using focused synthetic-aperture acoustic beams," J. Acoust. Soc. Am. **113**, 2599–2610 (2003).
[12]V. K. Kinra, P. T. Jaminet, C. Zhu, and V. R. Iyer, "Simultaneous measurement of the acoustical properties of a thin-layered medium: The inversion problem," J. Acoust. Soc. Am. **95**, 3059–3074 (1994).
[13]G. C. Gaunaurd, D. Brill, H. Haung, P. W. B. Moore, and H. C. Strifors, "Signal processing of the echo signatures returned by submerged shells insonified by dolphin 'clicks': Active classification," J. Acoust. Soc. Am. **103**, 1547–1557 (1998).
[14]A. Tesei, W. L. J. Fox, A. Maguer, and A. Løvik, "Target parameter estimation using resonance scattering analysis applied to air-filled, cylindrical shells in water," J. Acoust. Soc. Am. **108**, 2891–2900 (2000).
[15]S. D. Holland, S. V. Teles, and D. E. Chimenti, "Air-coupled, focused ultrasonic dispersion spectrum reconstruction in plates," J. Acoust. Soc. Am. **115**, 2866–2872 (2004).
[16]A. Tesei, M. Zampolli, and J. A. Fawcett, "Acoustic scattering from solid-filled spherical shells: Parametric study of elastic effects," in *Proceedings of the Eighth European Conference on Underwater Acoustics*, Carvoeiro, Portugal, 12–15 June 2006, pp. 157–162.
[17]M. Zampolli, A. Tesei, F. B. Jensen, N. Malm, and J. B. Blottman, "A computationally efficient finite element model with perfectly matched layers applied to scattering from axially symmetric objects," J. Acoust. Soc.

Am. **122**, 1472–1485 (2007).
[18]N. D. Veksler, *Resonance Acoustic Spectroscopy* (Springer, Berlin, 1993).
[19]S. G. Kargl and P. L. Marston, "Observations and modeling of the backscattering of short tone bursts from a spherical shell: Lamb wave echoes, glory, and axial reverberations," J. Acoust. Soc. Am. **85**, 1014–1028 (1989).
[20]P. L. Marston, "GTD for backscattering from elastic spheres and cylinders in water and the coupling of surface elastic waves with the acoustic field," J. Acoust. Soc. Am. **83**, 25–37 (1988).
[21]K. L. Williams and P. L. Marston, "Backscattering from an elastic sphere: Sommerfeld-Watson transformation and experimental confirmation," J. Acoust. Soc. Am. **78**, 1093–1102 (1985).
[22]B. T. Hefner and P. L. Marston, "Backscattering enhancements associated with subsonic Rayleigh waves on polymer spheres in water: Observation and modeling for acrylic spheres," J. Acoust. Soc. Am. **107**, 1930–1936 (2000).
[23]F. Padilla, M. de Billy, and G. Quentin, "Theoretical and experimental studies of surface waves on solid-fluid interfaces when the value of the fluid sound velocity is located between the shear and the longitudinal ones in the solid," J. Acoust. Soc. Am. **106**, 666–673 (1999).
[24]F. Chati, F. Leon, and G. Maze, "The generation, propagation, and detection of Lamb waves in plates using air-coupled ultrasonic transducers," Acoust. Lett. **24**, 1–4 (1996).
[25]G. S. Sammelmann, D. H. Trivett, and R. H. Hackman, "The acoustic scattering by a submerged, spherical. I. The bifurcation of the dispersion curve for the spherical antisymmetric Lamb wave," J. Acoust. Soc. Am. **85**, 114–124 (1989).
[26]A. Tesei, A. Maguer, W. L. J. Fox, R. Lim, and H. Schmidt, "Measurements and modeling of acoustic scattering from partially and completely buried spherical shells," J. Acoust. Soc. Am. **112**, 1817–1830 (2002).
[27]A. D. Pierce, *Acoustics. An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).
[28]C. Matteï, X. Jia, and G. Quentin, "Direct experimental investigations of acoustic modes guided by a solid-solid interface using optical interferometry," J. Acoust. Soc. Am. **102**, 1532–1539 (1997).
[29]G. J. Kühn and A. Lutsch, "Elastic wave mode conversion at a solid-solid boundary with transverse slip," J. Acoust. Soc. Am. **33**, 949–954 (1961).
[30]G. S. Murty, "A theoretical model for the attenuation and dispersion of Stoneley waves at the loosely bonded interface of elastic half spaces," Phys. Earth Planet. Inter. **11**, 65–79 (1975).
[31]S. I. Rokhlin and Y. Wang, "Analysis of boundary conditions for elastic wave interaction with an interface between two solids," J. Acoust. Soc. Am. **89**, 503–515 (1991).
[32]I. A. Viktorov, *Rayleigh and Lamb Waves. Physical Theory and Applications* (Plenum, New York, 1967).
[33]S. I. Rokhlin, D. K. Lewis, K. F. Graff, and L. Adler, "Real-time study of frequency dependence of attenuation and velocity of ultrasonic waves during the curing reaction of epoxy resin," J. Acoust. Soc. Am. **79**, 1786–1793 (1986).

Coupled hydrodynamic-acoustic modeling of sound generated by impacting cylindrical water jets

Xuemei Chen,[a] Steven L. Means, and William G. Szymczak
Acoustic Division, Naval Research Laboratory, Washington, DC 20375

Joel C. W. Rogers
Department of Mathematics, Polytechnic University, Brooklyn, New York 11201

(Received 17 April 2007; revised 4 February 2008; accepted 3 May 2008)

A coupled hydrodynamic-acoustic model describing acoustic propagation in a fluid containing multiple bubbles is proposed and applied to simulate noise generated by impacting water jets. The total pressure is decomposed into a "hydrodynamic" part and an "acoustic" part and computed using different schemes. The hydrodynamic pressure field is calculated independently using a generalized hydrodynamic model, and the pressure variations serve as sources in the wave equation for the acoustic pressure. A numerical algorithm developed to calculate wave propagation in an irregular region is used to account for the existence of the cavities. Noise generated by the impact of two cylindrical water jets is predicted. The computed near-field pressure is compared with the experimental data. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2936366]

PACS number(s): 43.30.Nb, 43.20.Bi, 43.30.Gv [WMC] Pages: 841–850

I. INTRODUCTION

It has been known for many years that the Knudsen noise spectrum is associated with gas bubble oscillations resulting from wave-breaking,[1] even at low sea states when there is little or no observable whitecapping.[2–5] Over the past decades, considerable attention has been devoted to the understanding of the bubble-related sound generation mechanisms. The collective oscillations of bubble clouds suggested by Carey and Bradley[6] and Prosperetti[7,8] appear to be a reasonable explanation of the potential noise source in the low (<1 kHz) frequency range.

The acoustic behavior of bubbly water involves more or less both situations of active sound generation and passive acoustic scattering, depending on the evolutionary stage the bubbles have attained. A detailed description of the life history of a bubble cloud has been documented in Monahan and Lu,[5] which identified two types of whitecaps, A and B, and two associated stages of bubble plumes, α and β, along with an older plume stage γ. Stage α is associated with the air entrainment process that gives rise to a bubble cloud, which can penetrate up to 1 m or more below the sea surface. The duration of active sound generation from entrained bubbles is of the same order as the active breaking duration, which usually lasts only 1–2 s and is many orders of magnitude shorter than that of the bubbly event.[9] Stages β and γ bubble plumes are stabilized by some still unclear mechanisms and are difficult to excite acoustically. The most important properties responsible for noise emission during a breaking event, such as the mechanism by which the bubble cloud is excited, the bubble spectrum generated in the wave-breaking process, and the physical parameters of bubble oscillation, are usually underestimated or ignored because of the short duration and their complicated nature.

In explaining how the bubbles are excited, several mechanisms are available. Medwin and Beaky[10] describe the oscillations as shock excited, by which they assume that the oscillations begin suddenly and are not forced from then on. Unfortunately, they did not give the nature of the "shock." Kerman[11] has proposed a model in which bubbles are excited by turbulent pressure fluctuation. It has been recognized by many researchers that the forcing mechanism for bubble oscillations is air entrainment and bubble fission.[3,12] On the theoretical side, Oğuz[13] proposed an effective pressure model by considering a distribution of bubbles at the air-water interface. Based on Oğuz's model and experimental observation showing that it is the newly entrained bubbles at the leading edge of the breaker that are the primary source of sound from breaking waves,[3,14] Means and Heitmeyer[15] developed a more physically realistic bubble cloud excitation model by treating newly formed bubbles as point sources. Numerical modeling is available to describe the evolution of two-dimensional (2D) surface waves[16] and the impact of two water masses.[17] Unfortunately, their modeling is unable to continue beyond the point at which the surface impacts on itself. In the absence of a quantitative physical theory, the study of wave breaking is based in large part on experiments in both the laboratory and the field, supplemented by process-oriented models.

The key quantity in estimating the low-frequency noise generated by the collective oscillations of bubble clouds is the knowledge of the space-time distribution of the entrained bubbles, which is usually referred to as the void fraction. In many theoretical studies, void fraction of the order $O(0.01)$ is widely used. Lamarre and Melville made 2D[18] and three-dimensional (3D)[19] measurements of laboratory breaking waves and field measurements[20] and concluded that void

[a]Author to whom correspondence should be addressed; electronic mail: xuemeichn@yahoo.com

fractions well above 1% are present for significant times in bubble clouds generated by breaking waves. During the very short time period right after the breaking, the void fraction could be as high as 20%,[21] even 30%.[18,19,22] The acoustic behavior of bubbles in water could be considerably underestimated if the transient large bubbles are not considered.

It can be concluded that the first couple of seconds after the breaking is very crucial in understanding the dynamic coupling between the surface-wave evolution and air entrainment. In view of the difficulty in gathering oceanic field data on the role of bubble clouds in low-frequency sound generation, a numerical model that relates the acoustic signature to the hydrodynamic properties will be a very useful tool in understanding the air entrainment mechanism and noise emission by pulsating bubble clouds.

In this paper, a coupled hydrodynamics-acoustics noise prediction model is developed based on the decomposition of variables. The total pressure is decomposed into a "hydrodynamic" part and an "acoustic" part. Accordingly, there are three solution steps, namely, hydrodynamics simulation, near-field source calculation and far-field noise prediction. The hydrodynamics step simulates the air entrainment and the subsequent bubble oscillation processes, and provides the hydrodynamic properties that are necessary for the acoustic pressure calculation. In order to account for the acoustic effects of air bubbles in the flow, a numerical algorithm has been developed in calculating acoustic wave propagation in irregular regions.

While a direct simulation resolving all the turbulent scales of wave breaking is still not available, one may start with an idealized experiment of two impacting cylindrical water jets. The model and experimental setup provided in Kolaini *et al.*[12] is used here as a benchmark problem.

II. FORMULATION

The pressure fluctuations resulting from a flow event could be classified as propagating and nonpropagating pressure fields. Sound waves are propagating pressure fluctuations and are typically several orders of magnitude smaller than the variations in the flow field that accounts for flow acceleration. Only a small portion of the energy propagates away from the flow as sound; the majority of pressure variations are confined to the flow itself, providing the volume forces necessary to balance fluctuations of local momentum. In a sound field, pressure fluctuation levels are of order $\bar{\rho}c_0\bar{u}$, while in a pseudo-sound field, which is the nonpropagating part, they are $\bar{\rho}\bar{u}^2$, being essentially independent of the sound speed c_0. The symbols $\bar{\rho}$, $\bar{u}$, and $\bar{c}_0$ represent the mean values of the density, fluid velocity, and sound speed, respectively. Experiments have shown that for frequencies greater than 500 Hz the sound radiated by breaking waves in the laboratory correlated with the energy dissipated,[23] and that on the order of 10^{-8} of the mechanical energy dissipated was radiated as sound.[14] Even in a more violent jet flow only about 10^{-4} of the turbulent energy actually escapes as sound.[24] As a consequence, the pressure and density perturbations responsible for sound radiation are invariably overwhelmed by the mean flow quantities. This realization led to the development of Lighthill's[25] classical theory of aerodynamical sound generation.

Since only the region in the vicinity of the sound source, in particular the wave breaking zone, is nonlinear, it is possible to divide the area of interest into near field and far field. In an unsteady compressible fluid there exists both hydrodynamic and acoustic pressure fluctuations; their relative strength usually depends on the location of the point of observation. For the situation considered here, the region where water impact occurs, the hydrodynamic pressure fluctuations dominate, whereas sufficiently far outside this region, the pressure in the sound field, emitted from the region of violent motion, is larger than the local hydrodynamic pressure fluctuations.

Extending Lighthill's idea of acoustic analogy, which relies on the separation of the acoustic perturbations and the solution of the resulting linearized Euler equations, the variables of the problem are decomposed into hydrodynamic part and the acoustic perturbation, i.e.,

$$P = P_h + P_a, \quad \rho = \rho_h + \rho_a, \tag{1}$$

where the subscript "h" represents the hydrodynamic part and "a" designates the acoustic perturbation. The hydrodynamic pressure is subject to the constraint $\rho_h \leqslant \rho_0$, where ρ_0 is the undisturbed fluid density.

In the absence of an external force, viscosity, and heat conduction, the general equations of fluid motion expressing mass and momentum balance in the fluid, assuming index summation convention, is formulated as

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u_i)}{\partial x_i} = 0, \tag{2}$$

$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial(\rho u_i u_j)}{\partial x_j} = -\frac{\partial P}{\partial x_i}, \tag{3}$$

where $u_i (i=1,2,3)$ stands for the components of the total fluid velocity, and ρ and P are the total fluid density and pressure.

Under isentropic conditions the pressure is a function of density only, and the compressibility of the fluid is given by

$$\kappa = \frac{1}{\rho}\frac{d\rho}{dP}. \tag{4}$$

By differentiating Eq. (2) with respect to t and Eq. (3) with respect to x_i and using Eq. (4) and the relation $c^2 = 1/\rho\kappa$, one obtains a version of the general wave equation[26]

$$\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} - \frac{\partial^2(\rho u_i u_j)}{\partial x_i \partial x_j} = \nabla^2 P. \tag{5}$$

Upon substituting Eq. (1) into Eq. (5), one has

$$\frac{1}{c^2}\frac{\partial^2 P_a}{\partial t^2} - \nabla^2 P_a = -\frac{1}{c^2}\frac{\partial^2 P_h}{\partial t^2} + \nabla^2 P_h + \frac{\partial^2(\rho u_i u_j)}{\partial x_i \partial x_j}. \tag{6}$$

Since the acoustic perturbations are much smaller than the hydrodynamic counterparts, one may use the approximation

$$\frac{\partial^2(\rho u_i u_j)}{\partial x_i \partial x_j} \cong \rho_0 \frac{\partial^2(u_i u_j)}{\partial x_i \partial x_j}. \tag{7}$$

It is noticed that if the compressibility is zero, which is the situation when the hydrodynamic simulation is conducted, the hydrodynamic pressure field is determined by the velocity field via the equation

$$\nabla^2 P_h = -\frac{\partial^2(\rho_h u_i u_j)}{\partial x_i \partial x_j} \cong -\rho_0 \frac{\partial^2(u_i u_j)}{\partial x_i \partial x_j}, \tag{8}$$

in which the resulting pressure is of the order $\rho_h u^2$. This pressure is truly hydrodynamic in that it is not influenced by the compressibility.

In a realistic situation, under subsonic conditions, Mach number $M = u/c \ll 1$, compressibility has only a very small effect, and the pressure field therefore is essentially an incompressible fluid. Notice that hydrodynamic pressure is characterized by the property that the pressure fluctuation is proportional to the square of the velocity fluctuation. This is in contrast to a sound field, in which the sound pressure is proportional to the first power of the velocity fluctuation, i.e., $\bar{\rho} c_0 \bar{u}$.

The wave equation

$$\frac{1}{c^2}\frac{\partial^2 P_a}{\partial t^2} - \nabla^2 P_a = -\frac{1}{c^2}\frac{\partial^2 P_h}{\partial t^2} \tag{9}$$

is then obtained, which is a nonhomogeneous linear equation. The fluid motion, in particular the pressure variations, serve as monopoles in driving the acoustic field.

As a consequence of this treatment, two regions of hydrodynamic pressure can be distinguished. One region is that in and around the most active field, where the pressure is essentially locally determined. This region contains all the hydrodynamic nonlinearities and is identified as the near field. The other is the more remote region that merely responds to the driving field and as a consequence, the hydrodynamic pressure (pseudo-sound) gives way to real sound at that distance. In this far-field region the hydrodynamic motion of the fluid is very small and the characteristic period of oscillation is long, compared with the period in the sound field which has been generated by the rapid fluctuations in the violent region.

A problem that arises from this treatment is the ignorance of the interaction between the hydrodynamic motion and the acoustic field. The errors incurred by neglecting the effects of acoustic radiation may be unreasonably large when the acoustic frequency is close to the bubble resonance.[27] It has been suggested by Leighton *et al.*[28] that the bubble oscillations are nonlinear within the breakers at the surf zone where the void fraction is the largest during the bubble cloud evolution. In recent papers,[29,30] an algorithm used to solve the linear propagation through liquids containing bubbles is presented and the treatment accounting for the acoustic radiation is discussed. A small correction to the bubble surface pressure can be made in the linear propagation formulation to account for the acoustic radiation through the modification of boundary conditions. The acoustic pressure is, in many cases, zero on the true free boundaries, such as the air-water interfaces. If the true and the hydrodynamic free boundaries were the same, the acoustic pressure would remain zero there. Accordingly, one can impose homogeneous Dirichlet boundary conditions for the acoustic pressure. In actuality, the acoustic radiation displaces the boundary slightly, therefore one gets an inhomogeneous Dirichlet condition for the acoustic pressure on the hydrodynamic free boundary, the inhomogeneity being proportional to the normal derivative of the hydrodynamic pressure times the displacement of the boundary due to acoustic effects.

As explained above, the variable decomposition allows for separate treatments of the large scale fluid motion and the small sound perturbation superimposed upon it. Two sets of equations governing the fluid motion and acoustic propagation, respectively, are obtained. Existing computational fluid dynamics (CFD) codes can be used to solve the flow motion while the wave propagation needs special treatment since cavities are present in the fluid. If the interactions between the fluid motion and acoustic field are neglected, the hydrodynamic simulation can be conducted independently without any feedback from the acoustic field.

A. Hydrodynamic simulation

The collision of water masses usually involves several complex features, such as free surface movement and deformation, free surface instability and breakup, and air cavity entrainment. For more general problems, a number of algorithms that are based on Lagrangian (moving) and Eulerian (fixed) grids have been proposed over the years. The most recent and successful one to simulate the full Navier–Stokes equation seems to be the front tracking method. Navier–Stokes equations are solved by the standard projection method with an explicit representation of the interface superposed on a Eulerian grid. On the other hand, in problems where vorticity is negligible, boundary element methods are preferable. However, despite the success these methods have achieved, all of these methods can only handle the evolving free surface up to the point of breaking. Liquid on liquid impacts correspond to a breakdown of classical numerical methods. In this paper, a computational model based on a generalized hydrodynamics theory designed to treat violent liquid collisions is used. This method has been successfully used to calculate the collision of two liquid cylinders,[31] shallow water plume formation,[32] and explosive cratering in water-covered sand.[33]

The generalized hydrodynamic model is based on a system of conservation laws subject to a one-sided density constraint. For convenience, the conservation laws stated above are repeated here. The equations are, in the presence of gravity,

$$(\rho_h)_t + \nabla \cdot (\rho_h \mathbf{u}) = 0, \quad \mathbf{x} \in \mathbf{D}, \tag{10}$$

$$(\rho_h \mathbf{u})_t + \nabla \cdot (\rho_h \mathbf{u}\mathbf{u}) = -\rho g \mathbf{k} - \nabla P_h, \tag{11}$$

subject to the density constraint

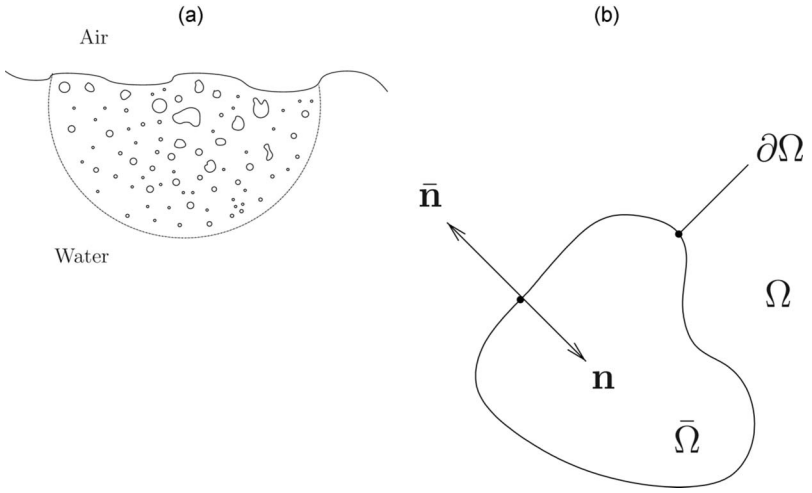


FIG. 1. (a) Sketch of the near field with a bubble cloud in an infinite water body Ω. (b) Normal $\mathbf{n}$ on bubble surface.

$$\rho_h \leqslant \rho_0, \tag{12}$$

where $\mathbf{D}$ is the whole computational region under consideration, including the liquid region in which $\rho(\mathbf{x},t)=\rho_0$ and the disjoint subset of the nonliquid region.

While this method could be classified as a volume of fluid (VOF) approach, it has several distinguishing features that do not exist in the regular method. For example, this formulation allows for the existence of regions of "sprays," in which $0<\rho<\rho_0$. The constraint (12) does not require that space be delineated into regions, separated by sharp boundaries, where $\rho=\rho_0$ or $\rho=0$, as is assumed in the classical theory. Such regions can be expected when the free surfaces become unstable.

Two versions of the computer code have been implemented for this model, BUB2D for both axially symmetric and two-dimensional problems and BUB3D for three-dimensional problems.

B. Wave propagation in fluid containing bubbles

The presence of bubbles in a fluid causes discontinuities in the physical properties of the fluid and thereby affects the passage of sound. In recent years, the acoustic behavior of bubbles in fluids has been a topic of increasing interest in studying propagation, attenuation, scattering, and reverberation.

Traditional numerical methods, such as finite difference schemes, involve space and time discretization and boundary condition specification. For a medium containing multiple complex moving boundaries, such as pulsating bubbles, imposing boundary conditions in a step-by-step fashion is not feasible.

In recent papers, Chen *et al.*[29,30] describe an algorithm for calculating linear wave propagation in irregular regions, addressing the problem when air bubbles are present in the medium. The algorithm depends in the first instance on the fact that moving boundaries can be replaced by time dependent sources. By replacing the boundaries with sources, the computational domain is extended to the whole space, including the regions occupied by the inhomogeneities. Although boundary conditions do not appear in the numerical scheme explicitly, when replacing the boundaries they must be specified on a physical base in order to determine the source function. When the interaction between the fluid motion and the acoustic field is neglected, homogeneous boundary conditions, corresponding to a zero pressure distribution on the bubble surface, should be imposed. In the following, the solution corresponding to the homogeneous Dirichlet boundary conditions will be presented.

Figure 1(a) is a sketch of a bubble cloud resulting from wave breaking. The hemispherical volume shows the hydrodynamic region to which the near-field acoustic pressure field refers. Let $\partial\Omega$ represent the total closed surfaces of the cavities with total interior space $\bar{\Omega}$ occupied by the air or air-water mixture. Ω denotes the liquid medium outside the cavities with a sound speed c. Therefore, $\bar{\Omega}\cup\Omega\subset\Re^3$. The scattered (or total) pressure field $P(\mathbf{x},t)$ in Ω is sought for given Cauchy conditions $P(\mathbf{x},t=0)$ and $P_t(\mathbf{x},t=0)$ such that (For convenience, hereafter P is used to denote acoustic pressure and the subscript "a" is omitted.)

$$\frac{1}{c^2}P_{tt}-\nabla^2 P=S_h(\mathbf{x},t),\quad \mathbf{x}\in\Omega,\quad t>0, \tag{13}$$

$$P(\mathbf{x},t)=0,\quad \mathbf{x}\in\partial\Omega, \tag{14}$$

$$\lim_{r\to\infty} r\left(P_r+\frac{1}{c}P_t\right)=0, \tag{15}$$

where S_h is the hydrodynamic source given in Eq. (9) and $S_h=-1/c^2\,\partial^2 P_h/\partial t^2$. Since one is interested here in the case of wave propagation in a fluid medium containing irregular air voids, homogeneous Dirichlet boundary conditions are imposed in Eq. (14) and the effects of acoustic radiation are excluded from consideration. Condition (15), in which r is the radial distance, is the Sommerfeld radiation condition.

Unlike the traditional finite difference or finite element schemes, the algorithm proposed in Chen *et al.*[29,30] takes a different approach by replacing the irregular surfaces with properly chosen sources and extending the computational domain to include the space occupied by the inhomogeneities. In their algorithm, the mathematical treatment of the bound-

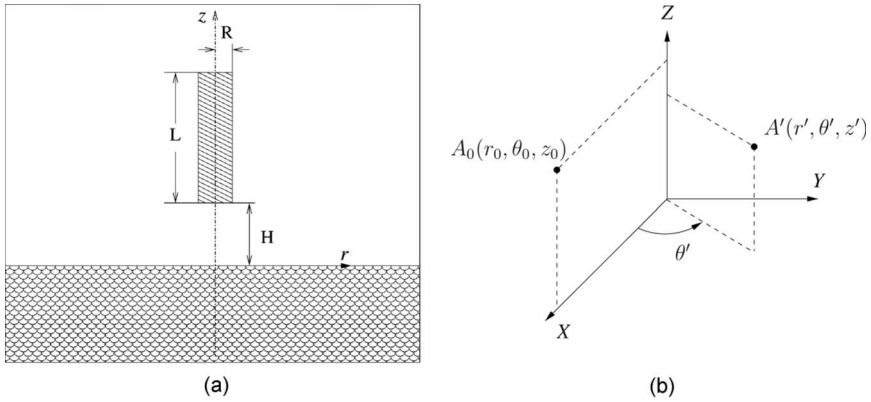


FIG. 2. (a) Experimental setup given in Kolaini *et al.* (1993). (b) Axisymmetric coordinate for the calculation.

ary replacement is characterized by the addition of a compensating source term S_∞ on the right hand side of the wave Eq. (13), i.e.,

$$\frac{1}{c^2}P_{tt} - \nabla^2 P = S_h(\mathbf{x},t) + S_\infty(\mathbf{x},t), \quad \mathbf{x} \in \Re^3, \quad t > 0, \quad (16)$$

where

$$S_\infty(\mathbf{x},t) = -\frac{1}{c}|\nabla_\chi| P_t - \nabla\chi \cdot (\nabla P)_+. \quad (17)$$

Equation (17) is the general form of the source term in three-dimensional space due to boundary replacement and

$$\chi = \begin{cases} 1 & \text{inside } \Omega \\ 0 & \text{otherwise} \end{cases}. \quad (18)$$

In Eq. (17) the subscript "+" indicates that ∇P is evaluated on the side of $\partial\Omega$ which is in Ω.

The determination of the source term S_∞ is not very straightforward. A one-dimensional analogy is a more intuitive approach that may help explain the procedure. This approach is described in Appendix A.

For the current problem where the near-field hydrodynamic fluctuation is also considered, one may think of the source as consisting of two parts, the hydrodynamic pressure fluctuation S_h and the source present at the irregular boundaries S_∞. Note that the boundary source part S_∞ only exists at the boundaries. It disappears in the liquid region where the normal wave equation is applicable. If we use $S(\mathbf{x},t)$ to denote the total source, we have

$$S(\mathbf{x},t) = S_h(\mathbf{x},t) + S_\infty(\mathbf{x},t). \quad (19)$$

Since the computational domain is extended to the whole free space, one may use a regular wave equation solver in infinite space with given sources. Here the inhomogeneous wave equation is solved by quadratures using Green's function. Let $P^n(\mathbf{x})$ and $P^{n+1}(\mathbf{x})$ denote the values of $P(\mathbf{x},t)$ at the nth time step $t=t^n$ and $(n+1)$th time step $t=t^{n+1}=t^n+\Delta t$, respectively. By using Green's function for the wave equation in three dimensions, one obtains a solution of Eq. (16),

$$P^{n+1}(\mathbf{x}) = \frac{c^2}{4\pi}\int_{t^n}^{t^n+\Delta t}\int_{\zeta\in S^2}(t^n + \Delta t - t')$$
$$\times S(\mathbf{x} + c(t^n + \Delta t - t')\boldsymbol{\zeta}, t')d\boldsymbol{\zeta}dt' + \frac{1}{4\pi}\int_{\zeta\in S^2}$$
$$\times P^n(\mathbf{x} + c\Delta t\boldsymbol{\zeta})d\boldsymbol{\zeta} + \frac{1}{4\pi}\int_{\zeta\in S^2}\Delta t P_t^n(\mathbf{x} + c\Delta t\boldsymbol{\zeta})d\boldsymbol{\zeta}$$
$$+ \frac{1}{4\pi}\int_{\zeta\in S^2} c\Delta t\boldsymbol{\zeta}\cdot\nabla P^n(\mathbf{x} + c\Delta t\boldsymbol{\zeta})d\boldsymbol{\zeta}, \quad (20)$$

where S^2 is the surface of a unit sphere centered at $\mathbf{x}(x,y,z)$.

Calculating $P^{n+1}(\mathbf{x})$ involves evaluating $S_\infty(\mathbf{x},t^n+\Delta t)$, which itself needs the information of $P^{n+1}(\mathbf{x})$. The one-dimensional analogy (Appendix A) tells us that $S_\infty(\mathbf{x},t^n+\Delta t)$ can be determined using earlier data. The mathematical derivation of the source is given in Appendix B.

III. NUMERICAL SIMULATION OF NOISE GENERATED BY IMPACTING CYLINDRICAL WATER JETS

A. Near-field results

The experimental setup described in the paper by Kolaini *et al.*[12] is sketched in Fig. 2(a). In particular, the case when $R=5.4$ cm, $H=15$ cm, and $L=45$ cm will be considered as a benchmark here. In the experiment, Kolaini *et al.* measured the pressure time history at a fixed position under the water at $r=40$ cm and $z=-20$ cm, where $z=0$ corresponds to the initial water tank air-liquid surface. It should be recognized that the measurement was conducted in the near field where the medium motion is dominated by the local hydrodynamic flow. The hydrophone is sensitive to both hydrodynamic as well as acoustic pressure fluctuations. Moreover, the existence of the water tank changed the emission spectrum so that no absolute measurements were available.

A bubble plume is generated by dropping a fixed cylindrical column of water into a still water tank. After the impact, the water jet opens a cavity as it descends into the water surface. As the cavity grows bigger and the leading edge advances, a "neck" is formed. Water surrounding the neck rushes radially toward the axis of symmetry, pinching off a large bubble plume. Detailed photographic images can be found in Kolaini *et al.*[12] and (Szymczak *et al.*).[31] The de-

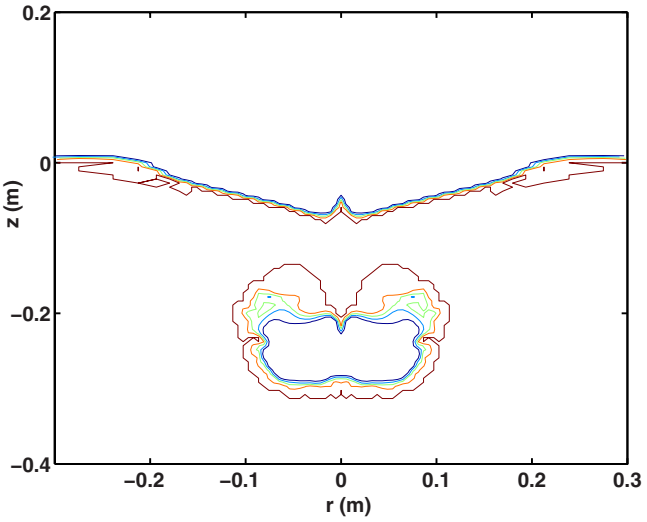


FIG. 3. (Color online) Calculated fluid density contour.

tached bubble plume undergoes volume pulsations, constituting a strong acoustic source radiating low-frequency sound.

As a first step of the noise prediction strategy outlined in the preceding section, the hydrodynamic simulation of the bubble entrainment by the impact of two cylindrical water jets and the subsequent bubble oscillations has been carried out by Szymczak *et al.*[31] using the generalized hydrodynamic model. In the calculation, axial symmetry is assumed so that the code BUB2D could be used. Figure 3 shows the simulated bubble density contours when the liquid wall has just closed.

Spray is evident in Fig. 3 where multiple density lines exist inside the bubble and at the air-liquid surface. This matches the observation obtained from cinematography in the experiment of Kolaini *et al.*,[12] which suggested that the "bubble plume" could be a large air-filled bubble with a roughened surface that contains air-water mixture.

Hydrodynamic pressure variations, in the form of the second order time derivative, as given in Eq. (9), serve as monopoles for the acoustic radiation. The hydrodynamic source term at the probe position $(r,z)=(40,-20)$ cm is plotted in Fig. 4(a). The acoustic pressure generated in this procedure is calculated using the formulation outlined in the previous section. The pressure time history at the probe position is shown in Fig. 4(b). The acoustic pressure, with an amplitude several orders of magnitude smaller than that of the hydrodynamic pressure, also exhibits a low-frequency oscillation. Figure 5 shows the comparison of the measured pressure (a) with the calculated total pressure (b) which is defined in Eq. (1) as the summation of the hydrodynamic pressure and the acoustic pressure. Generally speaking, both curves show low-frequency sinusoidal oscillations with damped amplitudes. Very similar to the measurement data, the calculated pressure has a second peak at $t\cong 1.05$ s. This is probably produced by the fallback of the liquid jet generated by the closing of the liquid wall. The energy spectrum comparison for case $L=45$ cm is displayed in Fig. 6. It shows some similar features, namely, the peaks occurring at the fundamental frequencies and the two frequencies are very close. The peak frequency of the energy spectrum of the

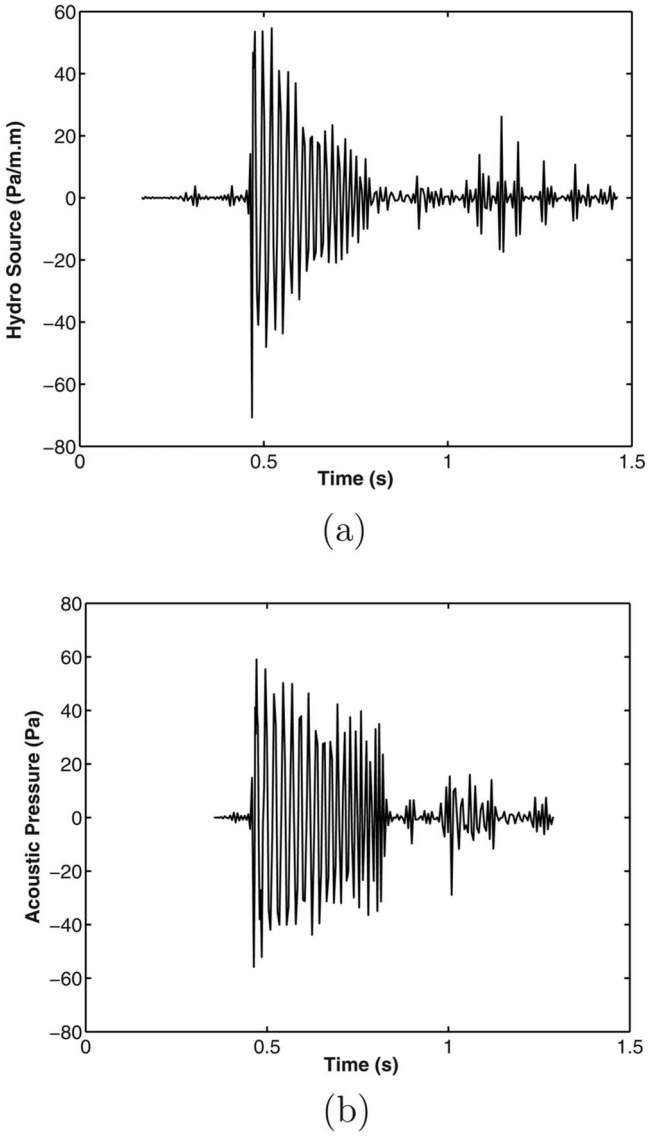


FIG. 4. (a) Hydrodynamic pressure variation $-1/c^2\,\partial^2(P_h)/\partial t^2$ at the probe position. (b) Computed acoustic pressure time history at the probe position.

experiment is reported as 43 Hz, and in the calculation it is 43.75 Hz. The calculated result shows a faster decay than the experimental data. This could be caused by the repeated liquid collisions, as well as numerical dissipation.[31]

B. Far-field noise

The calculation of far-field pressure can follow the same procedure as for the near field. A problem that arises for far-field calculations is the calculation efficiency. As stated in the mathematical formulation section, one may extend the calculation domain to infinity and include the effect of the presence of the boundary in the source term. That is to say, the water-atmosphere interface is also considered as part of the source. As the region increases, the calculation will become very inefficient. A practical way to avoid the very long free surface is to use the method of images. To that end, the water-atmosphere interface will be treated as a flat surface. From experimental observation and hydrodynamic simulation, surface deformation only occurs in a small limited re-

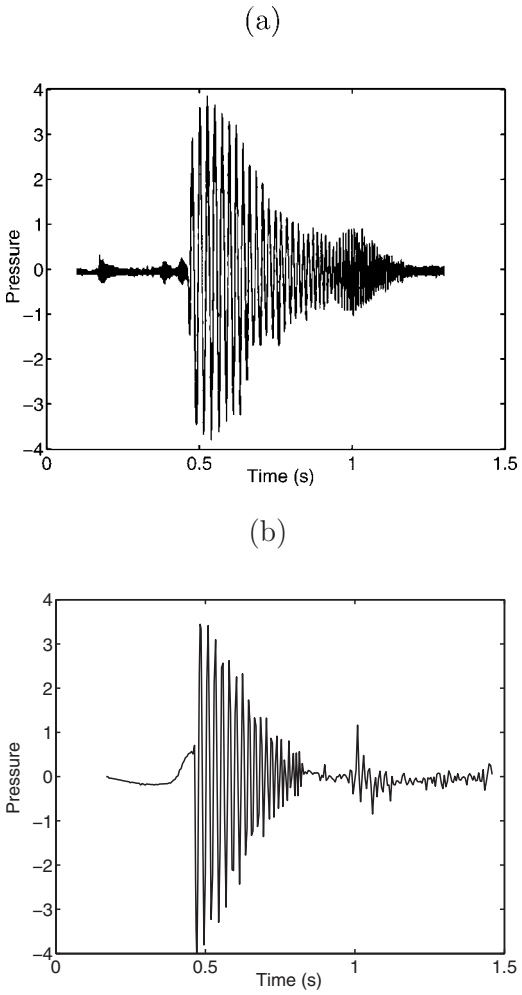


FIG. 5. Near-field pressure comparison: (a) Pressure measured at the probe, (b) calculated total pressure at the probe.

gion, i.e., the region very close to the bubble. For a surface that is only a few radii away from the bubble, the surface is relatively "flat." Therefore, it is reasonable to treat the interface as a flat surface for a much larger space.

The depth of submergence of the entrained bubbles is typically a few centimeters (this is true in the case of the benchmark problem). From the theoretical point of view, the presence of the neighboring free surface confers a dipole character to the basically monopole nature of the acoustic radiation. If $p_b(t)$ denotes the pressure on the liquid side of the bubble interface, each bubble will radiate a pressure field given, far from the bubble, by

$$P = 2d\cos\theta\left(\frac{R_b}{rc}\right)\dot{p}_b\left(t-\frac{r}{c}\right), \tag{21}$$

where d is the depth of submergence, R_b is the radius of the bubble, and r and θ are local spherical coordinates with origin at the free surface directly above the bubble, i.e., at the center of the dipole.

With the knowledge of the hydrodynamic source and acoustic pressure distribution in the near field, one can readily evaluate the pressure in the far field. For the axisymmetric case, the expression is

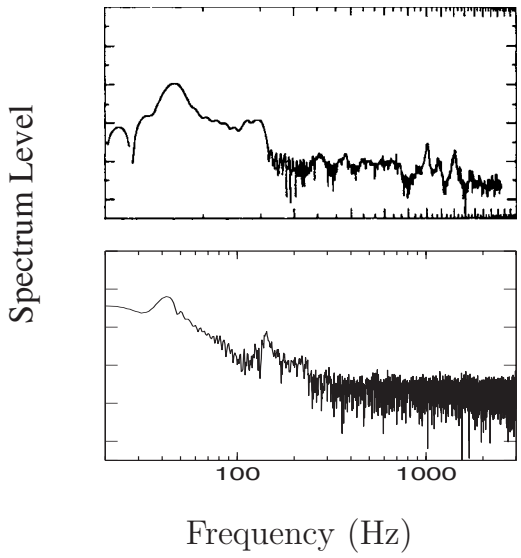


FIG. 6. Near-field spectrum comparison. Upper panel: measurements; lower panel: calculation.

$$P(r_0,z_0,t) = \frac{1}{4\pi}\int_r\int_z\int_0^{2\pi}\frac{S(r',z',t-|A'A_0|/c)}{|A'A_0|} \times d\theta' r' dr' dz', \tag{22}$$

where $|A'A_0|$ is the distance between a source point $A'(r',\theta',z')$ and a point in far field $A_0(r_0,\theta_0=0,z_0)$ in cylindrical coordinates [refer to Fig. 2(b)]. The distance is given by

$$|A'A_0| = [r_0^2 + (r')^2 - 2r_0 r'\cos\theta' + (z_0-z')^2]^{\frac{1}{2}}. \tag{23}$$

Figure 7 shows the theoretical and computed far-field acoustic pressure projected on *Range-Depth* plane at $t=0.8$ s. The results displayed are in the range of about 1–5 wavelengths. The distribution of the pressure field and the phase variations agree qualitatively.

IV. CONCLUSIONS

In this paper, a coupled hydro-acoustic formulation has been developed to relate the physical parameters of the flow motion to the acoustic signatures it produces. Based on variable decomposition, two sets of equations are obtained in solving the flow motion and the sound generation and propagation, respectively. Without considering the near-field interaction between the flow motion and the sound field, the hydrodynamic simulation is conducted independently. The generalized hydrodynamic formulation makes it possible to predict not only the entrainment of air, but also the subsequent pulsations which provide acoustic sources. A numerical algorithm designed for treating wave propagation in irregular regions is used to calculate wave propagation in fluid containing bubbles resulting from the jets impacting.

Near-field noise generated by impacting cylindrical water jets is predicted. Favorable results are obtained upon comparison with the measurements. The sound shows low-

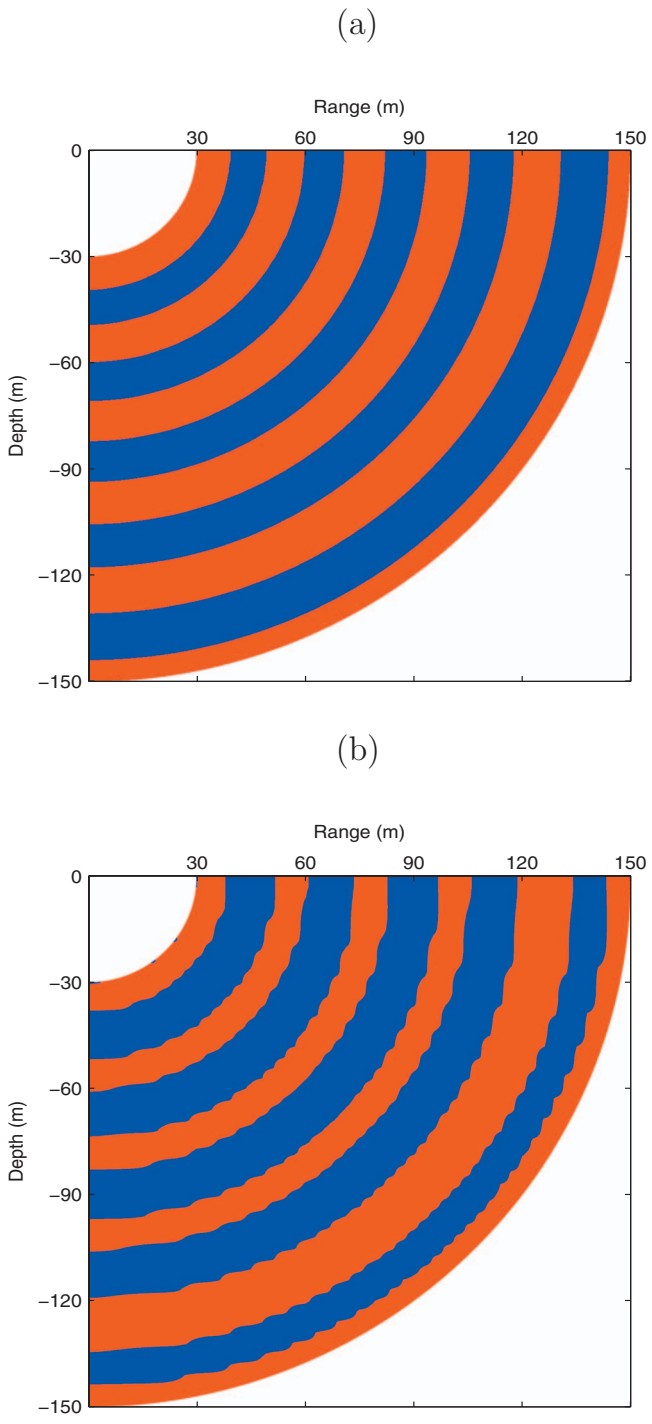


FIG. 7. Far-field pressure comparison: (a) Pressure estimated using Eq. (21) at $t=0.8$ s; (b) computed pressure at $t=0.8$ s. Red stripes denote positive pressure, and blue ones represent negative values.

frequency sinusoidal oscillations with a damped amplitude. The far-field acoustic pressure is also formulated using the sources as modeled within the near field.

It has been recognized that the existence of salt could reduce bubble coalescence and enhance bubble stability.[34,35] This property could lead to significant changes in bubble distribution and acoustic behavior. It implies that the bubble entrainment mechanism could be different in salt water from that in fresh water and suggests more in-depth studies of the forcing mechanism in open-ocean breaking waves are needed.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research base funding at the Naval Research Laboratory.

APPENDIX A: DETERMINATION OF SOURCE TERM S_∞

Consider a one-dimensional wave propagation problem in a half space $x>0$, i.e., the boundary is set at $x=0$. The mathematical description of the problem is

$$\frac{1}{c^2}\phi_{tt} - \phi_{xx} = 0, \quad x > 0, \quad t > 0, \tag{A1}$$

$$\phi(x,t) = 0, \quad x < 0, \tag{A2}$$

$$\phi(0^+,t) = 0, \tag{A3}$$

where "+" denotes the "fluid" side of the boundary, that is, the side in which the propagation takes place.

The wave equation could be rewritten in either of the two equivalent forms,

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)\left(\frac{1}{c}\frac{\partial \phi}{\partial t} - \frac{\partial \phi}{\partial x}\right) = 0 \quad \text{or}$$

$$\left(\frac{1}{c}\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right)\left(\frac{1}{c}\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x}\right) = 0. \tag{A4}$$

The D'Alembert solution indicates that the solution is of the form

$$\phi(x,t) = f_1(x+ct) + f_2(x-ct). \tag{A5}$$

Since $1/c\,\phi_t + \phi_x = 2f_1'(x+ct)$, we may think of $f_1(x+ct)$ and $1/c\,\phi_t + \phi_x$ as describing a "leftward" moving wave. The same applies to $1/c\,\phi_t - \phi_x$, likewise, $f_2(x-ct)$ and $1/c\,\phi_t - \phi_x$ describe "rightward" moving waves.

The leftward propagating signal which arrives at $(0^+,t)$ from $(x',t-x'/c)$ will generate a "kick" at $(0^+,t)$. This kick may be represented by a source $-\delta(x)(1/c\,\phi_t + \phi_x)$. Thus, we may regard the solution of Eqs. (A1)–(A3) as analogous to solving the problem

$$\frac{1}{c^2}\phi_{tt} - \phi_{xx} = S_\infty(x,t), \quad -\infty < x < +\infty, \quad t > 0, \tag{A6}$$

where

$$S_\infty(x,t) = -\delta(x)\left(\frac{1}{c}\phi_t(0^+,t) + \phi_x(0^+,t)\right). \tag{A7}$$

Similarly, the three-dimensional source could be derived as

$$S_\infty(\mathbf{x},t) = -\frac{1}{c}|\nabla_\chi|\phi_t - \nabla_\chi \cdot (\nabla_\phi)_+, \tag{A8}$$

where

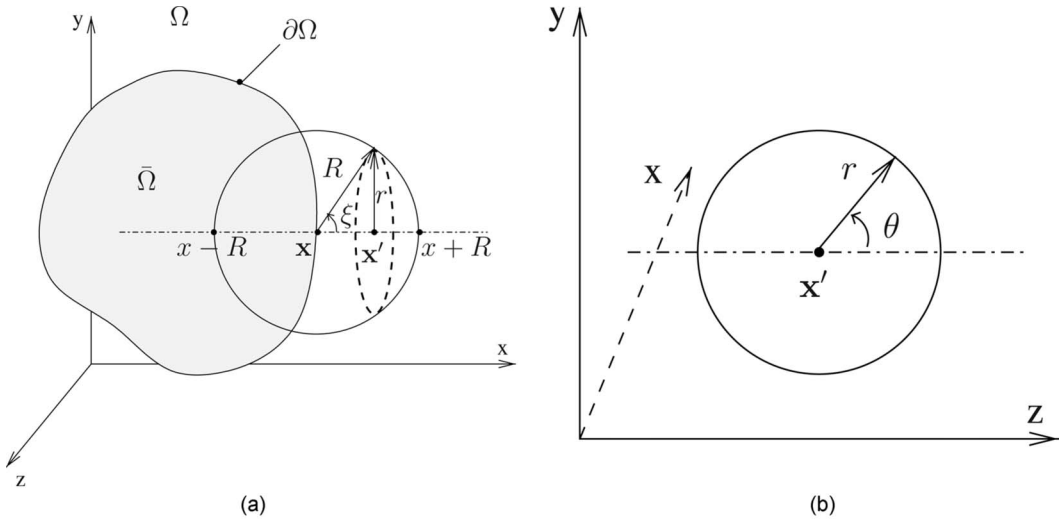


FIG. 8. Sketch of the coordinates in calculating $\mathbf{P}$ and $\hat{\mathbf{P}}$.

$$\chi = \begin{cases} 1 & \text{inside liquid} \\ 0 & \text{otherwise,} \end{cases} \tag{A9}$$

and the subscript "+" indicates that $\nabla\phi$ is evaluated on the side of the fluid.

Expression (A8) could be justified by the fact that

$$\nabla\chi = -\mathbf{n}\delta_{\partial\Omega}, \quad |\nabla\chi| = \delta_{\partial\Omega}, \tag{A10}$$

where the normal $\mathbf{n}$ points to $\bar{\Omega}$ [Fig. 1(b)] and $|\nabla_\chi|$ is the Dirac measure concentrated on $\partial\Omega$.

APPENDIX B: NUMERICAL CALCULATION OF THE SOURCE TERM

Define

$$\mathbf{P}(\mathbf{x},r,t) \equiv \frac{1}{4\pi r^2}\int_{|\mathbf{y}-\mathbf{x}|=r} P(\mathbf{y},t)dS \tag{B1}$$

and

$$\Psi(\mathbf{x},t^n,t) \equiv \frac{c^2}{4\pi}\int_{t^n}^{t}\int_{\zeta\in S^2}(t-t')S(\mathbf{x}+c(t-t')\boldsymbol{\zeta},t')d\boldsymbol{\zeta}dt', \tag{B2}$$

then $\mathbf{P}(\mathbf{x},r,t)$ is the average value of $P(\mathbf{x},t)$ over the sphere centered at $\mathbf{x}$ with radius r.

Let $\mathbf{P}^n(\mathbf{x},r) \equiv \mathbf{P}(\mathbf{x},r,t^n)$ and $\Psi^n(\mathbf{x}) \equiv \Psi(\mathbf{x},t^n,t^n+\Delta t)$, then the solution of wave Eq. (20) is expressed as

$$P^{n+1}(\mathbf{x}) = \mathbf{P}^n(\mathbf{x},c\Delta t) + \Delta t\mathbf{P}_t^n(\mathbf{x},c\Delta t) + c\Delta t\frac{\partial}{\partial R}\mathbf{P}^n(\mathbf{x},R)\bigg|_{R=c\Delta t} + \Psi^n(\mathbf{x}). \tag{B3}$$

One may further define $\hat{\mathbf{P}}$ as

$$\hat{\mathbf{P}}(\mathbf{x},r,t) \equiv \frac{1}{2\pi}\int_0^{2\pi} P(\mathbf{x}+\mathbf{j}r\sin\theta+\mathbf{k}r\cos\theta,t)d\theta, \tag{B4}$$

where r and θ are the local cylindrical coordinates with origin at $\mathbf{x}'(x',y',z')$ (Fig. 8).

$\mathbf{P}$ is represented as

$$\mathbf{P}(\mathbf{x},R,t) = \frac{1}{2}\int_0^{\pi}\hat{\mathbf{P}}(\mathbf{x}+\mathbf{i}R\cos\xi,R\sin\xi,t)\sin\xi d\xi = \frac{1}{2R}\int_{x-R}^{x+R}\hat{\mathbf{P}}(\mathbf{x}+\mathbf{i}(x'-x),(R^2-(x'-x)^2)^{\frac{1}{2}},t)dx'. \tag{B5}$$

For a point on the boundary $\partial\Omega$ with normal $\mathbf{n}=-\mathbf{i}$, the multiplier of the delta function in the source expression could be derived as

$$\begin{aligned}\frac{1}{c}P_t^{n+1}(\mathbf{x}) + P_x^{n+1}(\mathbf{x}) = {} & \frac{1}{c}P_t^n[\mathbf{x}+\mathbf{i}(c\Delta t)] + P_x^n[\mathbf{x}+\mathbf{i}(c\Delta t)] \\ & + \frac{1}{2}\int_{x-c\Delta t}^{x+c\Delta t}\Bigg[\hat{\mathbf{P}}_{xR'}^n(\mathbf{x}+\mathbf{i}(x'-x),R')\frac{x'-x+c\Delta t}{R'} \\ & + \frac{1}{c}\hat{\mathbf{P}}_{R't}^n(\mathbf{x}+\mathbf{i}(x'-x),R')\frac{x'-x+c\Delta t}{R'} \\ & + \hat{\mathbf{P}}_{R'R'}^n(\mathbf{x}+\mathbf{i}(x'-x),R') + \frac{1}{R}\hat{\mathbf{P}}_{R'}^n(\mathbf{x}+\mathbf{i}(x'-x),R')\Bigg]\Bigg|_{R'} \\ & \times dx' + \left(\frac{1}{c}\Psi_t^n(\mathbf{x}) + \Psi_x^n(\mathbf{x})\right),\end{aligned} \tag{B6}$$

where $R' = [(c\Delta t)^2-(x'-x)^2]^{1/2}$. It shows that $S_\infty(\mathbf{x},t^n+\Delta t)$

could be evaluated using the information of previous time step.

[1]G. M. Wenz, "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956 (1962).
[2]G. E. Updegraff and V. G. Anderson, "*In situ* acoustic signature of low sea state microbreaking," J. Acoust. Soc. Am., **85**, S146 (1989).
[3]H. C. Pumphrey and J. E. Ffowcs Williams, "Bubbles as sources of ambient noise," IEEE J. Ocean. Eng. **15**, 268–274 (1990).
[4]E. C. Monahan and I. G. O'Muricheartaigh, "Whitecaps and the passive remote sensing of the ocean surfaces," Int. J. Remote Sens. **7**, 627–642 (1986).
[5]E. C. Monahan and M. Lu, "Acoustically relevant bubble assemblages and their dependence on meteorological parameters," IEEE J. Ocean. Eng. **15**, 340–349 (1990).
[6]W. M. Carey and M. P. Bradley, "Low-frequency ocean surface noise sources," J. Acoust. Soc. Am. **78**, S1–S2 (1985).
[7]A. Prosperetti, "Bubble-related ambient noise in the ocean," J. Acoust. Soc. Am., **78**, S2 (1985).
[8]A. Prosperetti, "Bubble-related ambient noise in the ocean," J. Acoust. Soc. Am. **84**, 1042–1054 (1988).
[9]P. A. Hwang and W. J. Teague, "Low-frequency resonant scattering of bubble clouds," J. Atmos. Ocean. Technol. **17**, 847–853 (2000).
[10]H. Medwin and M. M. Beaky, "Bubble sources of the Knudsen sea noise spectra," J. Acoust. Soc. Am. **86**, 1124–1130 (1989).
[11]B. R. Kerman, "Underwater sound generation by breaking wind waves," J. Acoust. Soc. Am. **75**, 149–165 (1984).
[12]A. R. Kolaini, R. A. Roy, L. A. Crum, and Y. Mao, "Low-frequency underwater sound generation by impacting transient cylindrical water jets," J. Acoust. Soc. Am. **94**, 2809–2820 (1993).
[13]H. N. Oğuz, "A theoretical study of low-frequency oceanic ambient noise," J. Acoust. Soc. Am. **95**, 1895–1912 (1994).
[14]M. R. Loewen and W. K. Melville, "A model of the sound generated by breaking waves," J. Acoust. Soc. Am. **90**, 2075–2080 (1991).
[15]S. L. Means and R. M. Heitmeyer, "Low-frequency sound generation by an individual open-ocean breaking wave," J. Acoust. Soc. Am. **110**, 761–768 (2001).
[16]D. G. Dommermuth, D. K. P. Yue, R. J. Rapp, E. S. Chan, and W. K. Melville, "Deep-water breaking waves: A comparison between potential theory and experiments," J. Fluid Mech. **89**, 432–442 (1987).
[17]H. N. Oğuz, A. Prosperetti, and A. R. Kolaini, "Air entrapment by a falling water mass," J. Fluid Mech. **294**, 181–207 (1995).
[18]E. Lamarre and W. K. Melville, "Air entrainment and dissipation in breaking waves," Nature (London) **351**, 469–472 (1991).
[19]E. Lamarre and W. K. Melville, "Void-fraction measurements and sound-speed fields in bubble plumes generated by breaking waves," J. Acoust. Soc. Am. **95**, 1317–1328 (1994).
[20]E. Lamarre and W. K. Melville, "Instrumentation for the measurement of void fraction in breaking waves: Laboratory and field results," IEEE J. Ocean. Eng. **17**, 204–215 (1992).
[21]Q. Wang and E. C. Monahan, "The influence of salinity on the spectra of bubbles formed in breaking wave simulations," in *Sea Surface Sound '94*, edited by M. J. Buckingham and J. R. Potters, World Scientific, Singapore, pp. 312–319 (1995).
[22]G. B. Deane, "Sound generation and air entrainment by breaking waves in the surf zone," J. Acoust. Soc. Am. **102**, 2671–2689 (1997).
[23]W. K. Melville and R. J. Rapp, "The surface velocity field in steep and breaking waves," J. Fluid Mech. **189**, 1–22 (1988).
[24]J. E. Ffowcs Williams, "Hydrodynamic noise," Annu. Rev. Fluid Mech. **1**, 197–222 (1969).
[25]M. J. Lighthill, "On sound generated aerodynamically: 1. General theory," Proc. R. Soc. London, Ser. A **211**, 564–578 (1952).
[26]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, 1986).
[27]J. B. Keller and M. Miksis, "Bubble oscillations of large amplitude," J. Acoust. Soc. Am. **68**, 628–633 (1980).
[28]T. G. Leighton, S. D. Meers, and P. R. White, "Propagation through nonlinear time-dependent bubble clouds and the estimation of bubble populations from measured acoustic characteristics," Proc. R. Soc. London, Ser. A **460**, 2521–2550 (2004).
[29]X. M. Chen, S. L. Means, J. C. W. Rogers, and W. G. Szymczak, "Numerical simulation of noise generated by impacting cylindrical water jets," *Proceedings of the 7th European Conference on Underwater Acoustics*, Delft, The Netherlands (2004).
[30]X. M. Chen, J. C. W. Rogers, S. L. Means, and W. G. Szymczak, "An algorithm on direct simulating linear wave propagation in irregular regions," J. Comput. Acoust. (accepted).
[31]W. Szymczak, S. L. Means, and J. C. W. Rogers, "Computations of bubble formation and pulsations generated by impacting cylindrical water jets," J. Eng. Math. **48**, 375-389 (2004).
[32]W. G. Szymczak and J. M. Solomon, "Computations and experiments of shallow depth explosion plumes," NSWCDD/TR-94/156, Dahlgren, Virginia: NSWC (1996).
[33]W. G. Szymczak and J. C. W. Rogers, "Generalized hydrodynamics with viscoplasticity for channeling in saturated sand," NRL/FR/7130-00-9946, Washington, DC: NRL (2000).
[34]A. R. Kolaini, "Effects of salt on bubble acoustic radiation in water," J. Acoust. Soc. Am. **105**, 2181–2186 (1999).
[35]W. M. Carey, J. W. Fitzgerald, E. C. Monahan, and Q. Wang, "Measurements of the sound produced by a tipping trough with fresh and salt water," J. Acoust. Soc. Am. **93**, 3178–3192 (1993).

Stability analysis of thermally induced spontaneous gas oscillations in straight and looped tubes

Yuki Ueda[a)]
Department of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei, Tokyo 184-8588, Japan

Chisachi Kato
Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguroku, Tokyo 153-8505, Japan

(Received 5 December 2007; revised 8 May 2008; accepted 9 May 2008)

A gas in a tube spontaneously oscillates when the temperature gradient applied along the wall of the tube is higher than the critical value. This spontaneous gas oscillation is caused by the thermal interaction between the gas and the tube wall. The stability limit of the thermally induced gas oscillation is numerically investigated by using the linear stability theory and a transfer matrix method. It is well known that an acoustic wave excited by the spontaneous gas oscillation occurring in a looped tube is different from that in a straight tube with two ends; a traveling acoustic wave is induced in a looped tube, whereas a standing acoustic wave is caused in a straight tube. The conditions for the stability limits in both tube types were calculated. The calculated and measured conditions were compared and were found to be in good agreement. Calculations performed by varying the value of the Prandtl number of the gas were used to determine the reasons for the existence of the stability limits of the looped and straight tubes.
 [DOI: 10.1121/1.2939134]

PACS number(s): 43.35.Ud [RR] Pages: 851–858

I. INTRODUCTION

When the temperature gradient imposed along the wall of a tube exceeds a critical value, the gas in the tube begins to oscillate, so that an acoustic wave is excited. This spontaneous gas oscillation is sustained by the energy conversion from heat to acoustic power occurring through the thermal contacts between the gas and the wall of the tube.

The spontaneous gas oscillation can be classified into two types based on the characteristics of the thermally induced acoustic wave. One type of gas oscillation occurs in a tube with two ends, i.e., a straight tube, which excites a standing acoustic wave. The other type of gas oscillation occurs in a looped tube, which causes a traveling acoustic wave.

Rott[1,2] and Rott and Zouzoulas[3] analytically investigated the stability limit of gas oscillation in a straight tube by employing a strict linearization of all basic equations, and they derived the stability curves for the temperature ratio against the tube radius relative to the thickness of the viscous boundary layer. Yazaki *et al.* compared their experimentally obtained stability curves with Rott's analytical ones; this comparison revealed a good agreement between the experimentally and analytically obtained curves.[4] A number of studies have investigated the stability curves in a straight tube.[5–9] This provides a basic understanding of the spontaneous gas oscillation in a straight tube.

The spontaneous gas oscillation in a looped tube has recently attracted more attention compared with that in a tube with two ends; thus, most of the recent work in thermoacoustics has dealt with spontaneous gas oscillation in a looped tube rather than that in a straight tube.[10–12] This is because a traveling acoustic wave can realize an efficient energy conversion.[13,14] Yazaki *et al.* measured the conditions for the stability limit of the spontaneous gas oscillation in a looped tube, and they studied the stability curve of the looped tube.[15] Tanaka[16] and Penelet *et al.*[17] succeeded in calculating the conditions for the stability limit. However, in the case of the looped tube, very few studies have numerically investigated the conditions for the stability limit in detail.

In this paper, we calculate the conditions for the stability limit of the thermally induced gas oscillation in both looped and straight tubes by using the linear stability theory and a transfer matrix method. It will be shown that the obtained calculation results qualitatively agree with the experimental results. By using the calculation in which the Prandtl number of a gas is varied, we address the reasons for the existence of the stability limit in looped and straight tubes.

In Sec. II, the numerically investigated looped and straight tubes are mentioned; in Sec. III, the method for calculating the stability limit of the thermally induced spontaneous gas oscillation is described. Section IV provides the calculation results of the stability curve and compares these results with the experimental results. The calculation results are analyzed in detail in Sec. V. The summary of the results is provided in Sec. VI.

II. CALCULATION MODEL

A. Looped tube and straight tube

The numerically investigated tubes are schematically illustrated in Fig. 1. The total length of the looped and straight

[a)]Electronic mail: uedayuki@cc.tuat.ac.jp

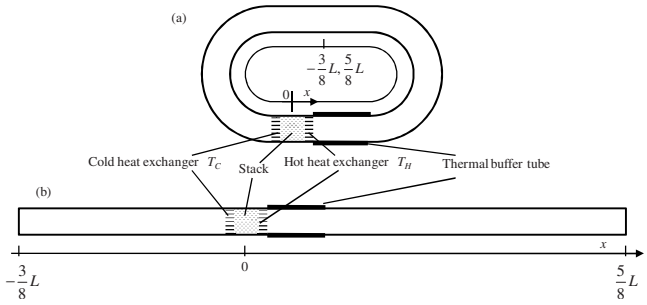


FIG. 1. Schematic illustrations of (a) a looped tube and (b) a straight tube.

tubes is denoted as L. Both tubes are filled with atmospheric air and can be divided into six components: waveguide 1, cold heat exchanger, *stack*, hot heat exchanger, thermal buffer tube, and waveguide 2. The radius of the waveguides and thermal buffer tube is denoted as r_{tube}, and the length of the thermal buffer tube is denoted as L_{tb}. The stack is composed of a porous material with several narrow circular channels. The inner radius of the narrow circular channels in the stack is denoted as r_{stack}. The length of the stack is L_{stack}. The stack is placed between cold and hot heat exchangers; these exchangers are composed of flat plates set in parallel with a spacing of $2r_{\text{hx}}$. The length of the plates is L_{hx}. The porosity of the stack is ϵ_{stack}, and that of the heat exchangers is ϵ_{hx}. In this study, porosity is defined as the ratio of the total cross-sectional area of the flow channels in the stack/heat exchanger to the area of the tube (πr_{tube}^2).

The coordinate x is defined along the axis of the tube. This coordinate is used for both looped and straight tubes. The origin of x is set at the center of the stack. In the case of the straight tube, the closed ends are located at $x=-3L/8$ and at $x=5L/8$ so that the center of the stack will be set at the midpoint of a pressure node and a velocity node of a nonviscous standing wave with a wavelength $\lambda=L$. Due to this condition, the frequency of the gas oscillation occurring in the straight tube is the same as that occurring in the looped tube. In the case of the looped tube, $x=-3L/8$ and $x=5L/8$ represent the same position.

B. Temperature distribution along the axes of the tubes

The calculation method proposed below requires the temperature distribution to be along the tube axis. The mean temperature T_m of the gas in both tubes is assumed to be distributed, as shown in Fig. 2. T_m is maintained at a constant value except from $x=-X_1$ to X_1 (in the stack) and from $x=X_2$ to X_3 (in the thermal buffer tube), where $X_1=L_{\text{stack}}/2$, $X_2=X_1+L_{\text{hx}}$, and $X_3=X_2+L_{\text{tb}}$.

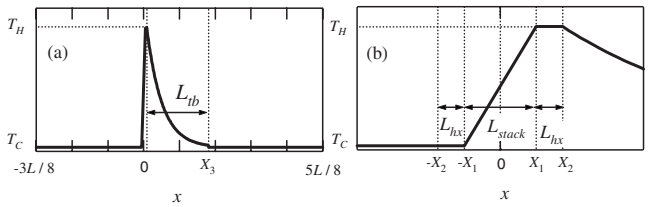


FIG. 2. Temperature distribution (a) along the tube and (b) near the stack.

In the stack (from $x=-X_1$ to X_1), T_m increases linearly from T_C to T_H [see Fig. 2(b)]. The reason for the assumption of linear temperature distribution in the stack is as follows: the stacks used in the experiments performed by several researchers are essentially insulated. Therefore, the temperature distribution in the absence of spontaneous gas oscillation would become linear due to thermal conduction along the axial direction.

In the thermal buffer tube (from $x=X_2$ to X_3), T_m decreases exponentially,

$$T_{m,\text{tb}}(x) = (T_H - T_C)\exp\left(-\frac{x-X_2}{R}\right) + T_C. \tag{1}$$

The coefficient in this equation, R, was varied in the calculation. At $x=X_3$, T_m discontinuously changes, as can be seen in Fig. 2(a). This is because $T_{m,\text{tb}}(X_3)-T_C=(T_H-T_C)\exp(-L_{\text{tb}}/R)>0$. We numerically confirmed that when $L_{\text{tb}}/R>4$, the temperature difference $T_{m,\text{tb}}(X_3)-T_C$ has a negligible influence on the calculated stability limit of the present tubes.

III. CALCULATION METHOD

This section describes the method for calculating the stability limit of the thermally driven spontaneous gas oscillation in the tubes. This calculation is based on Rott's thermoacoustic theory[1] and the transfer matrix method for oscillatory pressure P and volume velocity U. The time variation of the pressure and volume velocity is given by the factor $\exp(i\omega t)$. Generally, ω is represented as a complex quantity; its real and imaginary parts represent the angular frequency of the gas oscillation and the logarithmic amplification, respectively. At the stability limit, the logarithmic amplification becomes zero. Therefore, ω has only a real part.

A. Transfer matrix

In this subsection, the calculation method of the transfer matrix of the components (which are the waveguides, heat exchangers, stack, and thermal buffer tube) is described. Using Rott's acoustic approximation,[1] the momentum and continuity equations in a flow channel are written[18] as

$$\frac{dP}{dx} = -\frac{1}{A}\frac{i\omega\rho_m}{1-\chi_\nu}U, \tag{2}$$

$$\frac{dU}{dx} = -\frac{i\omega A[1+(\gamma-1)\chi_\alpha]}{\gamma P_m}P + \frac{\chi_\alpha-\chi_\nu}{(1-\chi_\nu)(1-\sigma)}\frac{1}{T_m}\frac{dT_m}{dx}U, \tag{3}$$

respectively, where A is the cross-sectional area of the channel, and ρ_m, P_m, γ, and σ are the mean density, the mean pressure, the ratio of specific heats, and the Prandtl number of the working gas, respectively. Here, the assumption that the heat capacity of the channel wall is considerably larger than that of the working gas is used. χ_α and χ_ν are complex functions that allow us to describe the three-dimensional phenomena in the channel using the two one-dimensional equations. For example, in a circular channel with radius r,

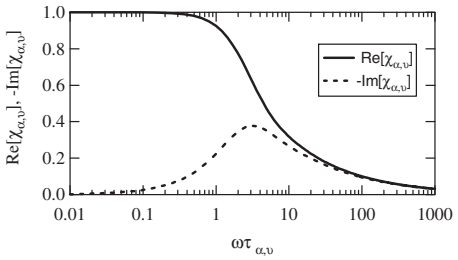


FIG. 3. Complex functions χ_α and χ_ν.

$$\chi_\alpha = \frac{2J_1(Y_\alpha)}{Y_\alpha J_0(Y_\alpha)}, \tag{4}$$

$$\chi_\nu = \frac{2J_1(Y_\nu)}{Y_\nu J_0(Y_\nu)},$$

$$Y_\alpha = (i-1)\sqrt{\omega\tau_\alpha}, \quad Y_\nu = (i-1)\sqrt{\omega\tau_\nu}, \tag{5}$$

where J_1 and J_0 are the first- and zeroth-order Bessel functions, respectively, and τ_α and τ_ν are the thermal and viscous relaxation times, respectively. τ_α is defined as $\tau_\alpha = r^2/(2\alpha)$ and τ_ν as $\tau_\nu = r^2/(2\nu)$, where α and ν are the thermal diffusivity and kinematic viscosity, respectively. $\omega\tau_\alpha$ and $\omega\tau_\nu$ can also be described by r and the thicknesses of the thermal boundary layer, δ_α, and of the viscous boundary layer, δ_ν, as $\omega\tau_\alpha = (r/\delta_\alpha)^2$ and $\omega\tau_\nu = (r/\delta_\nu)^2$, respectively. $\omega\tau_\alpha$ and $\omega\tau_\nu$ are related by $\sigma = \nu/\alpha$ as $\omega\tau_\alpha/\sigma = \omega\tau_\nu$. Figure 3 shows χ_α and χ_ν of the circular channel as a function of $\omega\tau_\alpha$ and $\omega\tau_\nu$, respectively.

Equations (2) and (3) can be modified in a matrix form as follows:

$$\frac{d}{dx}\begin{pmatrix} P(x,t) \\ U(x,t) \end{pmatrix} = C(x)\begin{pmatrix} P(x,t) \\ U(x,t) \end{pmatrix},$$

$$C(x) \equiv \begin{pmatrix} 0 & -\frac{1}{A}\frac{i\omega\rho_m}{1-\chi_\nu} \\ -\frac{i\omega A[1+(\gamma-1)\chi_\alpha]}{\gamma P_m} & \frac{\chi_\alpha - \chi_\nu}{(1-\chi_\nu)(1-N_{\mathrm{Pr}})}\frac{1}{T_m}\frac{dT_m}{dx} \end{pmatrix}. \tag{6}$$

Here, it should be noted that ρ_m, γ, χ_ν, χ_α, and σ depend on T_m, i.e., on x.

When the (2,2) element of C (c_{22}) becomes zero, the matrix C represents the well-known wave equation of a plane sound wave propagating in a channel.[19] Hence, c_{22}, which is the second term on the right-hand side of Eq. (3), plays an important role in the stability of spontaneous gas oscillation.

1. Without temperature gradient

When $dT_m/dx = 0$. Eq. (6) can be solved analytically. This is because γ, χ_ν, χ_α, and σ are independent of x, and the second term on the right-hand side of Eq. (3) vanishes. When the pressure and volume velocity at point x_0 are denoted by P_0 and U_0, respectively, the solution can be expressed as

$$\begin{pmatrix} P(x,t) \\ U(x,t) \end{pmatrix} = M_{\mathrm{I}}(x,x_0)\begin{pmatrix} P_0(x_0,t) \\ U_0(x_0,t) \end{pmatrix},$$

$$M_{\mathrm{I}}(x,x_0) \equiv \begin{pmatrix} \cos[k(x-x_0)] & \frac{-i\omega\rho_m \sin[k(x-x_0)]}{Ak(1-\chi_\nu)} \\ \frac{Ak(1-\chi_\nu)\sin[k(x-x_0)]}{i\omega\rho_m} & \cos[k(x-x_0)] \end{pmatrix}. \tag{7}$$

Here, k is the complex wave number given by

$$k = \frac{\omega}{a}\sqrt{\frac{1+(\gamma-1)\chi_\alpha}{1-\chi_\nu}}, \tag{8}$$

where a is the adiabatic sound speed. Equation (7) shows that when P and U are specified for one point, the distributions of P and U along the tube without the temperature gradient can be calculated.

2. With temperature gradient

When $dT_m/dx \neq 0$, it is difficult to solve Eq. (6) analytically. Hence, it is computationally integrated. By applying a forward difference scheme using the fourth-order Runge–Kutta method to Eq. (6),

$$\begin{pmatrix} P(x+\Delta x,t) \\ U(x+\Delta x,t) \end{pmatrix} = (E + \Delta x C'(x))\begin{pmatrix} P(x,t) \\ U(x,t) \end{pmatrix},$$

$$C'(x) = \tfrac{1}{6}(\mathrm{RK}_A + 2\mathrm{RK}_B + 2\mathrm{RK}_C + \mathrm{RK}_D),$$

$$\mathrm{RK}_A = C(x),$$

$$\mathrm{RK}_B = C(x+\Delta x/2)\left(E + \frac{\Delta x}{2}\mathrm{RK}_A\right),$$

$$\mathrm{RK}_C = C(x+\Delta x/2)\left(E + \frac{\Delta x}{2}\mathrm{RK}_B\right),$$

$$\mathrm{RK}_D = C(x+\Delta x)(E + \Delta x \mathrm{RK}_C) \tag{9}$$

can be obtained, where E is a unit matrix. Hence,

$$\begin{pmatrix} P(x,t) \\ U(x,t) \end{pmatrix} = M_{\mathrm{II}}(x,x_0)\begin{pmatrix} P_0(x_0,t) \\ U_0(x_0,t) \end{pmatrix},$$

$$M_{\mathrm{II}}(x,x_0) \equiv (E + \Delta x C'_{n-1})(E + \Delta x C'_{n-2}) \cdots (E + \Delta x C'_1)(E + \Delta x C'_0) \tag{10}$$

is obtained. Here, n is the number of partitions between x_0 and x, Δx is defined as $(x-x_0)/n$, and C'_j represents C' at $x = x_0 + j\Delta x$. This equation shows that despite the temperature gradient being imposed along the thermal buffer tube/stack, when P and U are provided at one position, the distributions of P and U can be calculated.

B. Method for calculating stability limit

Since T_m is maintained at the constant value from $x=-3L/8$ to $-X_2$, [see Figs. 2(a) and 2(b)], the transfer matrix $M_{\mathrm{I},1}$ of this region was calculated by using M_{I}. Similarly, the transfer matrices for $x=-X_2$ to $-X_1$ (the cold heat exchanger), for $x=X_1$ to X_2 (the hot heat exchanger), and for $x=X_3$ to $5L/8$ are calculated and are denoted by $M_{\mathrm{I,chx}}$, M_{Ihhx}, and $M_{\mathrm{I},2}$, respectively.

Equation (10) was used to compute the transfer matrices $M_{\mathrm{II},s}$ for $x=-X_1$ to X_1 (the stack) and $M_{\mathrm{II,tb}}$ for $x=X_2$ to X_3 (the thermal buffer tube). The partition number $n=100$ was used because it was confirmed that the calculated stability limits with $n=100$, 200, and 400 are almost similar.

Using $M_{\mathrm{I},1}$, $M_{\mathrm{I,chx}}$, $M_{\mathrm{I,hhx}}$, $M_{\mathrm{I},2}$, $M_{\mathrm{II},s}$, and $M_{\mathrm{II},l}$, the transfer matrix of the looped and straight tubes, M_{all}, is written as

$$M_{\mathrm{all}} = M_{\mathrm{I},2} M_{\mathrm{II},l} M_{\mathrm{I,hhx}} M_{\mathrm{II},s} M_{\mathrm{I,chx}} M_{\mathrm{I},1}. \tag{11}$$

Note that in the present study, we did not take account of nonlinear effects, such as the "minor loss,"[20] occurring at the connecting point between the components. By using M_{all}, the oscillatory pressure P_a and volume velocity U_a at $x=-3L/8$ are related to the oscillatory pressure P_b and volume velocity U_b at $x=5L/8$ as

$$M_{\mathrm{all}} \begin{pmatrix} P_a \\ U_a \end{pmatrix} = \begin{pmatrix} P_b \\ U_b \end{pmatrix}. \tag{12}$$

By using this equation, we calculated the stability limit of the thermally induced gas oscillation.

1. For the looped tube

Since $x=-3L/8$ and $x=5L/8$ represent the same position in the looped tube, Eq. (12) can be rewritten as

$$M_{\mathrm{all}} \begin{pmatrix} P_a \\ U_a \end{pmatrix} = \begin{pmatrix} P_a \\ U_a \end{pmatrix}. \tag{13}$$

The solution (P_a, U_a) of Eq. (13) is nonzero if the determinant of the matrix $(M_{\mathrm{all}}-E)$ is zero, i.e., if

$$(m_{11}-1)(m_{22}-1) - m_{12}m_{21} = 0, \tag{14}$$

where E is the unit matrix and m_{ij} is the element of M_{all}. Therefore, by solving Eq. (14), we can achieve the condition of the stability limit of the spontaneous gas oscillation induced in the looped tube.

2. For the straight tube

For the straight tube, Eq. (12) can be modified as

$$M_{\mathrm{all}} \begin{pmatrix} P_a \\ 0 \end{pmatrix} = \begin{pmatrix} P_b \\ 0 \end{pmatrix}. \tag{15}$$

This is because the volume velocity at $x=-3L/8$ and $x=5L/8$ must be zero due to the closed ends. P_a and P_b in Eq. (15) are nonzero if m_{21} is zero, i.e., if

$$m_{21} = 0. \tag{16}$$

Therefore, Eq. (16) determines the condition of the stability limit in the straight tube.

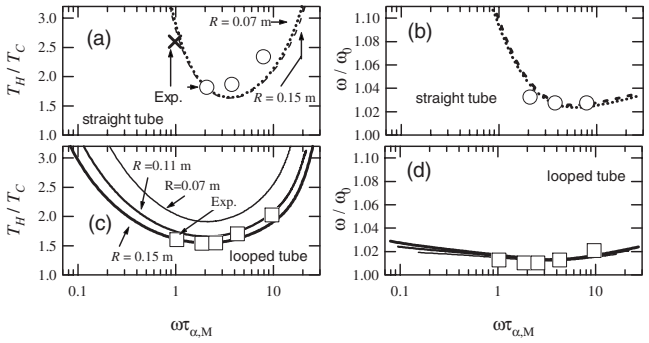


FIG. 4. The conditions of the stability limit of [(a) and (b)] the straight tube and [(c) and (d)] the looped tube. The temperature ratio T_H/T_C and frequency ratio ω/ω_0 of the stability limit are plotted as a function of $\omega\tau_{\alpha,M}$.

IV. CALCULATION RESULTS

A. Calculation condition

In the calculation described below, we chose r_{stack} and T_H as the variable parameters and defined $L=2.8$ m, $L_{\mathrm{stack}}=35$ mm, $L_{\mathrm{hx}}=13$ mm, $L_{\mathrm{tb}}=0.60$ m, $r_{\mathrm{tube}}=20$ mm, $r_{\mathrm{hx}}=0.5$ mm, $\epsilon_{\mathrm{hx}}=0.67$, and $T_C=295$ K. These values are the same as those in the experiment described later in Sec. IV B. Although the porosity $\epsilon_{\mathrm{stack}}$ of the stacks used for the experiment were between 0.69 and 0.87, $\epsilon_{\mathrm{stack}}$ in the calculation was assumed to be the same as ϵ_{hx}, i.e., $\epsilon_{\mathrm{stack}}=0.67$. The coefficient R in Eq. (1) was varied from 0.07 to 0.15 m. The reason why the values of R were selected to be in this range is as follows. We experimentally measured the temperature distribution along the thermal buffer tube: when the gas oscillation did not take place, the measured R was 0.07 m; on the other hand, when the gas oscillation occurred, R in the case of the looped tube had values between 0.07 and 0.15 m, whereas R for the straight tube was approximately 0.07 m.

In order to calculate the values of ν, α, ρ_m, γ, and c_p, we used polyfunctions fitting their temperature dependences.[21] The number of polynomial terms of the polyfunctions was 6.

B. Stability limit

The left-hand sides of Eqs. (14) and (16) were computed as a function of real ω with various T_H and r_{stack}; further, the combinations of ω/ω_0, T_H/T_C, and $\omega\tau_{\alpha,M}$ satisfying Eqs. (14) and (16) were determined. ω/ω_0, T_H/T_C, and $\omega\tau_{\alpha,M}$ are dimensionless parameters corresponding to ω, T_H, and r_{stack}, respectively. Here, ω_0 is defined as $\omega_0=2\pi a_0/L$, where $a_0=340$ m/s, and $\omega\tau_{\alpha,M}$ is defined using thermal diffusivity at the center of the stack $x=0$, where the temperature $T_M=(T_H+T_C)/2$. Although the combinations thus obtained revealed that spontaneous gas oscillation occurs in several frequency modes, we focus on the calculation results of a mode at $\omega/\omega_0 \sim 1$.

In Fig. 4, the calculated conditions of the stability limit are denoted by lines; the calculated T_H/T_C and ω/ω_0 are shown as a function of $\omega\tau_{\alpha,M}$. First, we see the calculated T_H/T_C of the straight tube. The calculated results for the straight tube with $R=0.07$ m and that with $R=0.15$ m are shown in Fig. 4(a) by dotted and dashed lines, respectively. However, it is difficult to distinguish the dashed line from the dotted line. This indicates that T_H/T_C of the straight tube is

almost independent of the value of R. The value of T_H/T_C takes a minimum value at $\omega\tau_{\alpha,M}\sim 3$, and the T_H/T_C curve is U shaped. Figure 4(b) shows that ω/ω_0 is also independent of the value of R and increases rapidly with a decrease in $\omega\tau_{\alpha,M}$ from $\omega\tau_{\alpha,M}\sim 3$. These results are identical to the characteristics of the curves obtained by Rott and Zouzoulas[3] and Yazaki *et al.*[4]

Next, we focus on the calculation results of the looped tube, which are shown in Figs. 4(c) and 4(d). Figure 4(c) shows that the value of T_H/T_C of the looped tube depends on the value of R. However, it was found that independent of the value of R, T_H/T_C at a given R takes a minimum value near $\omega\tau_{\alpha,M}=2$. Moreover, it was shown that the shapes of the three T_H/T_C curves of the looped tube are very similar.

The calculated $\omega\tau_{\alpha,M}$ dependences of T_H/T_C of the looped tube are similar to those of the straight tube for $\omega\tau_{\alpha,M}>3$. However, as can be seen from Figs. 4(a) and 4(c), for $\omega\tau_{\alpha,M}<3$, the $\omega\tau_{\alpha,M}$ dependences of T_H/T_C of the looped tube are different from those of the straight tube; as $\omega\tau_{\alpha,M}$ decreases, T_H/T_C of the looped tube increases gradually compared with that of the straight tube. As a result, when $\omega\tau_{\alpha}\ll 3$, T_H/T_C at the stability limit of the looped tube would be considerably lower than that of the straight tube, regardless of the value of R. Figures 4(b) and 4(d) show that the curves of ω/ω_0 of the looped tube also differ from those of the straight tube for $\omega\tau_{\alpha,M}<3$.

C. Comparison of calculation with experiment

In order to compare the calculated and experimental values of T_H/T_C and ω/ω_0, we performed the measurements of T_H/T_C and ω/ω_0. We constructed the looped and straight tubes shown in Figs. 1(a) and 1(b). The looped tube consisted of stainless-steel tubes and four 90° elbows, whereas the straight tube consisted of stainless-steel tubes. Ceramic honeycombs comprising square channels were used as stacks in the experiments, whereas the stacks used in the calculations were assumed to have circular channels. Half of the hydraulic diameter[22] of the stacks was used as r_{stack} for the experiments. We considered that the difference in the geometry of the channels in the experiments and calculations had a negligible effect on the stability limit because the relation between $\chi_{\alpha,\nu}$ and $\omega\tau_{\alpha,\nu}$ of a square channel is very similar to that of a circular channel.[18,23] The hot and cold heat exchangers used in the experiments comprised flat brass plates set in parallel with a spacing. The thickness of the flat brass plates was 0.5 mm. The dimensions of the constructed tubes (the total tube length L, the stack length L_{stack}, the length of the heat exchangers, L_{hx}, the radius of the tubes, r_{tube}, and the spacing in the heat exchangers, $2r_{\mathrm{hx}}$,) were the same as those in the calculation.

The hot heat exchanger was heated by an electric heater wound around it. The cold heat exchanger was cooled with chilled water (≈295 K) passing through the tube wound around it. It was experimentally confirmed that the temperature difference across the cross section of the hot heat exchanger was up to 20 K, whereas the temperature difference across the cross section of the cold heat exchanger was up to 5 K. We defined the averaged temperatures across the cross section of the hot and cold heat exchangers as T_H and T_C, respectively.

We measured the pressure by using a pressure sensor mounted on the tube wall near the stack with increasing T_H and determined the condition of the stability limit experimentally. In Figs. 4(a) and 4(c), the measured T_H/T_C at the stability limit in the straight tube and that in the looped tube are plotted with open circles and open squares, respectively. When stacks with $r_{\mathrm{stack}}=0.30$ and 0.40 mm were used, spontaneous gas oscillation was not observed in the straight tube for $T_H/T_C<2.6$, which is the experimental upper limit of T_H/T_C. When $T_H/T_C=2.6$ and $\omega/\omega_0=1.05\pm 0.05$ are assumed, $\omega\tau_{\alpha,M}$ of the stack with $r_{\mathrm{stack}}=0.40$ mm is evaluated to be 1.04 ± 0.05. These results prove that the gas in the straight tube is stable in the region ($\omega\tau_{\alpha,M}<1.04$, $T_H/T_C<2.6$). In order to show this result, we plotted the multiplication sign × at $(\omega\tau_{\alpha,M},T_H/T_C)=(1.04,2.6)$ in Fig. 4(a).

As observed in Fig. 4(a), the experimentally obtained T_H/T_C of the straight tube quantitatively agrees with the calculated T_H/T_C. Figure 4(c) shows that the measured T_H/T_C of the looped tube takes the minimum value at $\omega\tau_{\alpha,M}\sim 2$ and gradually increases compared with that of the straight tube below $\omega\tau_{\alpha,M}=2$. This result is the same as the calculated results. Therefore, we consider that the measured T_H/T_C of the looped tube is also in qualitative agreement with the calculated T_H/T_C.

The symbol in Figs. 4(b) and 4(d) show the measured ω/ω_0 at the stability limits. These figures also show that for both tube types, the measured ω/ω_0's agree well with the calculated ratios. Based on the agreement of T_H/T_C and ω/ω_0, we conclude that the proposed calculation method succeeds in modeling the stability limit of the thermally induced spontaneous gas oscillation in looped and straight tubes.

V. ANALYSIS OF THE STABILITY LIMIT

A. Temperature ratio T_H/T_C at the stability limit

As mentioned above, the calculated T_H/T_C's of both looped and straight tubes take minimum values, and the T_H/T_C curves are U shaped. In this subsection, the reasons for the existence of the stability limit are determined.

The $\omega\tau_{\alpha,M}$ dependency of T_H/T_C for $\omega\tau_{\alpha,M}>3$ can be understood from the $\omega\tau_{\alpha,\nu}$ dependency of $\chi_{\alpha,\nu}$. As shown in Fig. 3, $\mathrm{Re}[\chi_{\alpha,\nu}]$ and $\mathrm{Im}[\chi_{\alpha,\nu}]$ approach zero for $\omega\tau_{\alpha,\nu}\gg 3$. This reduces the value of the second term on the right-hand side of Eq. (3); when $\chi_{\alpha,\nu}$ is zero, the second term is also zero. Since the second term is the origin of the driving force for spontaneous gas oscillation, T_H/T_C required for the excitation of the oscillation increases when $\omega\tau_{\alpha,M}$ increases from 3.

Understanding the $\omega\tau_{\alpha,M}$ dependency of T_H/T_C for $\omega\tau_{\alpha,M}<3$ is more difficult than understanding that for $\omega\tau_{\alpha,M}>3$. This is because P and U are complex and when $\omega\tau_{\alpha,\nu}$ decreases below 3, $\mathrm{Re}[\chi_{\alpha,\nu}]$ increases whereas $\mathrm{Im}[\chi_{\alpha,\nu}]$ decreases. In order to understand the $\omega\tau_{\alpha,M}$ dependency of T_H/T_C for $\omega\tau_{\alpha,M}<3$, we consider the physical meaning of $\omega\tau_{\alpha,\nu}$. A decrease in $\omega\tau_{\nu}$ simply increases the

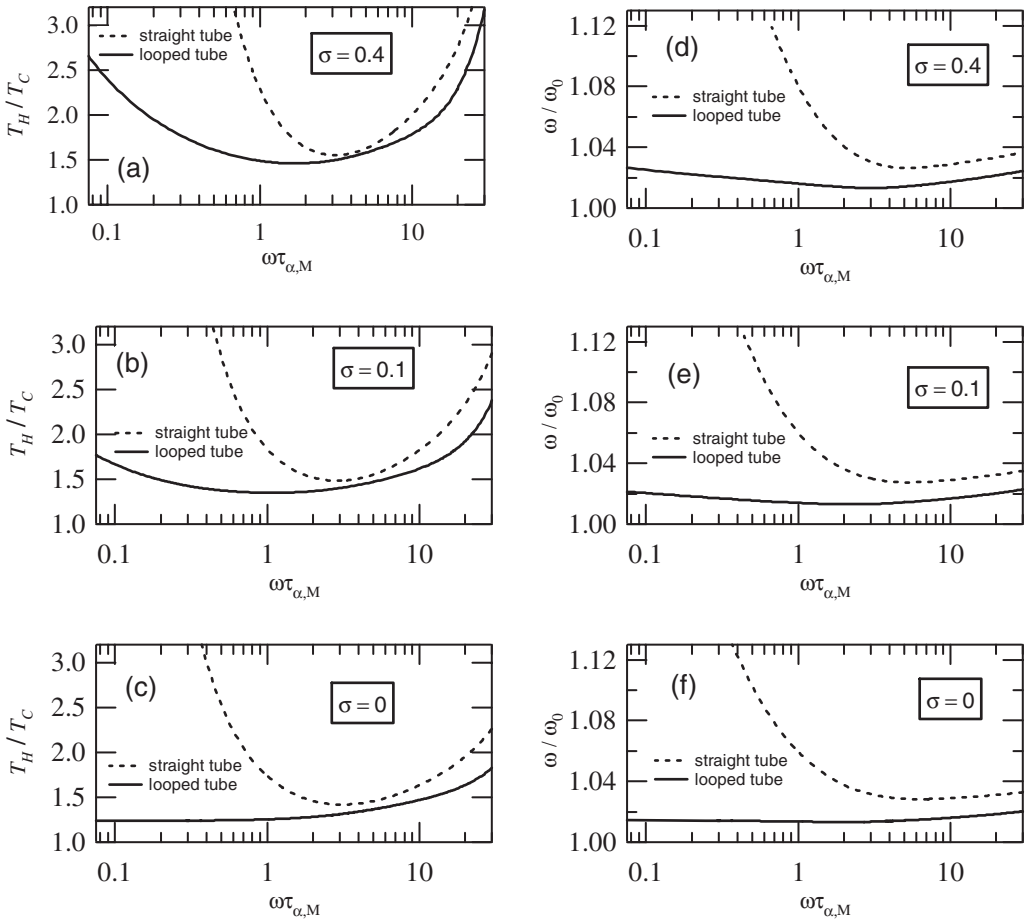


FIG. 5. Stability curves in the looped tube (solid line) and straight tube (dashed line). The stability curves are calculated with a variety of the value of the Prandtl number σ in the stack.

effect of the viscosity. On the other hand, a decrease in $\omega\tau_\alpha$ modifies the thermal process of heat exchange between the gas and channel wall due to the effect of the thermal conductivity of the gas. When $\omega\tau_\alpha \ll 3$, the thermal relaxation time τ_α is much shorter than the cyclic period $2\pi/\omega$; therefore, heat exchange occurs instantaneously, i.e., isothermally (reversibly). When $\omega\tau_\alpha \sim 3$, heat exchange occurs irreversibly because $\tau_\alpha \sim 2\pi/\omega$. Based on the discussion regarding entropy oscillation along the radial direction of a flow channel, Tominaga stated that irreversibility and isothermal reversibility of the heat exchange process are expressed by the values of $\mathrm{Im}[\chi_\alpha]$ and $\mathrm{Re}[\chi_\alpha]$, respectively.[24]

In order to separate the effects of viscosity and thermal conductivity on the stability limit, we calculated the stability limits by setting the value of the Prandtl number σ to 0.71, 0.40, 0.10, and 0. It should be noted that because σ was changed only in the stack, the viscosity outside the stack did not change. The calculated T_H/T_C is shown in Figs. 5(a)–5(c). Since the shape of the stability curves is independent of the values of R, as mentioned in Sec. IV B, only the calculated results with $R=0.15$ m are shown. Since σ for air is nearly 0.71, the curves of the calculated T_H/T_C with $\sigma=0.71$ are almost the same as those shown in Fig. 4(a) by the solid line. Hence, the calculation results with $\sigma=0.71$ are not shown.

In the case investigated by Rott[2] and Yazaki *et al.*,[4] it was claimed that when $\omega\tau_{\alpha,M} \ll 3$, the existence of the stability limit in the tube having two ends can be attributed to the viscosity. However, as shown by the dashed lines in Figs. 4(a) and 5(a)–5(c), T_H/T_C of the straight tube increases with $\omega\tau_{\alpha,M} < 3$, even though σ decreases. This implies that in the present case, the stability limit of the T_H/T_C curve of the straight tube is not necessarily attributable to the viscosity. T_H/T_C of the straight tube takes a minimum value near $\omega\tau_{\alpha,M}=3$ independently of the value of σ. This fact indicates that for the excitation of the gas oscillation in the straight tube, the value of $\omega\tau_\alpha=3$ is important. As shown in Fig. 3, when $\omega\tau_\alpha \sim 3$, $-\mathrm{Im}[\chi_\alpha]$ attains the maximum value and when $\omega\tau \ll 3$, it becomes zero. As mentioned above, $\mathrm{Im}[\chi_\alpha]$ represents the irreversibility of the heat exchange process. Therefore, we can say that the irreversibility of the heat exchange process is indispensable for the excitation of the spontaneous gas oscillation in the straight tube and that when $\omega\tau_{\alpha,M}$ decreases from 3, T_H/T_C of the straight tube increases due to the decrease in the irreversibility.

The solid lines in Figs. 4(c) and 5(a)–5(c) indicate that in the case of the looped tube, the optimum value of $\omega\tau_{\alpha,M}$ for realizing minimum T_H/T_C decreases with a decrease in σ; further, T_H/T_C monotonically decreases at $\sigma=0$. This observation reveals two facts: (1) T_H/T_C increase when $\omega\tau_{\alpha,M} < 3$ in the case of the looped tube, which is shown in Fig. 4(c) by lines, is caused by the viscosity in the stack and (2)

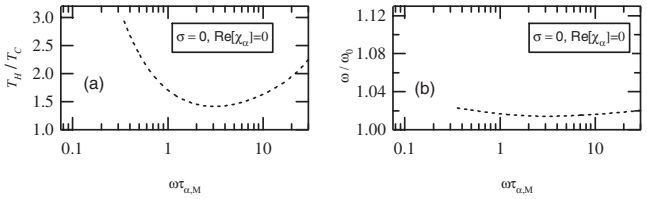


FIG. 6. Stability curves of the straight tube, which is calculated with the assumption that the real part of χ_α, i.e., Re$[\chi_\alpha]$, and σ are both zero.

the isothermal reversibility in the heat exchange process, which is represented by Re$[\chi_\alpha]$, is important for the excitation of the spontaneous gas oscillation in the looped tube. The second fact is considered to contribute to achieving a higher efficiency in the energy conversion occurring in a looped tube as compared with that in a straight tube.

B. Angular frequency ratio ω/ω_0 at the stability limit

The calculated ω/ω_0 with various values of σ is shown in Figs. 5(d)–5(f). When $\sigma=0.71$, ω/ω_0 is not shown due to the above-mentioned reason for not showing T_H/T_C when $\sigma=0.71$. The solid lines in Figs. 4(c), 4(d), and 5(a)–5(g) indicate that the $\omega\tau_{\alpha,M}$ dependence of ω/ω_0 of the looped tube is closely related to T_H/T_C at the stability limit; as T_H/T_C decreases, ω/ω_0 is reduced. This can be considered as proof that ω/ω_0 is determined by the value of the mean sound speed a_m in the tubes because a_m is a function of T_H under a given value of R. In contrast, the $\omega\tau_{\alpha,M}$ dependence of ω/ω_0 of the straight tube cannot be explained by using only a_m. As shown in Figs. 4(a) and 4(b), the T_H/T_C value of the straight tube at $\omega\tau_{\alpha,M}=10$ and 1.5 is 2.1; however, ω/ω_0 at $\omega\tau_{\alpha,M}=10$ and 1.5 is 1.03 and 1.06, respectively. Moreover, ω/ω_0 calculated with various values of σ indicates that the $\omega\tau_{\alpha,M}$ dependence of ω/ω_0 of the straight tube weakly depends on σ, i.e., the value of the viscosity.

The $\omega\tau_{\alpha,M}$ dependence of ω/ω_0 of the straight tube can also be explained by the $\omega\tau_\alpha$ dependence of χ_α. This is because when $\sigma=0$, only χ_α depends on $\omega\tau_\alpha$. In order to split the effects of the imaginary and real parts of χ_α, we set Re$[\chi_\alpha]$ to zero, keeping $\sigma=0$, and calculate the stability limit of the straight tube. The stability limit thus calculated is shown in Fig. 6.

As mentioned above, the T_H/T_C curve of the straight tube for $\omega\tau_{\alpha,M}<3$ exists due to the decrease in $-$Im$[\chi_\alpha]$. Hence, the T_H/T_C curve shown in Fig. 6(a) is U shaped and is very similar to the T_H/T_C curve shown by the dashed line in Fig. 5(c), even though Re$[\chi_\alpha]=0$. However, as observed in Fig. 6(b), the rapid increase in ω/ω_0 with a decrease in $\omega\tau_{\alpha,M}$, which is observed below $\omega\tau_{\alpha,M}<3$ in Fig. 5(f), vanishes; due to this, ω/ω_0 at Re$[\chi_\alpha]=0$ and $\sigma=0$ depends on T_H/T_C. These facts imply that the rapid increase in ω/ω_0 is caused by the $\omega\tau_\alpha$ dependence of Re$[\chi_\alpha]$, while the T_H/T_C curve of the straight tube is determined by the $\omega\tau_\alpha$ dependence of Im$[\chi_\alpha]$.

VI. SUMMARY

The stability limits of the spontaneous gas oscillation thermally induced in the looped and straight tubes were numerically investigated. These numerical results were found to be in qualitative agreement with the measured results. The stability limits were calculated with various values of the Prandtl number of the gas in order to determine the causes of the existence of the stability limit. As a result, it was found that the temperature ratio required for the gas oscillation in the looped tube is determined by the viscosity and isothermal reversibility of the thermal process of heat exchange occurring between the gas and tube wall; however, the temperature ratio required in the straight tube is determined by the irreversibility of the thermal process. On the other hand, the frequency of the gas oscillation occurring in the straight tube depends on the reversibility, whereas the sound speed mainly contributes to the frequency of the gas oscillation occurring in the looped tube.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture in Japan under the Grant-in-Aid for Scientific Research (Grant No. 17-10613, 2006) and the Grant-in-Aid for Division of Young Researchers.

[1]N. Rott, "Damped and thermally driven acoustic oscillations," Z. Angew. Math. Phys. **20**, 230–243 (1969).
[2]N. Rott, "Thermally driven acoustic oscillations. Part 2: Stability limit for helium," Z. Angew. Math. Phys. **24**, 54–72 (1973).
[3]N. Rott and G. Zouzoulas, "Thermally driven acoustic oscillations," Z. Angew. Math. Phys. **27**, 197–224 (1976).
[4]T. Yazaki, A. Tominaga, and Y. Narahara, "Experiments on thermally driven acoustic oscillations of gaseous helium," J. Low Temp. Phys. **41**, 45–60 (1980).
[5]W. Arnott, J. Belcher, R. Raspet, and H. Bass, "Stability analysis of a helium-filled thermoacoustic engine," J. Acoust. Soc. Am. **96**, 370–375 (1994).
[6]A. Atchley and F. Kuo, "Stability curves for a thermoacoustic prime mover," J. Acoust. Soc. Am. **95**, 1401–1404 (1994).
[7]N. Sugimoto, "Thermoacoustic oscillations and their stability analysis from a viewpoint of the boundary-layer theory (in Japanese)," Nagare **24**, 381–393 (2005).
[8]N. Sugimoto and M. Yashida, "Marginal condition for the onset of thermoacoustic oscillations of a gas in a tube," Phys. Fluids **19**, 074101 (2007).
[9]T. Yazaki, S. Takashima, and F. Mizutani, "Complex quasiperiodic and chaotic states oeserved in thermally induced oscillations of gas columns," Phys. Rev. Lett. **58**, 1108–1111 (1987).
[10]S. Backhaus, E. Tward, and M. Petach, "Traveling-wave thermoacoustic electric generator," Appl. Phys. Lett. **85**, 1085–1087 (2004).
[11]E. Luo, W. Dai, Y. Zhang, and H. Ling, "Thermoacoustically driven refrigerator with double thermoacoustic-stirling cycles," Appl. Phys. Lett. **88**, 074102 (2007).
[12]T. Yazaki, T. Biwa, and A. Tominaga, "A pistonless stirling cooler," Appl. Phys. Lett. **80**, 157–159 (2002).
[13]P. Ceperley, "A piston-less stirling engine," J. Acoust. Soc. Am. **65**, 1508–1513 (1979).
[14]S. Backhaus and G. W. Swift, "A thermoacoustic stirling engine," Nature (London) **399**, 335–338 (1999).
[15]T. Yazaki, A. Iwata, T. Mackawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128–3131 (1998).
[16]H. Tanaka, "Investigation of thermoacoustic phenomena," MS thesis, The University of Tokyo, Tokyo, 2002.
[17]G. Penelet, S. Job, V. Gusev, P. Lotton, and M. Bruneaua, "Dependence of sound amplification on temperature distribution in annular thermoacoustic engines," Acust. Acta Acust. **91**, 567–577 (2005).
[18]G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Pennsylvania, 2002).
[19]T. Yazaki, Y. Tashiro, and T. Biwa, "Measurements of sound propagation

in narrow tubes," Proc. R. Soc. London, Ser. A **463**, 2855–2862 (2007).
[20]R. S. Wakeland and R. M. Keolian, "Measurements of the resistance of parallel-plate heat exchangers to oscillating flow at high amplitudes," J. Acoust. Soc. Am. **115**, 2071–2074 (2004).
[21]*JASM Data Book:Thermophysical Properties of Fluids* (the Japan Society of Mechanical Engineers, Tokyo, 1983).
[22]The hydraulic diameter d_h is defined as $d_h=4V_{\mathrm{gas}}/A_{\mathrm{gas\text{-}solid}}$, where V_{gas} is the gas volume and $A_{\mathrm{gas\text{-}solid}}$ is the gas-solid contact surface area.
[23]W. Arnott, H. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped poro cross sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991).
[24]A. Tominaga, "Therммodynamic aspect of thermoacoustic phenomena," Cryogenics **35**, 727–440 (1995).

Modeling of wave dispersion along cylindrical structures using the spectral method

Florian Karpfinger[a] and Boris Gurevich[b]
Department of Exploration Geophysics, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia

Andrey Bakulin[c]
WesternGeco, 10001 Richmond Ave., Houston, Texas 77042

(Received 20 November 2007; accepted 9 May 2008)

Algorithm and code are presented that solve dispersion equations for cylindrically layered media consisting of an arbitrary number of elastic and fluid layers. The algorithm is based on the spectral method which discretizes the underlying wave equations with the help of spectral differentiation matrices and solves the corresponding equations as a generalized eigenvalue problem. For a given frequency the eigenvalues correspond to the wave numbers of different modes. The advantage of this technique is that it is easy to implement, especially for cases where traditional root-finding methods are strongly limited or hard to realize, i.e., for attenuative, anisotropic, and poroelastic media. The application of the new approach is illustrated using models of an elastic cylinder and a fluid-filled tube. The dispersion curves so produced are in good agreement with analytical results, which confirms the accuracy of the method. Particle displacement profiles of the fundamental mode in a free solid cylinder are computed for a range of frequencies.
 [DOI: 10.1121/1.2940577]

PACS number(s): 43.40.At, 43.58.Ta, 43.35.Cg, 43.20.Hq [LLT] Pages: 859–865

I. INTRODUCTION

Modeling different wave modes propagating along a cylindrical borehole is important for the understanding and quantitative interpretation of borehole sonic and seismic measurements. Numerous different modes such as head waves, trapped modes, and surface waves can be observed in these structures. All these modes are frequency dependent. In sonic-logging recordings these modes overlap and are often hard to identify. In order to separate different borehole modes it is useful to analyze their dispersive characteristics.

Traditionally, mode dispersion was studied by finding roots of analytical dispersion equations. The method has a long history. By the end of the 19th century Pochhammer[1] and Chree[2] investigated the wave propagation in free elastic rods. These solutions are presented in detail by Love (Ref. 3, Sec. 201) and Kolsky (Ref. 4, Chap. 3). Numerical solutions to the Pochhammer–Chree equation are presented, for example, by Bancroft.[5]

Another case which was investigated by different authors is that of a hollow[6–8] and fluid-filled[9–11] cylindrical shell.

The root-finding method, however, becomes difficult to implement when the number of cylindrical layers and/or modes of interest becomes large.[12] The separation of different roots in the complex plane becomes even more challenging when inelastic effects need to be taken into account, such as in the case of a cylinder filled with a viscoelastic fluid[13–15] or poroelastic structures.[16–18]

An alternative approach to model two-dimensional circular structures was recently introduced by Adamou and Craster[19] based on spectral methods. The idea of this method is to solve the underlying differential equations by numerical interpolation using orthogonal polynomials and spectral differentiation matrices (DMs). The advantage of this approach is that it is much faster and easier to implement than conventional root-finding methods, especially for attenuating, poroelastic, or anisotropic structures.

In this paper we extend the concept of the spectral method for wave propagation along circular cylindrical structures, and compare the results with known analytical solutions. In Sec. II, the underlying equations in cylindrical coordinates and the eigenvalue problem are formulated for a free solid cylinder. In Sec. III, the solution of the eigenvalue problem for an elastic cylinder is described using the spectral method. Numerical results are presented in the form of dispersion curves. In Sec. IV the approach is extended to multiple layers. The dispersion curves are displayed for the case of a fluid filled tube. In Sec. V, displacement profiles are computed for various frequencies of the fundamental mode propagating in the free solid cylinder.

II. THEORY

A. Equations of motion

We first introduce the spectral method for the simplest case of axisymmetric wave propagation along a free solid bar. The dynamics of the cylinder is analyzed in cylindrical

[a]Electronic mail: florian.karpfinger@postgrad.curtin.edu.au
[b]Electronic mail: boris.gurevich@geophy.curtin.edu.au. Also at CSIRO Petroleum, Bentley, Western Australia
[c]Electronic mail: abakulin@slb.com

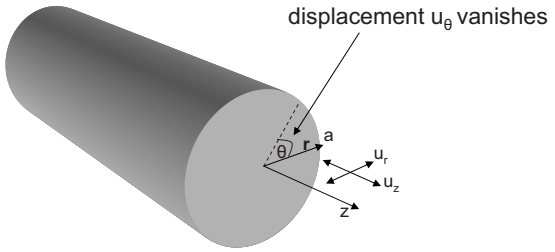


FIG. 1. Geometry of a free solid bar, displaying the coordinate system which reduces to (r,z) and the displacement field (u_r,u_z) for axisymmetric wave propagation.

coordinates (r,θ,z) (Fig. 1). For axisymmetric motion the transverse component u_θ of the displacement field $\mathbf{u}=(u_r,u_\theta,u_z)$ is identically zero, while its radial and axial components u_r and u_z are independent of θ.

For the analysis of the wave propagation it is convenient to introduce displacement potentials

$$u_r = \partial_r\phi - \partial_z\psi_\theta, \tag{1}$$

$$u_z = \partial_z\phi + r^{-1}\partial_r(r\psi_\theta), \tag{2}$$

where ϕ is the scalar potential and ψ_θ is the transverse component of the vector potential $\boldsymbol{\psi}$, ∂_x is a shortcut for the partial derivative $\partial/\partial x$. For axisymmetric motion ψ_θ is the only nonzero component of $\boldsymbol{\psi}$:

$$\psi = (0,\psi_\theta,0)^T, \tag{3}$$

where ψ_θ can in turn be written as

$$\psi_\theta = -\partial_r\eta, \tag{4}$$

so that

$$\boldsymbol{\psi} = \nabla\times(\eta\mathbf{e}_z), \tag{5}$$

where $\mathbf{e}_z$ is the unit vector in z direction.

The equations of axisymmetric motion can be written in the form (see Ref. 20, Sec. 2.13)

$$\nabla^2\phi = \frac{1}{v_p^2}\partial_t^2\phi, \tag{6}$$

$$\left(\nabla^2 - \frac{1}{r^2}\right)\psi_\theta = \frac{1}{v_p^2}\partial_t^2\psi_\theta, \tag{7}$$

where v_p is the P-wave velocity, v_s is S-wave velocity, t is time, and ∇^2 is Laplace operator,

$$\nabla^2 = \partial_r^2 + r^{-1}\partial_r + \partial_z^2. \tag{8}$$

The motion of the cylinder can be found from the solution of Eqs. (6) and (7) subject to the boundary conditions on the displacements and stress tractions on the free surface of the cylinder. The displacements are given by Eqs. (1) and (2). The normal and tangential stress tractions are related to displacements using the Hooke's law,

$$\sigma_{rr} = \lambda\Delta + 2\mu\partial_r u_r, \tag{9}$$

and

$$\sigma_{rz} = \mu(\partial_z u_r + \partial_r u_z), \tag{10}$$

where $\Delta = \partial_r u_r + \partial_r r^{-1} + \partial_z u_z$ denotes the dilatation in cylindrical $r-z$ coordinates, λ and μ are the Lame parameters.

We consider the propagation of an infinite train of sinusoidal waves along the z axis of the cylinder, which is a harmonic function of z and t of the form

$$\phi = \Phi e^{i(k_z z-\omega t)}, \quad \psi_\theta = \Psi e^{i(k_z z-\omega t)}, \tag{11}$$

where ω is the angular frequency, k_z the axial wave number, and U and W are the amplitudes which are functions of r and θ. From Eq. (11) it follows that $\partial_t\phi = -\omega\phi$ and $\partial_z\phi = ik_z\phi$, etc.

B. Helmholtz equations

The two wave equations [Eqs. (6) and (7)] transformed into the $\omega-k_z$ domain by introducing Eq. (11) and dropping $e^{i(k_z z-\omega t)}$ become

$$\underbrace{\left(\partial_r^2 + r^{-1}\partial_r + \frac{\omega^2}{v_p^2}\right)}_{\mathfrak{L}_{v_p}}\Phi = k_z^2\Phi, \tag{12}$$

$$\underbrace{\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \frac{1}{r^2} + \frac{\omega^2}{v_s^2}\right)}_{\mathfrak{L}_{v_s}}\Psi = k_z^2\Psi. \tag{13}$$

Equations (12) and (13) are now ordinary differential equations containing derivatives with respect to r only and coefficients depending on frequency ω and axial wave number k_z. The aim is to find a relation between ω and k_z. This means finding a k_z for a given ω or vice versa. This can be done by solving Eqs. (12) and (13) as an eigenvalue problem so that the wave number k_z^2 represents the eigenvalue and the potentials $\Phi(r)$ and $\Psi(r)$ are the eigenvectors. Alternatively, we could rearrange Eqs. (12) and (13) so that the terms with k_z appear on the left-hand side and with ω on the right-hand side, which will give an eigenvalue problem for ω^2. For linear elasticity both approaches must give identical results.[19] However for more complicated media (say, viscoelastic or poroelastic) it is advantageous to look for k_z as a function of ω as coefficients of governing equations may themselves explicitly depend on ω.

C. Boundary conditions

The solution of Eqs. (12) and (13) should be solved subject to boundary conditions on the surface of the cylinder. In order to apply the boundary conditions, the displacements and stress components have to be expressed independent of the axial wave number k_z.

The radial and axial displacement components u_r and u_z can be expressed by substituting Eq. (11) into Eqs. (1) and (2),

$$u_r = \partial_r\Phi - \hat{\Psi}, \tag{14}$$

$$\hat{u}_z = \underbrace{-k_z^2\Phi}_{\mathfrak{L}_{v_p}\Phi} + (\partial_r + r^{-1})\hat{\Psi}, \tag{15}$$

where $\hat{\Psi} = ik_z\Psi$ and $\hat{u}_z = ik_z u_z$.

These expressions are used to make the stress components σ_{rr} and σ_{rz} [Eqs. (9) and (10)] solely dependent on the potentials Φ and $\hat{\Psi}$. This yields after some manipulations

$$\sigma_{rr} = \left[-\lambda\left(r^{-2} + \frac{\omega^2}{v_p^2}\right) + 2\mu\partial_r^2\right]\Phi + 2\mu\partial_r\hat{\Psi}, \tag{16}$$

$$\hat{\sigma}_{rz} = -2\mu\left(\partial_r^3 + r^{-1}\partial_r^2 - r^{-2}\partial_r + \frac{\omega^2}{v_p^2}\partial_r\right)\Phi + \mu\left(2\partial_r^2 + 2r^{-1}\partial_r - 2r^{-2} + \frac{\omega^2}{v_s^2}\right)\hat{\Psi}, \tag{17}$$

where $\hat{\sigma}_{rz} = ik_z\sigma_{rz}$. Equations (12)–(17) fully describe the problem of any vibrating cylindrical structures in the r–z plane.

The classical way to solve such problems would be the so-called *root-finding* approach. A general solution to Eqs. (12) and (13) is found, which is a combination of Bessel functions of different order. Substituting the solution into the boundary conditions yields a homogeneous system of linear algebraic equations. In order for this system to have nontrivial solutions, the determinant of its matrix M must be equal to zero, $\det M(\omega, k_z) = 0$. This is called the frequency equation. The roots of this equation yield the dispersion relation $\omega(k_z)$. Since wave solutions in cylindrical coordinates contain various Bessel functions, it is often quite difficult to isolate and determine the various roots. Solving the frequency equation gets even more complicated in the case of leaky modes or lossy structures where solutions of the dispersion relation should be found in the complex plane.

In Sec. III an alternative approach, based on the spectral method, is presented.

III. SPECTRAL METHOD FOR AN ELASTIC CYLINDER

The spectral method bypasses the difficulties and solves the underlying Helmholtz equations numerically. For elastic wave propagation this was first implemented by Adamou and Craster,[19] who investigated circumferential waves in an elastic annulus (motion independent of r and z, see Fig. 1). In this study we extend the spectral method to axisymmetric longitudinal modes.

Subsequently the method is straightforwardly extended to the case of arbitrary n-layered fluid–solid media. The eigenvectors correspond to the potentials Φ and Ψ which are used to compute the mode shapes.

A. Polynomial interpolation

In order to solve the eigenvalue problem (12) and (13) numerically represent functions $\Phi(r)$ and $\Psi(r)$ by Chebychev polynomials. The advantage of this approach is that the derivatives of these polynomials can be computed exactly using so-called differentiation matrices. Consider a function $f(r)$ evaluated at N interpolation points, which is represented by a vector $\mathbf{f}$ of length N. This interpolated function $\mathbf{f}^{(m)}$ is connected to its mth derivative $\mathbf{f}$ through

$$\begin{pmatrix} f_1^{(m)} \\ f_2^{(m)} \\ \vdots \\ f_N^{(m)} \end{pmatrix} \approx \underbrace{\begin{pmatrix} D_{11}^{(m)} & D_{12}^{(m)} & \cdots & D_{1N}^{(m)} \\ D_{21}^{(m)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ D_{N1}^{(m)} & \cdots & \cdots & D_{NN}^{(m)} \end{pmatrix}}_{D^{(m)}} \cdot \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}. \tag{18}$$

This is, the mth derivative of $\mathbf{f}$ can be calculated by multiplying $\mathbf{f}$ with the $N \times N$ matrix $D^{(m)}$, which represents the DM. The N interpolation points, which are, in our case, evaluated along the radius r of the cylinder, are the N maxima of the Chebyshev polynomial of the Nth order. The Chebyshev DMs are calculated using the recursive formula for the derivatives of Chebyshev polynomials. The advantage of this approach is that the derivatives of the polynomials can be computed exactly.

The DMs may be generated using the MATLAB routine CHEBDIF.[21] The discretized $\mathbf{r}$ vector and the calculated DMs are now used to represent the differential operator $\mathfrak{L}_{vp}$ as an $N \times N$ matrix,

$$L_{v_p} = D^{(2)} + \mathrm{diag}\left(\frac{1}{r}\right)D^{(1)} + \left(\frac{\omega^2}{v_p^2}\right)\mathbf{I}, \tag{19}$$

where $\mathrm{diag}(g(r))$ represents a matrix with the elements of a vector $\mathbf{g}(\mathbf{r})$ on the leading diagonal and zeros elsewhere. $\mathbf{I}$ is the identity tensor of size $N \times N$. In the same way matrix representations for all equations of motion as well as displacement and stress components are constructed.

B. Eigenvalue problem

The Helmholtz equations (12) and (13) can be combined as a matrix equation of the following form:

$$\underbrace{\begin{pmatrix} \mathfrak{L}_{v_p} & 0 \\ 0 & \mathfrak{L}_{v_s} \end{pmatrix}}_{\mathfrak{L}} \underbrace{\begin{pmatrix} \Phi \\ \hat{\Psi} \end{pmatrix}}_{\boldsymbol{\Theta}} = k_z^2 \begin{pmatrix} \Phi \\ \hat{\Psi} \end{pmatrix}. \tag{20}$$

To solve Eq. (20) as an eigenvalue problem numerically, the differential operator matrix $\mathfrak{L}$ has to be discretized in analogy to Eq. (19). Equation (20) can now be expressed in terms of DMs where L now is a matrix of size $2N \times 2N$ matrix,

$$\underbrace{\begin{pmatrix} L_{v_p} & 0 \\ 0 & L_{v_s} \end{pmatrix}}_{L} \boldsymbol{\Theta} = k_z^2 \boldsymbol{\Theta}, \tag{21}$$

where

$$L_{vp} = D^{(2)} + \mathrm{diag}\left(\frac{1}{r}\right)D^{(1)} + \left(\frac{\omega^2}{v_p^2}\right)\mathrm{I}, \tag{22}$$

$$L_{vs} = D^{(2)} + \mathrm{diag}\left(\frac{1}{r}\right)D^{(1)} - \left(\frac{1}{r^2}\right)\mathrm{I} + \left(\frac{\omega^2}{v_p^2}\right)\mathrm{I}. \tag{23}$$

Furthermore, the boundary conditions, also expressed in form of DMs have to be substituted. For a free solid bar, the

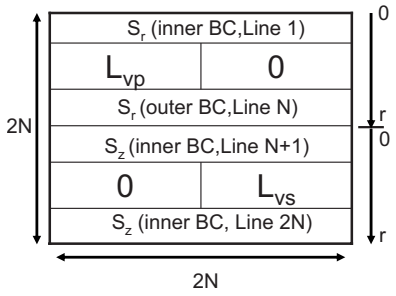


FIG. 2. Structure of the $\tilde{L}$ matrix for a cylinder.

stress-free boundary conditions are assumed at $r=a$, which means $\sigma_{rr}|_{r=a}=\sigma_{rr}|_{r=a}=0$. σ_{rr} is the normal stress in radial direction and σ_{rz} is the radial shear stress acting in z direction.

The expressions for the stress components σ_{rr} [Eq. (16)] and σ_{rz} [Eq. (17)] can also be expressed using the DMs: The resulting equations can be written in a matrix form

$$\begin{pmatrix} \sigma_{rr} \\ \hat{\sigma}_{rz} \end{pmatrix} = \underbrace{\begin{pmatrix} S_{r\Phi} & S_{r\Psi} \\ S_{z\Phi} & S_{z\Psi} \end{pmatrix}}_{S} \begin{pmatrix} \Phi \\ \hat{\Psi} \end{pmatrix}, \tag{24}$$

where submatrices $S_{r\Phi}S_{r\Psi}S_{z\Phi}S_{z\Psi}$ are

$$S_{r\Phi} = -\lambda[\mathrm{diag}(r^{-2}) + \omega^2/v_p^2] + 2\mu D^{(2)}, \tag{25}$$

$$S_{z\Phi} = 2\mu D^{(1)}, \tag{26}$$

$$S_{r\hat{\Psi}} = -2\mu\left[D^{(3)} + \mathrm{diag}\left(\frac{1}{r}\right)D^{(2)} \right. \tag{27}$$

$$\left. - \mathrm{diag}\left(\frac{1}{r^2}\right)D^{(1)} + (\omega^2/v_p^2)I \right], \tag{28}$$

$$S_{z\hat{\Psi}} = \mu\left[2D^{(2)} + 2\,\mathrm{diag}\left(\frac{1}{r}\right)D^{(1)} \right. \tag{29}$$

$$\left. - \mathrm{diag}\left(\frac{1}{r^2}\right) + (\omega^2/v_s^2)I \right]. \tag{30}$$

The last step is to embed appropriate boundary conditions in the matrix representation and replace matrix L with matrix $\tilde{L}$ as shown in Fig. 2. The lines in the L matrix in Eq. (21) corresponding to $r=a$ will be replaced by the corresponding lines of the S matrix. In order to fulfill the stress free boundary conditions, the corresponding values on the right-hand side have to be set equal to zero. In addition, for the lines at $r=0$ the same has to be done. The reason for that is that due to the singularities of the equations at $r=0$ we have to consider a hollow cylinder with a very small inner radius, which is a limiting case for a solid cylinder.

This can be done by introducing a matrix Q on the right-hand side of Eq. (21),

$$\tilde{L}\boldsymbol{\Theta} = k_z^2 Q\boldsymbol{\Theta}, \tag{31}$$

which is a $2N\times 2N$ matrix and defined as follows:

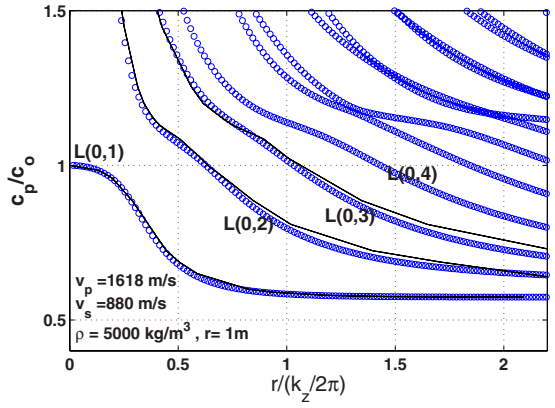


FIG. 3. (Color online) Dispersion curves of an elastic cylinder (circles). x axis: Wave-number-radius product, y axis: Phase velocity $v_{\mathrm{ph}}=\omega/k_z$ normalized by the bar velocity $v_0^2=E/\rho$ where E is the Young's modulus and ρ is density [compare with Davies (Ref. 22, Sec. 11, Fig. 13, lines)].

$$Q = \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix}. \tag{32}$$

Here M is a diagonal matrix which has the following form:

$$M = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & & 0 \end{pmatrix}. \tag{33}$$

Equation (31) is a generalized eigenvalue problem, which means that we cannot find the inverse M^{-1} as $\det(M)=0$. But generalized eigenvalue problems can be solved using the MATLAB routine EIG$(\tilde{L},Q)$ for instance.

In the next section this approach can be extended to n arbitrary cylindrical fluid and solid layers.

C. Dispersion curves

Let us illustrate the results produced by this approach in the form of dispersion curves (Fig. 3). To compare with previous results obtained by root-finding techniques, we use a model presented by Davies.[22] In Fig. 3 the dispersion curves for a free solid bar are computed (circles) with the parameters shown in the picture. These curves are in good agreement (lines) with the dispersion curves provided in Davies,[22] (Fig. 4) which were calculated analytically using root-finding techniques. The fundamental mode $L(0,1)$ behaves like a pure extensional mode for low frequencies and propagates with the velocity $\sqrt{E/\rho}$ where E is the Young's modulus and ρ is density. For higher frequencies the mode propagates like a Rayleigh wave on the cylinder surface. The higher modes $(L(0,1)\ldots L(0,n))$ have cut-off frequencies, which means they do not exist below these frequencies. For very high frequencies they tend to propagate close to the Rayleigh velocity.

IV. MULTIPLE LAYERS

The above-described approach can be extended to n cylindrical fluid and solid layers (see Fig. 4). Each of the n

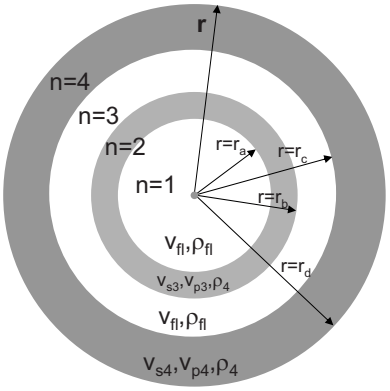


FIG. 4. Geometry of a model with four cylindrical layers. The layer index is $n=1-4$ numbered from the center to the surface of the bar. The layers are either nonviscous fluid $(v_{\mathrm{fl}},\rho_{\mathrm{fl}})$ or elastic solid $(v_{pn},v_{sn},\rho_{\mathrm{fl}})$.

layers has P- and S-wave velocities $(v_{p_1},\ldots,v_{p_n},v_{s_1},\ldots,v_{s_n})$ and densities $\rho_1,\ldots,\rho_n$. In this work we represent the fluid layers as solids with very small shear velocity. For each of the layers the matrix L_n is constructed in analogy to Eq. (21). These matrices are combined in a diagonal matrix of the size $n2N\times n2N$ which has the form

$$L=\begin{pmatrix} L_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & L_n \end{pmatrix}. \tag{34}$$

The same procedure has to be followed for the stress components S_n [see Eq. (24)] and for each layer n which are finally combined in a matrix S of same size as L. A similar matrix U is computed for the displacement components.

For the case of layering, additional conditions of continuity of displacements and stresses have to be introduced,

$$[\sigma_{rr}]|_{r=r_i}=[\sigma_{rz}]|_{r=r_i}=0, \tag{35}$$

$$[u_r]|_{r=r_i}=[u_z]|_{r=r_i}=0, \tag{36}$$

where r_i indicates the position of the interface of the nth layer with $i=a,b,\ldots$.

The interface conditions are introduced as the vanishing differences of the displacements and stresses of the corresponding layers. It is convenient to apply the conditions as illustrated in Fig. 5, which shows in which lines of the L matrix for n layers all boundary conditions have to be introduced. The stress-free boundary conditions on the inner and outer boundary are introduced in the L matrix the same way as for a free cylinder in the rows $(1, N+1, n2N-N$ and $n2N)$.

This means that the elements of S and U representing the interpolation points of the inner and outer boundary and the interfaces replace the corresponding rows in the L matrix, which is now referred to as $\tilde{L}$. The eigenvalue problem can now be formulated analogous to Eq. (31) and solved using a generalized eigenvalue routine.

Dispersion curves: Fluid-filled cylinder. The second example (Fig. 6) is a two-layer model: A fluid-filled cylinder. The dispersion curves were originally calculated by Del Grosso and McGill.[9] Here the dispersion curves (lines) were computed by Sidorov using the *root-finding* technique analogous to that of Ref. 9. Again we were able to reproduce these results accurately using the spectral method. The dispersion curves referred to as ETn are for stress free surface boundary conditions, while the Rn modes were computed for rigid surface boundary conditions.

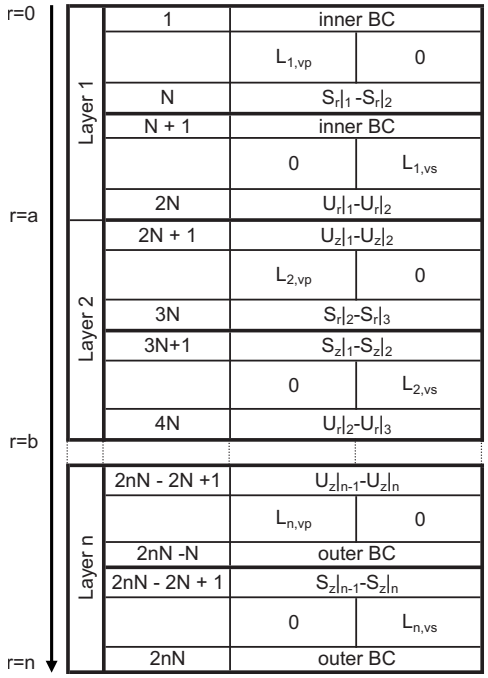


FIG. 5. Structure of the matrix $\tilde{L}$ for n layers.

In the case of a stress free surface there exist two fundamental modes starting from zero frequency: The first one (ET0) is commonly referred to as a tube wave or Stoneley wave, while the second (ET1) is an analog of a (longitudinal) plate or extensional wave. Mode ET1 only weakly depends on the fluid properties and disappears when the thickness of the cylinder wall increases to infinity or the outer boundary of the cylinder becomes rigid (Rn).

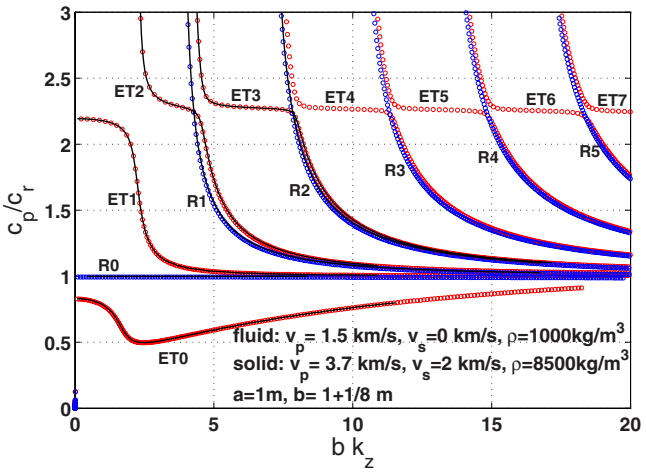


FIG. 6. (Color online) Dispersion curves for a hollow cylinder filled with nonviscous fluid. Thickness of the cylinder wall: 0.125 m. Modes in elastic tube with stress-free outer boundary: ET_n, whereas modes for pipe with rigid outer boundary: R_n. Phase velocity v_{ph} is normalized by the velocity of the fluid $(v_{p,\mathrm{fl}})$ [compare with Del Grosso and McGill (Ref. 9), lines].

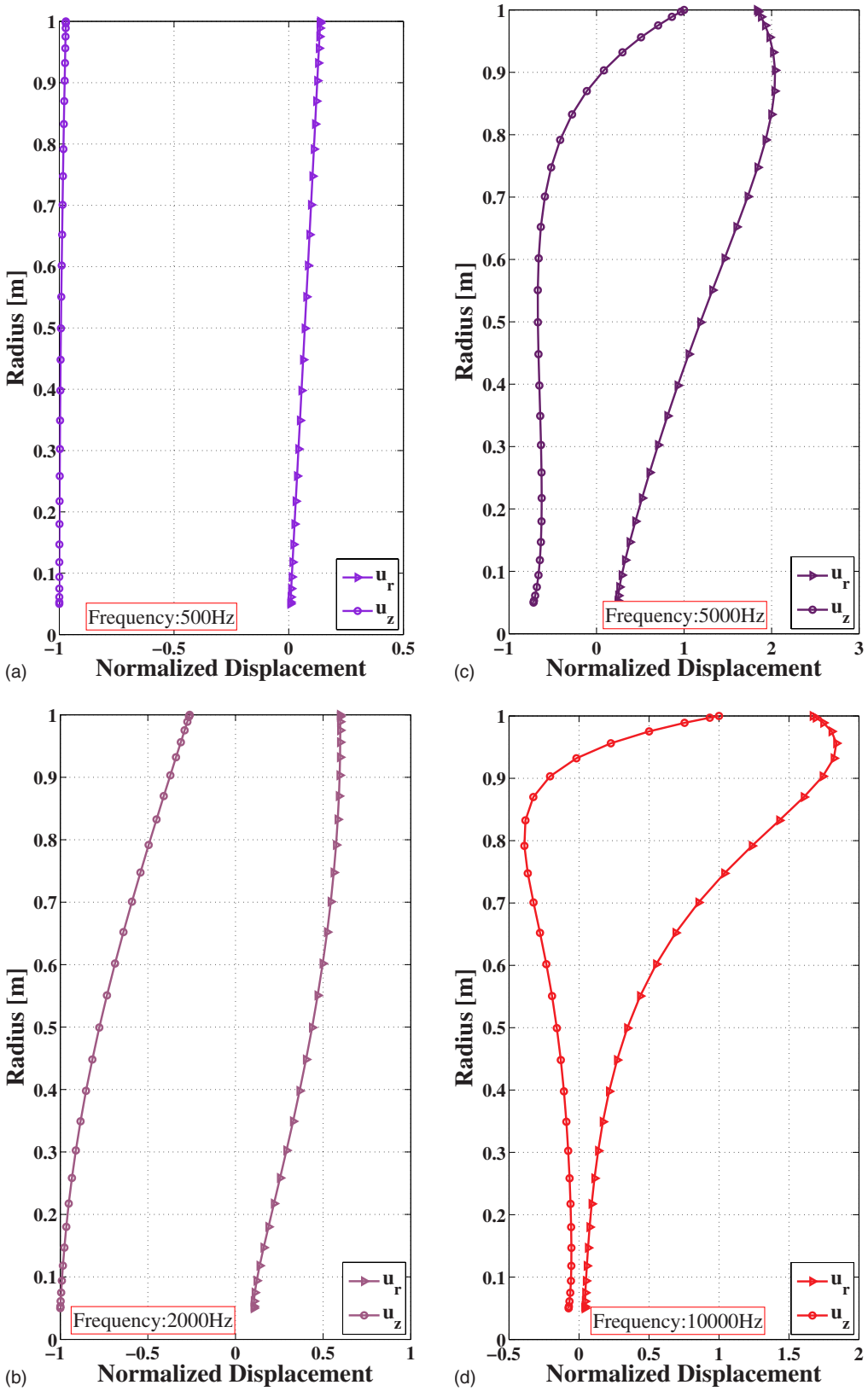


FIG. 7. (Color online) Particle displacement profiles of the fundamental longitudinal mode $L(0,1)$ for (a) 500 Hz, (b) 2000 Hz, (c) 5000 Hz, (d) 10 000 Hz. x axis: Normalized $u_r=|u_r|$ (triangle) and $u_z=i|u_z|$ (circle) displacement. y axis: Bar radius from 0 m (center of bar) to 1 m (surface of bar).

V. PARTICLE DISPLACEMENT PROFILES

In addition to eigenvalues representing dispersion curves, we also obtain eigenvectors representing the potentials Φ and Ψ. They allow the computation of the mode shapes, that is the distribution of field quantities for each mode like displacements, stresses, power flow, etc., along the radius of the cylinder. Figure 7 shows the displacements (u_r, u_z) which can be easily computed using the eigenvectors and Eqs. (4) and (5). In order to display the particle displacement profiles u_r and u_z are calculated along the radius for a

certain frequency. These values are normalized by the maximum absolute value of the u_z displacement. Finally the radial displacement is plotted as $u_r = |u_r|$ and the longitudinal displacement as $u_z = \mathrm{Im}|u_z|$.

For the illustration of the displacement profiles we have chosen the fundamental mode $L(0,1)$ propagating in a free solid cylinder (see Fig. 1). The particle motions u_r and u_z are computed for four different frequencies. Figures 7(a)–7(d) display the displacement profiles for u_r and u_z for the different frequencies.

For low frequencies (500 Hz) [Fig. 7(a)] the wave propagates like a longitudinal body wave. Consequently the particle motion is in the axial direction mainly and uniform throughout the radius of the cylinder. The radial displacement is very small.

In Fig. 7(b) we can see that for 2000 Hz the u_r displacement has already significantly increased all over the cross section. It only remains zero in the center of the cylinder. At the same time the u_z displacement decreases but keeps its maximum value in the center.

For a higher frequency [5000 Hz; Fig. 7(c)] it can be observed that the shape of the displacement profiles evolves slowly toward the typical pattern of Rayleigh modes. Close to the surface ($r = 0.85-1$ m) the motion is already Rayleigh-like. Only toward the center of the bar is the u_z component still relatively strong.

Finally in Fig. 7(d) we get the typical particle motion profile of Rayleigh waves. In contrast to Fig. 7(c) obviously the amplitudes of both displacement components decrease significantly for $r < 0.8$ m.

VI. CONCLUSIONS

We extended and implemented the spectral method for propagation of axisymmetric modes in a cylindrical bar. The method was also generalized to n-layered cylindrical fluid–solid structures. Numerical examples for a free solid cylinder and a fluid-filled tube were given in the form of dispersion curves and particle displacement profiles. Traditional techniques require finding complex roots of nonlinear equations that involve special functions. In contrast, spectral method demands only solving generalized eigenvalue problem without involving special functions. This represents a great simplification that becomes particularly advantageous for complex rheologies like viscoelastic, anisotropic, and poroelastic structures.

There are a lot of directions for further work. One scope is the extension to more complicated media like viscoleasticity and poroelasticity. The approach can also be extended for anisotropic and heterogeneous media. Of great importance will also be to allow unbounded structures, representing a borehole surrounded by infinite rock formation. Finally it would be of great importance to be able to compute the full wave form using the spectral method.

ACKNOWLEDGMENTS

The authors are grateful to Professor Boris Kasthan (St. Petersburg State University, Russia), who suggested the idea of applying the spectral method to the problem at hand. The authors also thank Shell International Exploration and Production for support of this work and Alexander Sidorov (St. Petersburg State University, Russia) for the computation of some dispersion curves using his *root-finding* program. Professor Richard Craster (Imperial College London) is thanked for his helpful advice.

[1] L. Pochhammer, "On the propagation velocities of small oscillations in an unlimited isotropic circular cylinder," J. Reine Angew. Math. **81**, 324–336 (1876).
[2] C. Chree, "The equations of an isotropic elastic solid in polar and cylindrical coordinates, their solutions and applications," Trans. Cambridge Philos. Soc. **14**, 250–369 (1889).
[3] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944).
[4] H. Kolsky, *Stress Waves in Solids* (Dover, New York, 1963).
[5] D. Bancroft, "The velocity of longitudinal waves in cylindrical bars," Phys. Rev. **59**, 588–593 (1941).
[6] D. C. Gazis, "Three-dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation," J. Acoust. Soc. Am. **31**, 568–573 (1959).
[7] D. C. Gazis, "Three-dimensional investigation of the propagation of waves in hollow circular cylinders. II. Numerical results," J. Acoust. Soc. Am. **31**, 573–578 (1959).
[8] J. A. McFadden, "Radial vibrations of thick-walled hollow cylinders," J. Acoust. Soc. Am. **26**, 714–715 (1954).
[9] V. A. Del Grosso and R. E. McGill, "Remarks on 'Axially symmetric vibrations of a thin cylindrical elastic shell filled with nonviscous fluid' by Ram Kumar, Acustica 17 [1968], 218," Acustica **20**, 313–314 (1968).
[10] T. Lin and G. Morgan, "Wave propagation through a fluid contained in a cylindrical, elastic shell," J. Acoust. Soc. Am. **28**, 1165–1176 (1956).
[11] S. I. Rubinow and J. B. Keller, "Wave propagation in a fluid-filled tube," J. Acoust. Soc. Am. **50**, 198–223 (1971).
[12] V. N. R. Rao and J. K. Vandiver, "Acoustics of fluid filled boreholes with pipe: Guided propagation and radiation," J. Acoust. Soc. Am. **105**, 3057–3066 (1997).
[13] G. W. Morgan and J. P. Kiely, "Wave propagation in a viscous liquid contained in a flexible tube," J. Acoust. Soc. Am. **26**, 323–328 (1954).
[14] J. Vollmann, R. Breu, and J. Dual, "High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. II. Experimental results versus theory," J. Acoust. Soc. Am. **102**, 909–920 (1997).
[15] J. Vollmann and J. Dual, "High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. I. Theory and numerical results," J. Acoust. Soc. Am. **102**, 896–908 (1997).
[16] J. G. Berryman, "Dispersion of extensional waves in fluid-saturated porous cylinders at ultrasonic frequencies," J. Acoust. Soc. Am. **74**, 1805–1812 (1983).
[17] G. H. F. Gardner, "Extensional waves in fluid-saturated porous cylinders," J. Acoust. Soc. Am. **34**, 36–39 (1962).
[18] C. J. Wisse, D. M. J. Smeulders, G. Chao, and M. E. H. van Dongen, "Guided wave modes in porous cylinders: Theory," J. Acoust. Soc. Am. **122**, 2049–2056 (2007).
[19] A. T. I. Adamou and R. V. Craster, "Spectral methods for modelling guided waves in elastic media," J. Acoust. Soc. Am. **116**, 1524–1535 (2004).
[20] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North Holland, Amsterdam, 1973).
[21] J. A. C. Weideman and S. C. Reddy, "A MATLAB differentiation matrix suite," ACM Trans. Math. Softw. **26**, 465–519 (2000).
[22] R. M. Davies, "A critical study of Hopkinson pressure bar," Proc. R. Soc. London, Ser. A **240**, 375–457 (1948).

Guided wave propagation and mode differentiation in hollow cylinders with viscoelastic coatings

Jing Mu and Joseph L. Rose[a)]
Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 14 July 2007; revised 12 May 2008; accepted 14 May 2008)

Guided wave propagation theories have been widely explored for about one century. Earlier theories on single-layer elastic hollow cylinders have been very beneficial for practical nondestructive testing on piping and tubing systems. Guided wave flexural (nonaxisymmetric) modes in cylinders can be generated by a partial source loading or any nonaxisymmetric discontinuity. They are especially important for guided wave mode control and defect analysis. Previous investigations on guided wave propagation in multilayered hollow cylindrical structures mostly concentrate on the axisymmetric wave mode characteristics. In this paper, the problem of guided wave propagation in free hollow cylinders with viscoelastic coatings is solved by a semianalytical finite element (SAFE) method. Guided wave dispersion curves and attenuation characteristics for both axisymmetric and flexural modes are presented. Due to the fact that dispersion curve modes obtained from SAFE calculations are difficult to differentiate from each other, a mode sorting method is established to distinguish modes by their orthogonality. Theoretical proof of the orthogonality between guided wave modes in a viscoelastic coated hollow cylinder is provided. Wave structures are also calculated and discussed in view of wave mechanics in multilayered cylindrical structures containing viscoelastic materials. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2940586]

PACS number(s): 43.40.At, 43.20.Mv, 43.20.Bi, 43.20.Ks [YHB] Pages: 866–874

I. INTRODUCTION

The work on wave propagation in bounded structures by Lamb (1917) and Rayleigh (1945), etc. in the early part of the last century sparked the beginning of research on guided waves theories. Ever since then, it has been recognized that there exists a lot more wave modes in free waveguide structures than those in a bulk medium (Graff, 1991; Rose, 1999). These guided wave modes are, in general, much more complex than bulk waves, as they have to satisfy the boundary conditions of the waveguides. Around the 1950s, accompanied with the rapid development of modern computers, there emerged a massive amount of studies on guided wave dispersion curve computations on multilayered structures. The two computational methods that have been widely used until now are the transfer matrix approach developed by Thomson (1950) and the global matrix method by Knopoff (1964). A good summary of these matrix techniques can be found in Lowe (1995). A common feature of these matrix techniques is that a root searching routine has to be established in order to find the roots in the wave number and frequency domain. If the structure contains viscoelastic material, then the root searching process has to be performed on both real and imaginary wave number domains for each frequency. This can possibly make the root searching process time consuming. Moreover, roots may be missed in the searching process at high frequency when two modes approach very close to each other or matrix size becomes larger (Shorter 2004).

In 1959, Gazis obtained the complete solution for harmonic guided wave modes propagating in an infinite hollow cylinder. This has been very beneficial for long range guided wave inspection on widely distributed pipelines. In practice, however, most of the pipelines used in industry are covered with viscoelastic coatings for various protection purposes. Therefore, the exploration of guided wave propagation in viscoelastic coated hollow cylinders becomes quite indispensable.

Due to the aforementioned computation difficulties encountered in matrix methods, theoretical calculation of the wave modes in hollow cylinders with viscoelatic coatings can be difficult. Some researchers have studied plate or plate-like structures with viscoelastic properties, e.g., Simonetti (2004) and Predoi *et al.* (2007). These plate solutions may be safely used to approximate hollow cylinders with a thickness to radius ratio less than 10% (Luo *et al.*, 2005). Therefore, it is still valuable to search for a complete solution to wave propagation in multilayered viscoelastic cylinders. The axisymmetric wave modes of such a structure have been provided by Barshinger and Rose (2004) using the global matrix method, but the flexural modes were not provided in their paper. Ma *et al.* (2006) investigated the fundamental torsional mode scattering from an axisymmetric sludge layer inside a pipe. In his study, the sludge layer is considered to be an elastic epoxy layer. Beard and Lowe (2003) used guided waves to inspect the integrity of rock bolts. They provided three lower order (circumferential order equals 1) flexural modes in their study at relatively low frequency range (less than 300 kHz). Therefore, to the authors' best knowledge, such a complete solution, including both axisymmetric and nonaxisymmetric modes (flexural modes with circumferential orders equal or larger than one) in multilayered hollow cylindrical structures covering a relative wide fre-

[a)]Electronic mail: jlresm@engr.psu.edu

 0001-4966/2008/124(2)/866/9/$23.00

quency range (up to mega Hertz), has never been reported in the literature before. To the authors' knowledge, there has been a lack of investigation on the characteristics of higher order flexural modes in viscoelastic multilayered hollow cylindrical structures. The flexural modes can be generated by any partial source loading or any nonaxisymmetric anomaly in a cylinder. The flexural modes have shown to be highly contributive in the analysis of source influence and guided wave mode control (e.g., focusing) in elastic hollow cylinders (Ditri and Rose 1992; Li and Rose 2001, 2002). Therefore, our goal in this paper is to understand and seek the complete solution to this problem using appropriate techniques. The semianalytical finite element method (SAFEM) was developed as an alternative way to tackle the wave propagation problem. Early employment of SAFEM in solving guided wave propagation problems can be found in Nelson *et al.* (1971) and Dong and Nelson *et al.* (1972). In recent years, SAFEM was also applied to the analysis of wave modes across a pipe elbow (Hayashi *et al.*, 2005) and in materials with viscoelastic properties (Shorter, 2004; Bartoli *et al.*, 2006).

In this paper, SAFEM is adopted to generate phase velocity and attenuation dispersion curves including both axisymmetric and flexural modes in hollow cylinders with viscoelastic coatings. Modal characteristics, such as wave structures and attenuation properties are provided and discussed. Further, driven by the fact that the wave modes obtained from SAFE calculations are difficult to differentiate from each other, a mode sorting algorithm based on modal orthogonality is accomplished. An orthogonality relation based upon SAFE formulation for elastic waveguides is developed by Damljanović and Weaver (2004). Different from Damljanović and Weaver, the orthogonality relation derived in this study is valid for both elastic and viscoelastic materials. It can be used in either single-layered or multilayered cylindrical structures. It is applicable not only for dispersion curves calculation by SAFE formulations but also for those obtained from analytical derivations.

II. SAFE FORMULATION

Let us start with the governing equation provided by the virtual work principle for a free hollow cylinder with as in Hayashi *et al.* (2003) and Sun (2004). Only linear elastic and viscoelastic material behaviors are considered here:

$$\int_V \delta\mathbf{u}^T \cdot \rho\ddot{\mathbf{u}}\, dV + \int_V \delta\boldsymbol{\varepsilon}^T \cdot \boldsymbol{\sigma}\, dV = 0, \tag{1}$$

where T represents matrix transpose, ρ is density, and $\ddot{\mathbf{u}}$ is the second derivative of displacement $\mathbf{u}$ with respect to time t. $\int_V dV$ is the volume integrals of the element, respectively. In cylindrical coordinates, $dV = r\, dr\, d\theta\, dz$. The first and second terms on the left-hand side are the corresponding increment of kinetic energy and potential energy.

Sun (2004) used two-dimensional (2D) SAFE to calculate the flexural modes in a single elastic cylinder. He meshed the cross section of the cylinder in both thickness direction r and circumferential direction θ. In this study, we incorporate the solution $e^{in\theta}$ in the circumferential direction, so that we can use exact representations in both the θ and z directions. The finite element approximation reduces to only one dimension r. This does not only improves the accuracy in the calculation for flexural modes with higher circumferential orders, but also reduces computation cost. For a harmonic wave propagating in the z direction, the displacement at any point $\mathbf{u}(r,\theta,z,t)$ can be represented by

$$\mathbf{u}(r,\theta,z,t) = \sum_{j=1}^{2} \mathbf{N}(r)\mathbf{U}^j e^{i(kz+n\theta-\omega t)}, \tag{2}$$

where $\mathbf{U}^j$ is the nodal displacement vector at the jth element and $\mathbf{N}(r)$ is the shape function in the thickness direction r. For a two-node element, $\mathbf{U}^j$ is a six-element vector and $\mathbf{N}(r)$ is a 3×6 matrix. The shape function matrix is chosen as follows:

$$\mathbf{N} = \begin{bmatrix} N_1 & 0 & 0 & N_2 & 0 & 0 \\ 0 & N_1 & 0 & 0 & N_2 & 0 \\ 0 & 0 & N_1 & 0 & 0 & N_2 \end{bmatrix}, \tag{3}$$

with

$$N_1 = \tfrac{1}{2}(1-\xi), \quad N_2 = \tfrac{1}{2}(1+\xi), \tag{4}$$

where $-1 \leqslant \xi \leqslant 1$ is the natural coordinates in the r direction.

Substituting Eqs. (2)–(4) into Eq. (1), one obtains Eq. (5) after simplification:

$$(\mathbf{K}_1^j + ik\mathbf{K}_2^j + k^2\mathbf{K}_3^j)\mathbf{U}^j - \omega^2\mathbf{M}^j\mathbf{U}^j = 0. \tag{5}$$

In Eq. (5), we have

$$\mathbf{B}_1 = \begin{bmatrix} \frac{\partial N_1}{\partial r} & 0 & 0 & \cdots & & \\ \frac{N_1}{r} & i\frac{n}{r}N_1 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & i\frac{n}{r}N_1 & \cdots & N_2 & \cdots \\ 0 & 0 & \frac{\partial N_1}{\partial r} & & & \\ i\frac{n}{r}N_1 & \frac{\partial N_1}{\partial r} - \frac{N_1}{r} & 0 & \cdots & & \end{bmatrix}_{6\times 6}, \tag{6}$$

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & N_1 & 0 & 0 & N_2 \\ 0 & N_1 & 0 & 0 & N_2 & 0 \\ N_1 & 0 & 0 & N_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{6\times 6}, \tag{7}$$

$$\mathbf{K}_1^j = \int_r \int_\theta \mathbf{B}_1^T \mathbf{C} \mathbf{B}_1 r \, dr \, d\theta,$$

$$\mathbf{K}_2^j = \int_r \int_\theta (\mathbf{B}_1^T \mathbf{C} \mathbf{B}_2 - \mathbf{B}_2^T \mathbf{C} \mathbf{B}_1) r \, dr \, d\theta, \tag{8}$$

$$\mathbf{K}_3^j = \int_r \int_\theta \mathbf{B}_2^T \mathbf{C} \mathbf{B}_2 r \, dr \, d\theta,$$

$$\mathbf{M}^j = \rho \int_r \int_\theta \mathbf{N}^T \mathbf{N} r \, dr \, d\theta,$$

and $\mathbf{C}$ is the stiffness matrix.

Assembling Eq. (5) for all elements in the cross-sectional area, the equation for the whole system can be written as

$$(\mathbf{K}_1 + ik\mathbf{K}_2 + k^2\mathbf{K}_3)\mathbf{U} - \omega^2 \mathbf{M}\mathbf{U} = \mathbf{f}, \tag{9}$$

where $\mathbf{K}_1$, $\mathbf{K}_2$ $\mathbf{K}_3$, and $\mathbf{M}$ are the $M \times M$ matrices where M equals three times the number of nodes. In order to decrease the order of k, Eq. (9) can be further reformulated as

$$[\mathbf{A} - k\mathbf{B}]\mathbf{Q} = 0, \tag{10}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{K}_1 - \omega^2 \mathbf{M} \\ \mathbf{K}_1 - \omega^2 \mathbf{M} & i\mathbf{K}_2 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{K}_1 - \omega^2 \mathbf{M} & 0 \\ 0 & -\mathbf{K}_3 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{U} \\ k\mathbf{U} \end{bmatrix}.$$

At a given frequency ω, the wave number k can be obtained by solving the eigenvalue problem in Eq. (10) using a standard eigenvalue routine. Then, the phase velocity and attenuation dispersion curves can be calculated from the real and imaginary parts of k. The corresponding wave structure $\mathbf{U}$ can be obtained from the upper half of the eigenvector $\mathbf{Q}$.

III. ORTHOGONALITY AND MODE SORTING

A common difficulty in producing dispersion curves calculated from the SAFE formulation is modal differentiation. Modal differentiation is tremendously helpful to interpret the behaviors of different modes. It also forms the basis of solving such further problems as wave scattering, source influence, and mode control. Of particular importance in mode sorting is to find the distinctive characteristics between different modes. A natural way of realizing this is to utilize the modal orthogonality, which is an intrinsic law followed by all guided wave modes. Damljanović and Weaver (2004) presented an orthogonality relation based on SAFE formulation for elastic waveguides and used it to solve a point source loading problem. Loveday and Long (2007) also employed this relation to sort the guided wave modes in a rail. Another way of sorting wave modes can be realized by identifying the similarity between wave modes at adjacent frequencies. However, we prefer to using orthogonality here not only due to the fact that orthogonality is a natural attribute of guided wave modes, but also because mode sorting based on orthogonality can help future research on wave scattering and source influence in multilayered cylinders.

The analytical orthogonality relations have been studied and used by researchers for decades. The orthogonality relation for the Rayleigh–Lamb modes of a 2D plate was obtained by Fraser (1976). Later, orthogonality relations of guided wave modes have been used as a powerful tool in the study of wave scattering (Kino, 1978; Engan, 1998; Shkerdin and Glorieux, 2004, 2005; Vogt *et al.*, 2003) and source influence (Ditri and Rose, 1992, 1994) analysis. In these studies, orthogonality relations are derived from either real or complex reciprocity relations. As has been pointed out by Auld (1990), both real and complex reciprocity relations are valid for elastic waveguides. If we set out from the real reciprocity relation, we will reach an orthogonality relation stating that the mode is orthogonal to all the other modes except a mode with the same modal behavior, but incoming from the opposite propagating direction, which is the case in this paper. On the other hand, starting out from the complex reciprocity relation yields an orthogonality relation stating that the mode is orthogonal to all the other modes except itself. Like the two reciprocity relations, these two types of orthogonality relations are both valid for elastic waveguides. However, only the real reciprocity relation is valid for viscoelastic materials and our purpose is to obtain an orthogonality relation that is applicable for viscoelastic waveguides. Thus, it is necessary to build the orthogonality relation base on the real reciprocity relation. The utilization of orthogonality relations is associated with guided wave modes and, therefore, is dependent on waveguide geometries. For example, to solve the source influence problem in hollow cylinders, the reciprocity relation formulated in three-dimensional (3D) cylindrical coordinates has to be used, whereas a reciprocity relation formulated in a 2D Cartesian coordinated may be used to analyze the wave field in a plate. The orthogonality relations for 3D solid cylinders and hollow cylinders have been provided by Vogt *et al.* (2003), Engan (1998), and Ditri and Rose (1992) by using the stress free boundary conditions under cylindrical coordinates. Here, for a multilayered viscoelastic cylinder, we show that the orthogonality relation is still valid by taking into account both the interface continuity conditions and stress free boundary conditions.

In this section, the analytical derivation of orthogonality of the modes in multilayered cylindrical structures containing both elastic and viscoelastic materials will be given. Different from the orthogonality relation developed by Damljanović and Weaver (2004), the orthogonality developed in this paper can be applied to multilayered waveguides containing any combination of elastic and viscoelastic materials. In addition, as the following orthogonality relation is derived analytically, it can be applied to dispersion curves obtained from either the SAFE formulation or from the matrix methods. In the following, the mode sorting process will be discussed accordingly.

Before we get started, it is worthwhile to introduce the notation used in the derivation to represent guided wave modes in cylindrical structures. Overall, we follow the notations used in Ditri and Rose (1992). Guided wave modes in a cylindrical waveguide can be represented by two indices, e.g., $T(m,n)$ or $L(m,n)$, where $m \in \{0,1,2,\ldots\}$ is the index of circumferential order, $n \in \{1,2,\ldots\}$ is the nth root of the characteristic equation of circumferential order m for torsional type (denoted by T) and longitudinal type (denoted by L) of the mode, respectively. Conventionally, modes with the same circumferential order m are often called a mode group or a mode family. In this paper, a mode group or family refers to longitudinal or torsional types of modes of the same n but with different circumferential orders. The reason to do this lays in the fact that longitudinal or torsional types of modes with the same n, but different circumferential orders, bear similar modal characteristics, such as phase velocities, attenuation, and wave structures in the thickness direction.

The orthogonality of normal modes in an elastic hollow cylinder was first developed by (Ditri and Rose 1992). The orthogonality of normal modes in multilayered hollow cylinders containing viscoelastic materials can be derived in a similar manner. Therefore, the derivation procedure will only be given briefly here. Let us start with the real reciprocity relation (Auld 1990)

$$\nabla \cdot (\mathbf{v}_1 \cdot \mathbf{T}_2 - \mathbf{v}_2 \cdot \mathbf{T}_1) = 0, \tag{11}$$

where $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{T}_1$, $\mathbf{T}_2$ are the particle velocities and stresses for two different wave modes (either torsional or longitudinal type) in a multilayered hollow cylinder. Without loss of generality, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{T}_1$, $\mathbf{T}_2$ can be represented as

$$\mathbf{v}_1 = \mathbf{v}_m^N e^{-i\beta_m^N z}, \quad \mathbf{T}_1 = \mathbf{T}_m^N e^{-i\beta_m^N z}, \tag{12}$$

$$\mathbf{v}_2 = \mathbf{v}_n^M e^{-i\beta_n^M z}, \quad \mathbf{T}_2 = \mathbf{T}_n^M e^{-i\beta_n^M z}, \tag{13}$$

where N and M are circumferential orders, n and m are indices of mode group, and β denotes the wave number of mode (M,n) or (N,m).

Substituting the previous expressions for $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{T}_1$, $\mathbf{T}_2$ into Eq. (11) yields

$$\nabla_{r\theta} \cdot (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) - i(\beta_m^N + \beta_n^M)(\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{e}}_z = 0, \tag{14}$$

where $\nabla_{r\theta}$ is the 2D divergence operator in cylindrical coordinates. By integrating both sides of Eq. (14) over the cross section D of the viscoelastic coated hollow cylinder, we obtain

$$\iint_D \nabla_{r\theta} \cdot (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) d\sigma = -4i(\beta_m^N + \beta_n^M) P_{mn}^{NM}, \tag{15}$$

where

$$P_{mn}^{NM} = -\frac{1}{4} \iint_D (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{e}}_z d\sigma. \tag{16}$$

By applying the Gauss divergence theorem to the left-hand side of Eq. (15), we have

$$\begin{aligned} \iint_D \nabla_{r\theta} \cdot (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) d\sigma = & \oint_{\partial_1 D} (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{n}}_1 ds + \oint_{\partial_2 D} (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{n}}_2 ds \\ & + \oint_{\partial_2 D} (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{n}}_1 ds + \oint_{\partial_3 D} (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{n}}_2 ds, \end{aligned} \tag{17}$$

where $\partial_1 D$ represents the inner boundary of the cross section of the elastic hollow cylinder, $\partial_2 D$ represents the interface between the cross sections of the elastic hollow cylinder and the viscoelastic coating, and $\partial_3 D$ represents the outer boundary of the cross section of the viscoelastic coating. The unit vectors $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ are defined as

$$\hat{\mathbf{n}}_1 = -\hat{\mathbf{r}}, \quad \hat{\mathbf{n}}_2 = \hat{\mathbf{r}}. \tag{18}$$

By using the fact that the displacements and normal stresses are continuous at the interface $\partial_2 D$, Eq. (17) can be simplified to

$$\begin{aligned} \iint_D \nabla_{r\theta} \cdot (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) d\sigma = & \oint_{\partial_1 D} (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{n}}_1 ds + \oint_{\partial_3 D} (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) \cdot \hat{\mathbf{n}}_2 ds. \end{aligned} \tag{19}$$

Further, by noticing that the tractions produced by the modes (M,n) and (N,m) vanish at the free boundaries of the waveguides for the right-hand side of Eq. (19), we obtain

$$\iint_D \nabla_{r\theta} \cdot (\mathbf{v}_m^N \cdot \mathbf{T}_n^M - \mathbf{v}_n^M \cdot \mathbf{T}_m^N) d\sigma = 0. \tag{20}$$

Combining Eqs. (15) and (20), we acquire

$$P_{mn}^{NM} = 0, \quad \beta_m^N \neq -\beta_n^M. \tag{21}$$

In addition to Eq. (21), direct evaluation of P_{mn}^{NM} using the orthogonality of the angular eigenfunctions $\cos(N\theta)$, $\sin(M\theta)$, etc. provides us with the orthogonality relation in multilayered hollow cylinders containing viscoelastic materials

$$P_{mn}^{NM} = 0, \quad \text{for } M \neq N \text{ or } \beta_m^N \neq -\beta_n^M. \tag{22}$$

Once the orthogonality relation is obtained, guided modes can be sorted by calculating P_{mn}^{NM} between the modes of adjacent frequencies. For instance, solving Eq. (10) yields N_1 modes at frequency ω_1 and N_2 modes at the adjacent frequency ω_2, where $\omega_2 - \omega_1 = \Delta\omega$ is the frequency step. P_{mn}^{NM} is then calculated between each mode at ω_2 and all of the modes at ω_1 using Eq. (17) to obtain N_1 results. The biggest value in the N_1 results indicates the two modes used for this orthogonality calculation belonging to the same mode in the dispersion curves. In fact, the other values will be very close to zero. Here, we do not have to know the values of m, n, and M, N. We only need to utilize the wave structures obtained from Eq. (10) for the two modes under concern, calculate

TABLE I. Material properties.

Material	C_L (mm/μs)	$\frac{\alpha_L}{\omega}$ (μs/mm)	C_S (mm/μs)	$\frac{\alpha_S}{\omega}$ (μs/mm)	ρ (g/cm^3)
Steel	5.85	⋯	3.23	⋯	7.86
E&C 2057/ Cat9 expoxy	2.96	0.0047	1.45	0.0069	1.6
Bitumastic 50	1.86	0.023	0.75	0.24	1.5

their corresponding stress components, and input these stress and displacement components into Eq. (16) to calculated P_{mn}^{NM}. It can be seen from this procedure that the orthogonality computation is performed $N_1 \times N_2$ times for the two adjacent frequencies. Strictly speaking, the orthogonality relation in Eq. (22) is valid among the modes under any single arbitrary frequency. However, we are performing it between two adjacent frequencies ω_1 and ω_2 as an approximation. As the behavior of the wave modes in the dispersion curves varies continuously, this approximation holds well for relatively small frequency steps. The frequency step used in our calculation is 10 kHz and the mode sorting procedure worked effectively.

IV. NUMERICAL RESULTS

A. Low attenuative material: E&C 2057 Cat9 expoxy

In order to verify our calculation, the dispersion curves for a 4 in. schedule 40 steel hollow cylinder coated with 0.02 in. E&C 2057 / Cat9 expoxy are calculated and compared with previous axisymmetric results given by Barshinger and Rose (2004). Nevertheless, the formulation provided in this paper is general. It is applicable to other pipe sizes and other viscoelastic and elastic cylinders. The materials that can be used in analytical matrix method can also be implemented in our calculations. It is simply a different input for material and pipe size, the solving procedure is the same. Even more complex material properties, for example, experimental material properties varying nonlinearly with frequency can be easily implemented. However, this is beyond the scope of this paper and will not be addressed in detail here. The numbers of elements used in our calculation for the elastic and viscoelastic layers are 24 and 4, respectively. The material properties are listed in Table I, where C_L and C_S are the longitudinal and shear bulk wave velocities, respectively. The terms α_L and α_S are attenuation parameters associated with longitudinal and shear bulk waves. They are frequency dependent. The viscoelastic material properties are experimentally measured. Details can be found in Barshinger (2001). Based on the correspondence principle (Christensen, 1981), wave propagation problems in elastic materials can be converted to those in viscoelastic materials by simply using complex material parameters. This makes the SAFE calculations for both elastic and viscoelastic cases essentially the same, except that the input parameters are different.

The dispersion curves of the longitudinal mode group (denoted by L) and the torsional mode group (denoted by T) after mode sorting for axisymmetric (circumferential order n equals zero) and flexural modes (circumferential order n ranges from one to ten) are plotted in Figs. 1 and 2, respectively. For clarity, a magnified inset for wave attenuation at low frequencies (below 300 kHz) is also shown in Fig. 2. In Figs. 1 and 2, the different modes are lined up very well with respect to frequency. Mode sorting can be very helpful when using guided wave modes. Especially when the number of wave modes is big (e.g., more than 20 modes), there may be

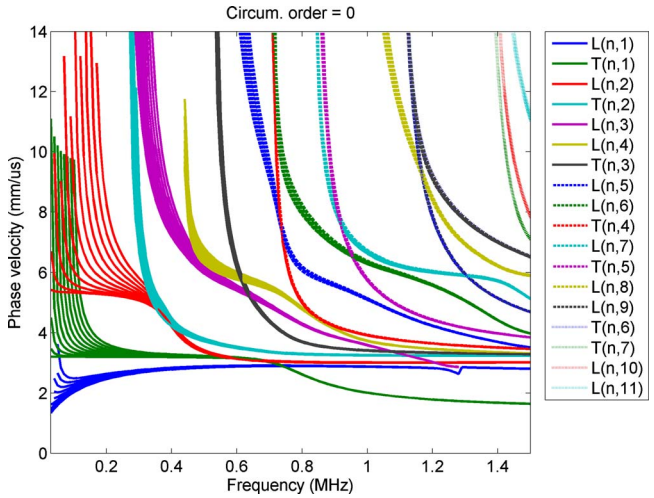


FIG. 1. (Color online) Phase velocity dispersion curves for guided wave modes with circumferential order n from 0 to 10 in a 4 in. schedule 40 steel hollow cylinder coated with 0.02 in. thick E&C 2057 Cat9 expoxy.

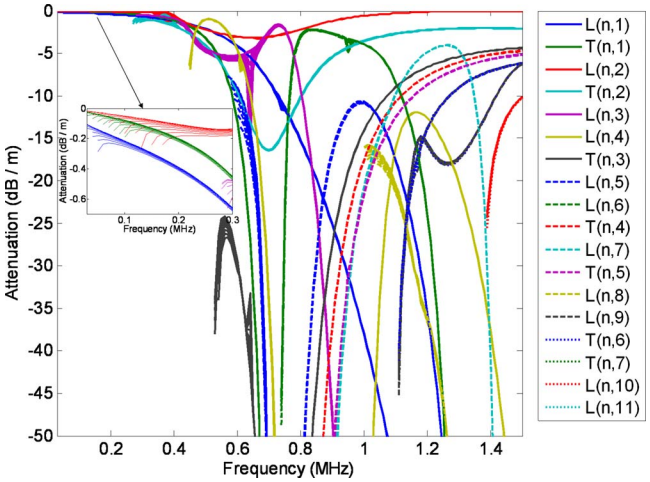


FIG. 2. (Color online) Attenuation dispersion curves for guided wave modes with circumferential order n from 0 to 10 in a 4 in. schedule 40 steel hollow cylinder coated with 0.02 in. thick E&C 2057 Cat9 expoxy.

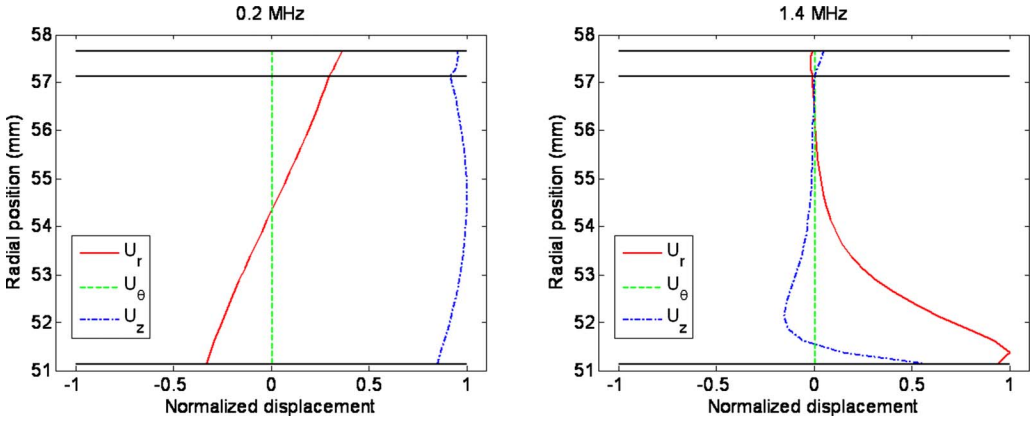


FIG. 3. (Color online) Normalized displacement distribution across hollow cylinder thickness for the L(0,2) mode at 0.2 and 1.4 MHz.

multiple mode crossings or mode branches, which makes it difficult to identify the trend of a specific mode under investigation. Our next example on highly attenuative material will illustrate this aspect even better. For well sorted wave modes, it will be very convenient to analyze the behavior of a specific mode with respect to frequency. On one hand, the orthogonality between different modes in natural waveguides provides us with a natural way of mode sorting. On the other hand, mode sorting can also serve as an effective means of verifying the above derivation and calculation of dispersion curves. The results in Figs. 1 and 2 validated our theory of orthogonality. This orthogonality mode sorting method is a powerful tool and can be applied to various waveguides such as multilayered plates, beams, rods, rails, and so on.

It can be seen from Fig. 2 that guided wave attenuation is majorly nonmonotonic with an increase in frequency. However, if observed carefully, several characteristics can be noticed: (1) The most striking characteristic in Fig. 2 is that the mode group $L(n,2)$ has low attenuation in almost the whole frequency bandwidth. Especially for frequency higher than 1.2 MHz, the attenuation of the mode group $L(n,2)$ converges to zero asymptotically. (2) The attenuation of mode groups $L(n,1)$ and $T(n,1)$ increases monotonically with an increase in frequency. The attenuations with these modes below 0.2 MHz are less than 1 dB/m. As a result, the two wave groups in this region are good choices in nondestructive testing. (3). Generally speaking, for almost all of the other mode groups having cutoff frequencies in the corresponding bare hollow cylinder problem, very high attenuation occurs close to their cutoff frequencies. This makes sense, as they become nonpropagating modes for frequencies lower than their cutoff frequencies in a hollow cylinder without viscoelastic coatings. As the frequency increases, the attenuation (the absolute value of attenuation in Fig. 2) decreases and reaches its minimum at a certain frequency for a certain mode. After that, the attenuation increases again with an increase of frequency. This is a significant characteristic as it occurs for almost all of the wave modes. Finding out where the minimal attenuation values occur for the different wave modes is crucial for nondestructive evaluation applications and would be an interesting subject for future work.

For comparison purposes, the phase velocity and attenuation dispersion curves for axisymmetric longitudinal and torsional modes are also calculated from the analytical formulation developed in Barshinger and Rose (2004). It is found that the longitudinal and torsional dispersion curves generated from the analytical matrix method match the axisymmetric results in Figs. 1 and 2 very well. However, the computational cost of the SAFE method is less even with the mode sorting procedure incorporated. For brevity, these comparison results will not be shown in this paper.

Wave structures are displacement amplitude distributions along the thickness direction for certain guided wave modes. The behavior of a specific guided wave mode is highly related to its wave structure. From the previous analysis regarding the phase velocity and attenuation dispersion curves, it is natural to consider the mode group $L(n,2)$ of most interest, since this group is the least attenuative at both low ($<$0.32 MHz) and high ($>$1 MHz) frequencies. Sample wave structures of $L(0,2)$ (axisymmetric mode) and $L(5,2)$ (flexural mode) at 0.2 and 1.4 MHz are shown in Figs. 3 and 4, respectively. The coated hollow cylinder size and material parameters are the same as those used in the dispersion curve calculations.

It can be seen that the wave structure of $L(0,2)$ at low frequency (0.2 MHz) has a similar distribution as the S0 mode in plate. The wave structure changes with frequency. At a frequency of 1.4 MHz, the wave structure of $L(0,2)$ changes to become similar to that of a surface wave. In addition, the displacement is mostly concentrated on the inner surface of the two layered hollow cylinder. This phenomenon explains why the attenuation of $L(0,2)$ approaches zero at high frequencies. Comparison of wave structures from $L(0,2)$ and $L(5,2)$ reveals that the displacements in the radial and axial directions (U_r and U_z) have very similar distributions for axisymmetric modes and flexural modes in a mode group. However, the displacements in the circumferential direction for $L(0,2)$ and $L(5,2)$ are similar at high frequency but different at low frequency. This can be expected because the wave velocities of the modes with different circumferential orders in a mode group are quite different at low frequency, but approach each other at high frequency.

From the earlier discussion, it can be seen that wave mode attenuation in a multilayered hollow cylinder is highly related to its energy concentration in the viscoelastic layer.

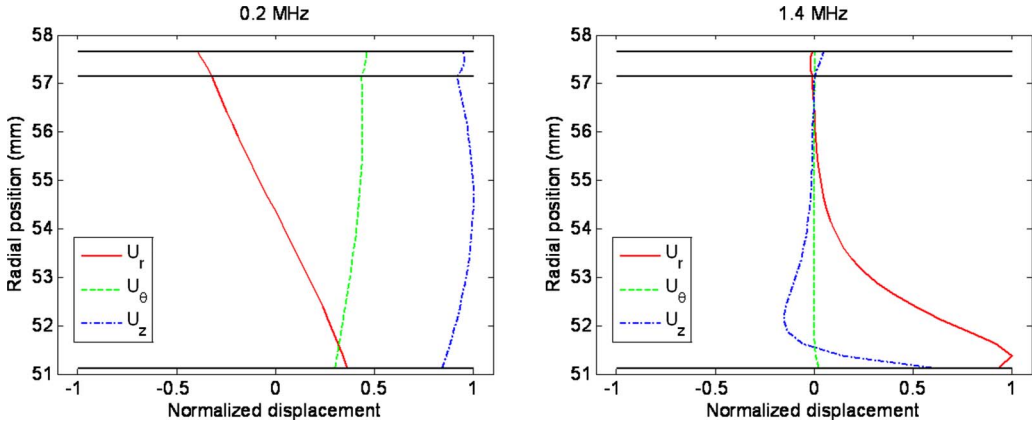


FIG. 4. (Color online) Normalized displacement distribution across hollow cylinder thickness for the L(5,2) mode at 0.2 and 1.4 MHz.

The more the energy is concentrated in the viscoelastic layer, the greater the attenuation for the wave mode. This conclusion is also drawn in Simonetti (2004) to approximate the guided wave attenuation by calculating the portion of energy contained in the viscoelastic layer of a multilayered plate structure.

B. Highly attenuative material: Bitumastic 50

The properties of Bitumastic 50 (Barshinger and Rose, 2004) are also given in Table I. It can be seen from Table I that Bitumastic 50 is much more attenuative, especially for shear waves, compared to E&C 2057/Cat9 epoxy.

The phase velocity dispersion curves and attenuation curves for a 4 in. schedule 40 steel pipe coated with 0.02 in Bitumastic 50 are calculated and shown in Figs. 5 and 6. A threshold is set for displaying the dispersion curves, so that the modes with an imaginary part of a wave number larger than 0.5 are not shown in Fig. 6. Comparing the phase velocity dispersion curves in Fig. 5 to those in Fig. 1, two major features can be observed in Fig. 5. First, the phase velocities of certain modes groups $L(n,3)$ and $L(n,6)$ decrease at their low frequencies. Second, at relatively high frequencies, some modes experience nonmonotonic variation with the increase of frequency, e.g., modes $L(n,3)$ and $L(n,5)$ in the circled areas B and A. Nonmonotonic change of phase velocity with frequency is also found in the dispersion curves of highly attenuative plates (Chan and Cawley, 1998). It is pointed out in the paper that the rising in phase velocity with the increase in frequency may be associated with the amplitude ratio between the longitudinal and shear partial waves in the waveguide.

Figure 6 shows the wave modes whose attenuations are smaller than 100 dB/m. A comparison between the attenuation curves in Figs. 2 and 6 reveals that, overall, the mode groups in a 4 in. schedule 40 pipe coated with Bitumastic 50 are more attenuative than the same mode groups in a 4 in. schedule 40 pipe but coated with E&C 2057/Cat9 epoxy of the same thickness. This agrees with the material properties of the two coatings.

Figure 7 shows the magnified axisymmetric phase velocity dispersion curves in the circled area A in Fig. 5. The modal behavior in this region is relatively more complex compared to that in the other regions. In Fig. 7, several modes cross each other and mode $L(0,5)$ changes nonmonotonically with the increase of frequency. Eight points are chosen on the three guided wave modes shown in Fig. 7, they are labeled A–H. The detailed information of the chosen modes including frequency, phase velocity values, and at-

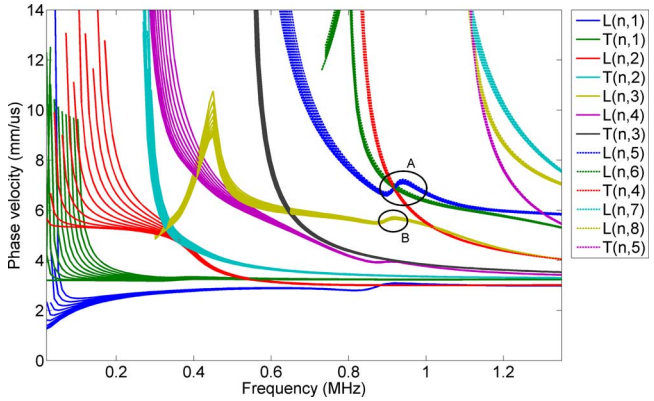


FIG. 5. (Color online) Phase velocity dispersion curves for guided wave modes with circumferential order n from 0 to 10 in a 4 in. schedule 40 steel hollow cylinder coated with 0.02 in. thick Bitumastic 50.

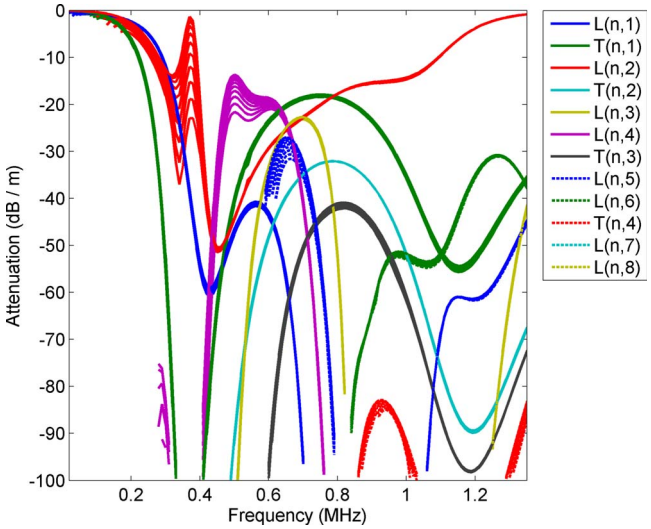


FIG. 6. (Color online) Attenuation curves for guided wave modes with circumferential order n from 0 to 10 in a 4 in. schedule 40 steel hollow cylinder coated with 0.02 in. thick Bitumastic 50.

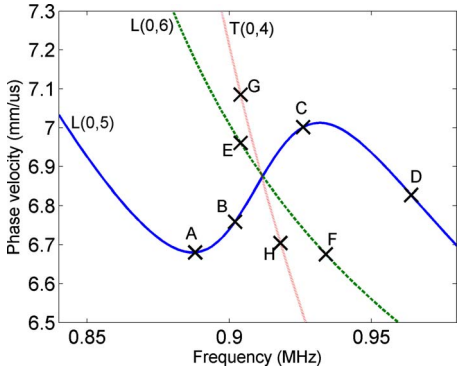


FIG. 7. (Color online) Magnified axisymmetric phase velocity dispersion curves in the circled area A in Fig. 5.

tenuation values are listed in Table II. It may be noticed from Table II that the attenuation values of mode $L(0,5)$ at points A, B, and C are much higher than the attenuation at point D. An inspection of the other modals reveals that the modes are generally more attenuative at the regions where phase velocity rising with the increase in frequency than the surrounding regions. The wave structures of the four points A to D on the mode $L(0,5)$ are plotted in Fig. 8. It can be clearly seen from Fig. 8 that the four wave structures are very similar to each other, which verifies that they are from the same mode and the mode $L(0,5)$ does evolve nonmonotonically with frequency in this region. Further investigation on modes $L(0,6)$ and $T(0,4)$ also reveals that the mode sorting in accurate in this region.

TABLE II. List of selected modes.

Label	Mode	Frequency (MHz)	Phase velocity (mm/μs)	Attenuation (dB/m)
A	L(0,5)	0.888	6.6802	−698.95
B	L(0,5)	0.902	6.7590	−875.21
C	L(0,5)	0.926	7.001	−797.84
D	L(0,5)	0.964	6.8276	−392.48
E	L(0,6)	0.904	6.9608	−71.052
F	L(0,6)	0.934	6.6752	−62.535
G	T(0,4)	0.904	7.0852	−98.064
H	T(0,4)	0.918	6.7044	−98.052

V. CONCLUSION

In this paper, the phase velocity and attenuation dispersion curves for a hollow cylinder with viscoelastic coating are developed by a SAFE formulation. It is analytically shown that the guided wave modes in such a multilayered cylinder containing viscoelastic materials are normal modes with orthogonality relations. Similar orthogonality relations can also be expected in various waveguides such as multilayered plates, rods, and rails. A mode sorting method based upon the orthogonality of normal modes is applied to the dispersion curves and excellent mode sorting results are obtained. The mode sorting results also validate the orthogonality relation. Numerical results are given to two different kinds of viscoelastic coating materials to verify the theoretical derivations and explore wave propagation characteristics in each case. Dispersion curves and sample wave structures are pro-

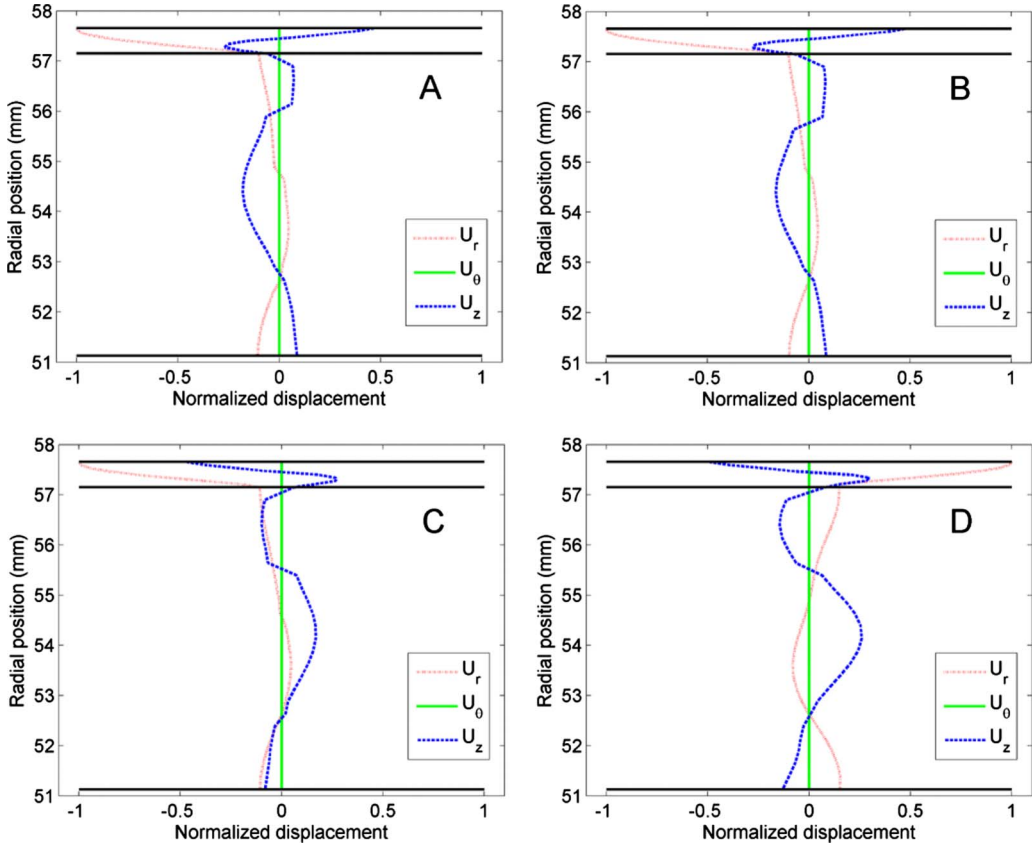


FIG. 8. (Color online) Wave structures of A, B, C, and D on mode L(0,5) in Fig. 7.

vided and the relation between wave structures and attenuation characteristics for flexural modes are discussed.

Auld, B. A. (**1990**). *Acoustic fields and waves in solids*, 2nd ed. (Krieger, Malabar, FL), Vol. **II**, pp. 153–154.

Barshinger, J. N. (**2001**). "Guided wave propagation in pipes with viscoelastic coatings," Ph.D. thesis, The Pennsylvania State University.

Barshinger, J. N., and Rose, J. L. (**2004**). "Guided wave propagation in an elastic hollow cylinder coated with a viscoelastic material," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1547–1556.

Bartoli, I., Marzani, A., Lanza di Scalea, F., and Viola, E. (**2006**). "Modeling wave propagation in damped waveguides of arbitrary cross-section," J. Sound Vib. **295**, 685–707.

Beard, M. D., and Lowe, M. J. S. (**2003**). "Non-destructive testing of rock bolts using guided ultrasonic waves," Int. J. Rock Mech. Min. Sci. **40**, 527–536.

Chan, C. W., and Cawley, P. (**1998**). "Lamb waves in highly attenuative plastic plates," J. Acoust. Soc. Am. **104**, 874–881.

Christensen, R. M. (**1981**). *Theory of Viscoelasticity: An Introduction* (Academic, New York).

Damljanović, V., and Weaver, R. L. (**2004**). "Forced response of a cylindrical waveguide with simulation of the wavenumber extraction problem," J. Acoust. Soc. Am. **115**, 1582–1591.

Ditri, J. J., and Rose, J. L. (**1992**). "Excitation of guided elastic wave modes in hollow cylinders by applied surface tractions," J. Appl. Phys. **72**, 2589–2597.

Ditri, J. J., and Rose, J. L. (**1994**). "Excitation of guided waves in generally anisotropic layers using finite sources," J. Appl. Mech. **61**, 330–338.

Dong, S., and Nelson, R. (**1972**). "On natural vibrations and waves in laminated orthotropic plates," J. Appl. Mech. **39**, 739–745.

Engan, H. E. (**1998**). "Torsional wave scattering from a diameter step in a rod," J. Acoust. Soc. Am. **104**, 2015–2024.

Fraser, W. B. (**1976**). "Orthogonality relation for the Rayleigh-Lamb modes of vibration of a plate," J. Acoust. Soc. Am. **59**, 215–216.

Gazis, D. C. (**1959**). "Three dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation," J. Acoust. Soc. Am. **31**, 568–573.

Graff, K. F. (**1991**). *Wave Motion in Elastic Solids* (Dover, New York).

Hayashi, T., Kawashima, K., Sun, Z., and Rose, J. L. (**2005**). "Guided wave propagation mechaninics across a pipe elbow," J. Pressure Vessel Technol. **127**, 322–327.

Hayashi, T., Song, W.-J., and Rose, J. L. (**2003**). "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," Ultrasonics **41**, 175–183.

Kino, G. S. (**1978**). "The application of reciprocity theory to scattering of acoustic waves by flaws," J. Appl. Phys. **49**, 3190–3199.

Knopoff, L. (**1964**). "A matrix method for elastic wave problems," Bull. Seismol. Soc. Am. **54**, 431–438.

Lamb, H. (**1917**). "On waves in an elastic plate," Proc. R. Soc. London, Ser. A **93**, 114-128.

Li, J., and Rose, J. L. (**2001**). "Excitation and propagation of nonaxisymmetric guided waves in a hollow cylinder," J. Acoust. Soc. Am. **109**, 457–464.

Li, J., and Rose, J. L. (**2002**). "Angular-profile tuning of guided waves in hollow cylinders using a circumferential phased array," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1720–1729.

Loveday, P. W., and Long, C. S. (**2007**). "Time domain simulation of piezoelectric excitation of guided waves in rails using waveguide finite elements," Proc. SPIE **6529**, 65290V.

Lowe, M. J. (**1995**). "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542.

Luo, W., Zhao, X., and Rose, J. L. (**2005**). "A guided wave plate experiment for a pipe," J. Pressure Vessel Technol. **127**, 345–350.

Ma, J., Simonetti, F., and Lowe, M. J. S. (**2006**). "Scattering of the fundamental torsional mode by an axisymmetric layer inside a pipe," J. Acoust. Soc. Am. **120**, 1871–1880.

Nelson, R. B., Dong, S. B., and Kalra, R. D. (**1971**). "Vibrations and waves in laminated orthotropic circular cylinders," J. Sound Vib. **18**, 429–444.

Predoi, M. V., Castaings, M., Hosten, B., and Bacon, C. (**2007**). "Wave propagation along transversely periodic structures," J. Acoust. Soc. Am. **121**, 1935–1944.

Rayleigh, J. (**1945**). *The Theory of Sound* (Dover, New York).

Rose, J. L. (**1999**). *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge).

Schwab, F. (**1970**). "Surface-wave dispersion computations: Knopoff's method," Bull. Seismol. Soc. Am. **60**, 1491–1520.

Shkerdin, G., and Glorieux, C. (**2004**). "Lamb mode conversion in a plate with a delamination," J. Acoust. Soc. Am. **116**, 2089–2100.

Shkerdin, G., and Glorieux, C. (**2005**). "Lamb mode conversion in an absorptive bi-layer with a delamination," J. Acoust. Soc. Am. **118**, 2253–2264.

Shorter, P. J. (**2004**). "Wave propagation and damping in linear viscoelastic laminates," J. Acoust. Soc. Am. **115**, 1917–1925.

Simonetti, F. (**2004**). "Lamb wave propagation in elastic plate coated with viscoelastic materials," J. Acoust. Soc. Am. **115**, 2041–2053.

Sun, Z. (**2004**). "Phased array focusing wave mechanics in tubular structures," Ph.D. thesis, The Pennsylvania State University.

Thomson, W. T. (**1950**). "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93.

Vogt, T., Lowe, M., and Cawley, P. (**2003**). "The scattering of guided waves in partly embedded cylindrical structures," J. Acoust. Soc. Am. **113**, 1258–1272.

Edge resonance in semi-infinite thick pipe: Numerical predictions and measurements

M. Ratassepp[a] and A. Klauson
Department of Mechanics, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia

F. Chati,[b] F. Léon, and G. Maze
Laboratoire d'Acoustique Ultrasonore et d'Electronique, UMR CNRS 6068, Université du Havre, Place Robert Schuman, BP 4006, 76610 Le Havre, France

(Received 29 December 2007; revised 16 May 2008; accepted 22 May 2008)

This paper presents theoretical and experimental studies of axisymmetric longitudinal guided wave $L(0,2)$ interaction with the free edge of the pipe. A numerical method based on normal mode superposition is applied to predict the edge resonance by an analysis of dispersion relations of separate modes. In parallel, the finite element analysis and experimental measurements prove the existence of edge resonance in the pipe in case of $L(0,2)$ wave incidence. It is shown that the edge resonance is mainly caused by the first pair of complex modes. Additionally the behavior of edge resonance phenomenon as a function of the curvature of the pipe is studied. The displacement amplitudes measured at the edge demonstrate that the edge resonance is affected by the frequency and thickness to midradius ratio of the pipe, and it is losing its strength in thicker pipes, as the growing difference between the outer and inner radii destroys symmetry. The reflected energy amplitudes show that at the resonance frequencies the incident wave is strongly converted to $L(0,1)$ and $L(0,3)$ modes, depending also on the curvature parameter of the pipe.
 [DOI: 10.1121/1.2945163]

PACS number(s): 43.40.Ey, 43.20.Ks, 43.40.At [RMW] Pages: 875–885

I. INTRODUCTION

The quantitative nondestructive evaluation of pipes using guided waves has moved toward the era of quick defect detection and size estimation.[1] Probably it is now, and will stay in the future, a challenging research area due to the huge number of practical interaction problems between guided wave and defects and the growing need to describe complicated wave propagation character in cylindrical structures. However, the scattering mechanism of the waves at defects is very complicated because the wave field near the defect is transformed and represents a diverse superposition of propagating and infinite number of nonpropagating modes. Resulting vibration depends on many parameters of the system such as the shape of the defect, the incident mode interacting the defect, and also the frequency of the excitation, and is therefore difficult to interpret. One phenomenon that appears at certain conditions of those parameters was already discovered a long time ago[2] but continues to be an intriguing and not fully understood fundamental problem. This is the resonance motion of the edge of different waveguides.

This problem has been thoroughly studied both theoretically and experimentally in plates[3–9] and solid cylinders.[10–13] Torvik[3] was supposedly the first to accurately describe this unusually strong localized motion in the vicinity of the plate edge in terms of nonpropagating modes. It was found that the edge resonance appeared as a result of the superposition of incident and reflected first symmetric Lamb mode S_0 and rapidly attenuating vibration modes that must satisfy the stress-free boundary condition at the end of the plate all at the same time. The first pair of nonpropagating modes caused a remarkable increase in the displacements of the plate end and clearly indicated the resonance phenomenon. A similar treatment can also be found in Zemanek's[12] work where a symmetric $L(0,1)$ mode in cylinders is used to study edge resonance. Thus it was realized that this phenomenon is not confined only to certain types of geometries of the waveguide but is inherent to structures that support symmetric dilatational vibration, and its behavior is driven by various parameters of the wave propagation medium. Therefore, researchers started later to investigate the resonance as a function of various parameters of the waveguide, attempting to explain the resonance phenomenon. Auld and Tsao[4] derived an efficient formula to predict edge resonance frequency for a plate by using a variational method, taking into account only a few trial field expressions. de Billy,[14] who studied rods of different cross sections, reported that the resonance frequency is not affected by the shape of the cross section but only by the value of the area of the waveguide. Grinchenko and Gorodetskaya[5] demonstrated that the form of the load at the end of the plate affects the strength of the edge mode. They found that the amplitude of the resonance could significantly increase when the load form chosen is consistent with the edge mode. In their experiment Le Clezio *et al.*[6] investigated the spatial nature of the edge mode near the end of the plate and observed the temporal nature of the resonance vibration. Besides, the *S*-parameter and the reciprocal work methods helped them to predict the phase shift of

[a]Electronic mail: madisr@staff.ttu.ee
[b]Electronic mail: farid.chati@univ-lehavre.fr

S_0 and the first pair of complex modes at resonance frequency and to show the dependence between Poisson ratio and the resonance amplitude as well as resonance frequency. Wilkie-Chancellier *et al.*[7] showed the influence of the oblique angle of the end side of the plate on the generation of the edge mode. A surprisingly small deviation of the end from the right angle led the incident S_0 mode to a complete conversion into antisymmetric A_0 and A_1 modes. Pagneux[8] focused on the coupling conditions between the propagating and nonpropagating waves in the case of the resonance in plates. Using the concept of complex resonance, he derived a simple empirical formula for the prediction of resonance frequency and proved the decoupling between real and complex modes mathematically for the particular nonzero Poisson ratio value. A similar study by Zernov *et al.*[9] showed that at specific values of Poisson ratio the edge mode is undamped and the energy leakage of the resonance mode to the plate is associated with the imaginary part of its complex eigenvalue.

However, only a few works[15–17] deal with the explanation of the resonance phenomenon in pipelike structures. In Refs. 15 and 16 authors used the theory of shells, which restricted the analysis to thin-walled structures. Grinchenko and Komissarova[17] investigated the edge resonance of the finite pipe when the first longitudinal mode is incident. The aim of the present study is to extend the analysis of edge resonance to thick pipes and additionally to investigate the influence of the curvature on axisymmetric longitudinal wave propagation characteristics in a semi-infinite pipe. On the basis of linear theory of elasticity the properties of different wave modes (propagating and nonpropagating) are expressed in the form of frequency dependent dispersion curves, presenting phase velocities and wave numbers for different thickness to inner radius ratios. The interaction of the incident mode $L(0,2)$ with the edge of the pipe is solved numerically following the approach of Torvik and others.[3,22,23] The influence of the curvature on the excitation of edge mode in the case of incident $L(0,2)$ is thoroughly investigated, and the results are compared with the results of the finite element (FE) method. Finally, an experimental determination of edge resonance is made in an aluminium pipe in order to verify the computational predictions.

The outline of the paper is as follows. In Sec. II, the computational scheme for complex axisymmetric modes and for propagating mode interaction with the edge is described and FE modeling procedure is given. To help in the reading of the paper the computational details are given in the Appendixes. In Sec. III the experimental setup is described. In Sec. IV the edge mode in a pipe is verified experimentally. Also some results are presented to show the effect of the curvature to the edge resonance phenomenon. It is shown that decreasing the curvature radius of the pipe (or increasing the thickness), the resonance loses its strength and the incident wave is mode converted at the edge in pipes with a certain curvature. Concluding remarks can be found in Sec. V.

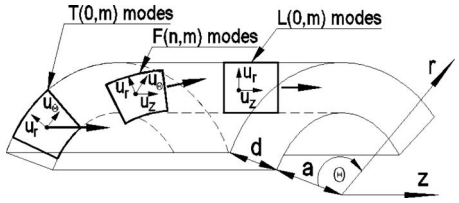


FIG. 1. Geometry of the problem and the representation of the wave families in the pipe.

II. MODELING PROCEDURES

A. Dispersion equation for axisymmetric longitudinal waves

The motion of a thick pipe can be modeled by using the elasticity theory approach developed by Gazis.[18] This approach allows the vibration of the pipe to be separated into three decoupled wave families, identified as axisymmetric longitudinal, axisymmetric torsional, and nonaxisymmetric flexural modes. Later, after Meitzler[19] we know these modes as $L(0,m)$, $T(0,m)$, and $F(n,m)$, respectively. The first index n indicates the circumferential order, and the second number m is a counter in order of appearance of different modes in one wave family. These modes are shown in Fig. 1 by their displacement components. It is known that axisymmetric longitudinal modes $L(0,m)$ are the counterparts of Lamb modes in a plate if the thickness of a pipe $d \ll a$, where a is the inner radius of a pipe. Therefore, the edge resonance must exist for a certain mode in the $L(0,m)$ wave family with similar propagation characteristics as the plate membrane mode S_0.[6] However, there is not much information about the vibration nature of a thick pipe $d \sim a$; moreover, the nature of formation of the edge mode in this structure is not quite clear. Subsequently we present the steps taken to familiarize oneself with the wave propagation character in thick pipes.

Traditionally this involves the analysis of the dispersion characteristics of the wave modes as a function of geometrical and material parameters of the pipe. These can be examined after solving the characteristic dispersion equation [a brief recall of the dispersion equation, displacements, and stresses for $L(0,m)$ modes is provided in Appendixes A and B]

$$\mathrm{Det}|G(c_L,c_T,a,d,f)| = 0.$$

Here, G is the matrix that gathers the boundary conditions of the pipe, c_L and c_T are velocities of longitudinal and transverse waves of the medium, and f is a frequency. In general, the roots of the dispersion equation can be real, imaginary, or complex. The properties of real (propagating) wave modes are well known, and they have been exploited in many applications.[20,21] Although several investigations have been performed on the explanation of imaginary and complex root (nonpropagating) wave modes in plates[7,22] and rods,[11] the nature of a complex frequency spectrum of the wave propagation in a pipe is not thoroughly interpreted.

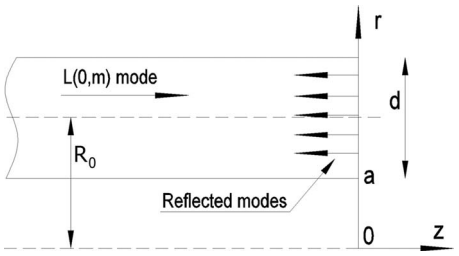


FIG. 2. $L(0,m)$ mode reflection at the edge of the pipe wall.

B. Computation of the reflection at the end of a semi-infinite pipe

Consider a two-dimensional axial cross section of a semi-infinite pipe wall as shown in Fig. 2. An incident mode $L(0,m)$ propagating in the $z>0$ direction interacts with the free end of the pipe at $z=0$. In order to characterize the wave field and the reflections at the free edge, we apply the normal mode expansion technique, also called the modal decomposition method.[3,22,23] A short treatment applied to the current problem is given in Appendix C.

In this method an arbitrary acoustic field of the waveguide structure is developed into an expansion of infinite series of wave modes, which must satisfy the boundary condition of the wave propagation medium. The reflected field near the edge of the pipe contains a finite number of propagating and infinite number of nonpropagating modes at any given frequency. The advantage of this method is that it is possible to estimate the contribution of each wave mode in this expansion. However, in order to obtain the amplitudes of the modes under interest, the linear system of an infinite number of boundary condition equations must be solved, which is obviously not possible. Previous studies[22] have shown that the solutions of reasonable accuracy can be obtained after making the system finite, considering a finite number of propagating and low order nonpropagating modes. The accuracy of the calculations can be checked by applying the energy conservation concept; the energy carried by the reflected propagating modes must be as close as possible to the energy of incident mode. Nonpropagating modes generated on the edge do not transmit energy along the pipe, and they attenuate with the distance from the edge.

C. Finite element study

The edge resonance has been successfully studied with the FE method in plates in case of S_0 incidence.[6] A similar study was performed to predict curvature effect on edge vibration of the pipe in case of incident $L(0,2)$ mode using the explicit procedure of program ABAQUS.[25] Due to the axial symmetry of the problem, a two-dimensional region representing a radial-axial section through the pipe was modeled, as shown in Fig. 3(a). The thickness of the wall of the pipe was 2.2 mm and the length varied from 250 to 750 mm. The mesh of the pipe consists of four noded linear quadrilateral axisymmetric elements with two degrees of freedom in each node (displacements in r and z directions). These elements satisfactorily describe the motion of axisymmetric longitudinal pipe modes, as was shown in previous studies.[26,27] On one side of the pipe the absorbing region[28] is applied to decrease the model size and neglect undesired reflections from the edge. The waves that enter this area are increasingly damped and eventually die out.

A number of geometries were set up in order to model the pipes with different thickness d to midradius R_0 ratios defined as

$$\Delta = d/R_0 \quad (R_0 = a + d/2),$$

so that $\Delta=2$ in the case of a solid cylinder. The thickness of the pipe always remained the same; only the position of the axis was modified to model the pipe with the desired radius. The pipes with the following geometries were studied: $\Delta=0.25$, 0.5, 0.75, and 1. In each model the density of the mesh was also changed according to the wavelength of propagating modes. At least 15 elements per wavelength were used, which is more than the lower limit of spatial discretization of eight elements per wavelength for accurate modeling. A summary of the FE models that were used in the study can be found in Table I.

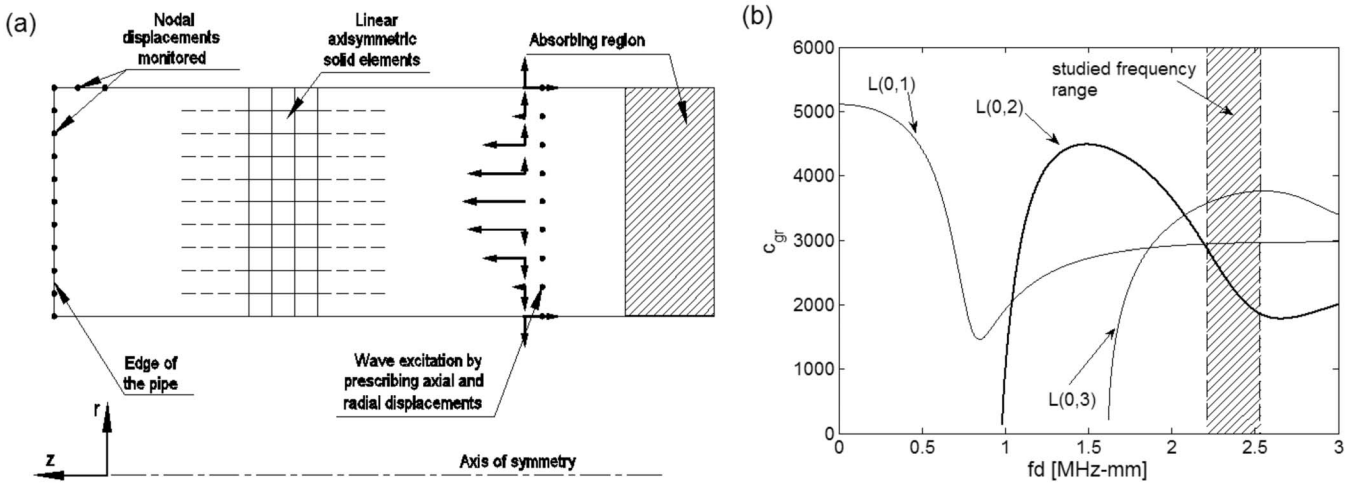


FIG. 3. (a) FE axisymmetric model of the pipe. (b) Group velocity dispersion curves for axisymmetric longitudinal modes propagating in an aluminium pipe; the curvature parameter is $\Delta=1$.

TABLE I. Summary of FE models used in the study.

Δ	Length (mm)	Elements through thickness	Tone burst (cycles)
0.25	250	15	20
0.5	350	15	40
0.75	450	20	50
1.0	750	20	60

The signal was generated 20 mm away from the edge of the pipe where the absorbing region was applied. The excitation of the $L(0,2)$ mode was achieved by "center mode shape" excitation technique,[29] which is used for pure mode generation in the nondispersive regime. The desired mode in the FE model was generated by scaling the tone bursts applied to each node through thickness according to the amplitude of displacement at that location in the exact mode shape profile at a center frequency of excitation. However, the group velocity curves in Fig. 3(b) show that it is not possible to generate an entirely pure mode with this technique because over the investigated frequency bandwidth the excited mode $L(0,2)$ is dispersive and the other lower order axisymmetric modes $L(0,1)$ and $L(0,3)$ are also generated. To avoid any other modes interfering with the results, a sufficiently long propagation distance was chosen to make the signals separable.

In each pipe model a different tone burst was used. The narrow band signals consisting of a tone burst multiplied by sinusoid window containing 20, 40, 50, and 60 cycles centered at resonance frequency were used for the excitation. The higher number of cycles was used in models where the wave propagation distance was longer and the reduction of distortion of the wave packet was needed. The resonance of the edge was supposed to be excited at a specific frequency in the frequency spectrum of the excitation. The rough estimation of this frequency was achieved by using the numerical modeling results, and the calculation was repeated with adjusted frequency.

The results of the simulations were obtained by monitoring nodal displacements at the free edge and close to the edge on the surface of the pipe. Both radial and axial displacement components were monitored to describe the edge mode. An illustrative time record of radial displacement of the outer corner of the edge measured at resonance frequency is plotted in Fig. 4(a), showing permanent high amplitude vibration of the edge. The resonance frequency was measured as that corresponding to the maximum value from the frequency spectrum, shown in Fig. 4(b).

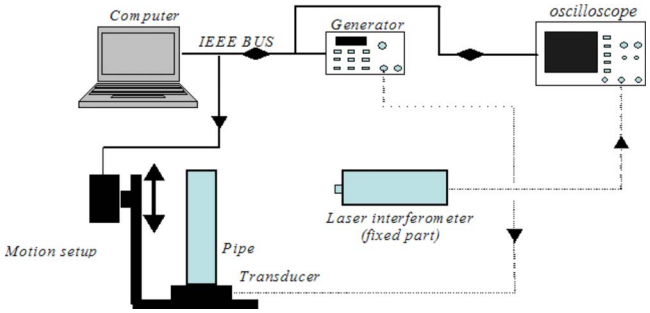


FIG. 5. (Color online) Scheme of the experimental setup.

III. EXPERIMENTAL PROCEDURE

The experimental setup is described in Fig. 5. Measurements were performed on an aluminium pipe of 230 mm length, with an inner radius of 7.85 mm and a wall thickness of $2.2^{\pm 0.1}$ mm ($\Delta \approx 0.25$). The pipe was vertically posed to the transducer. A metalscan gel layer was used to ensure a good ultrasound coupling. The set is vertically mobile by means of a motion setup, which allows measuring the displacement from different positions on the pipe surface.

Excitation of the $L(0,2)$ mode was achieved using the broadband piezoelectric transducer (Panametric V401) with a central frequency of 1 MHz. The transducer was excited using a narrow band signal consisting of a tone burst multiplied by a sinusoid window containing 20 cycles using the signal generator (3314 Generator). The detection of the signal was achieved using a laser interferometer (BMI heterodyne probe SH140) to measure the normal displacement on the surface of the tube. The measurements were taken at a series of

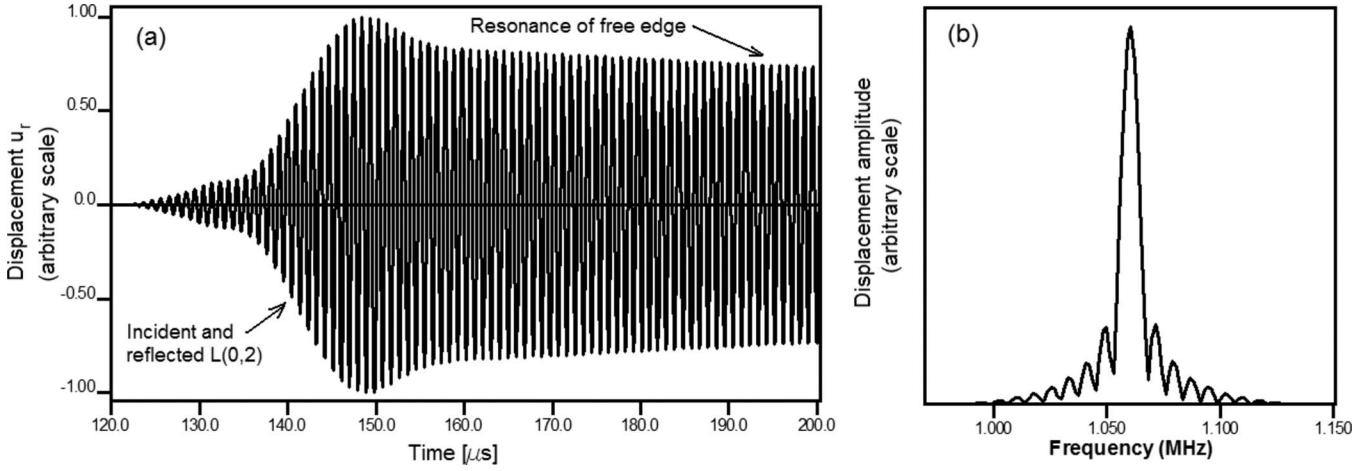


FIG. 4. (a) Typical time record showing the edge resonance in the case of $L(0,2)$; the center frequency-thickness product is 2.335 MHz mm. (b) Frequency spectrum of the signal shown in (a).

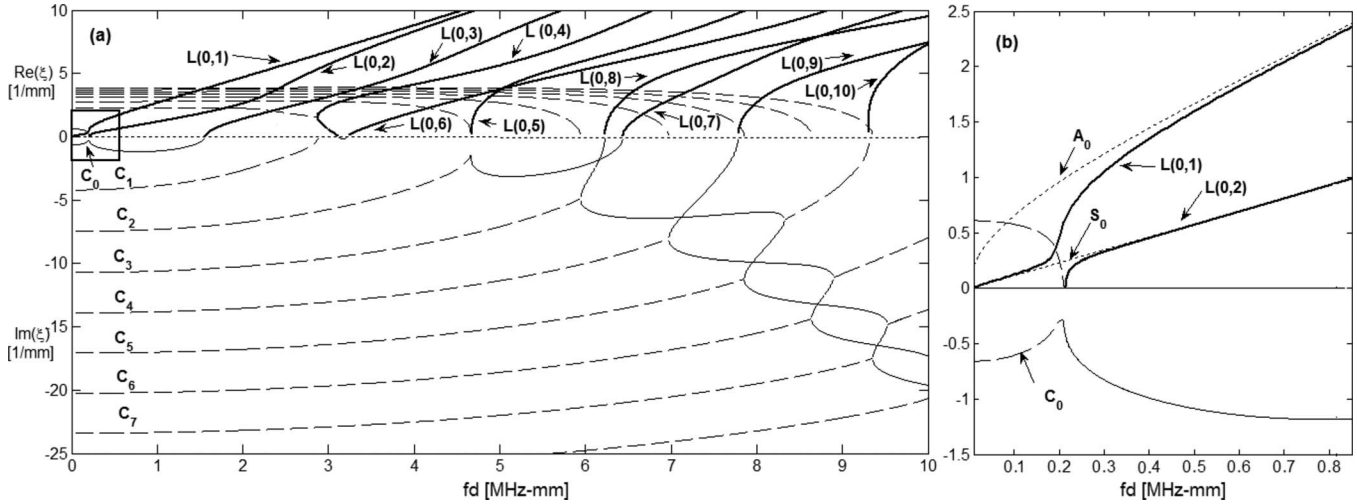


FIG. 6. (a) Axial wave number ξ dispersion curves for a pipe. (Bold solid line) real modes, (solid line) imaginary modes, and (dashed line) complex modes. (b) Zoomed area in a box from part (a). The curvature parameter is $\Delta=0.25$.

equally spaced positions along the pipe from 0.1 to 16.7 mm with the step 0.1 mm at the upper end. The measured signals by the laser interferometer were averaged (sweep average: 200) and displayed on a digital oscilloscope. Thereafter the obtained signal was recorded on a computer via the IEEE bus in order to realize the numerical treatments. This computer also allowed us to drive and to control the motion of the pipe.

IV. RESULTS AND DISCUSSION

This section presents wave propagation characteristics in the pipe as the function of curvature. The results are verified for various cases from the literature, and the existence of the edge resonance is proven experimentally. The main objective is to show how axisymmetric wave propagation and edge resonance pheonomenon is influenced by the curvature effect of the pipe.

A. Axisymmetric wave propagation in a thick pipe

For the modeling of guided wave interaction with discontinuities in the pipe, first their propagation character must be understood in an ideal undamaged pipe. A set of computations have been made for aluminium pipe as used in experiment (c_L=6440 m/s, c_T=3113 m/s, ρ=2765 kg/m^3, and d=2.2 mm).

Figure 6(a) represents a typical plot of wave numbers ξ for various possible propagating and nonpropagating modes as a function of frequency-thickness product fd in the pipes. The propagating modes are labeled as $L(0,1)$, $L(0,2)$,…. The nonpropagating modes with complex wave numbers are denoted as C_0, C_1, C_2,…, and their spatial attenuation along the propagation is characterized by Im(ξ). The behavior of these modes is very similar to Lamb modes in a plate, except in the region of low frequency where the mode with long wavelengths are curvature dependent. This is shown in Fig. 6(b), where the deviation of the lower order pipe modes $L(0,1)$ and $L(0,2)$ from Lamb modes A_0 (flexural mode) and S_0 (membrane mode) is in. The curve of the mode $L(0,1)$ slowly separates from Lamb mode A_0 as fd product decreases and finally tends toward the S_0 mode, but these two curves remain separate, as will be seen later. The mode $L(0,2)$, which is similar to the S_0 mode in plate, vanishes below the cutoff frequency fd=0.21 MHz mm, and the complex branch C_0 appears. The curvature effect deepens when curvature radius decreases. This is shown in Fig. 7, where the phase velocity curves are presented for different thickness to middle radius ratios Δ. For example, the decrease in the inner radius a toward zero of the pipe causes the bending type mode $L(0,1)$ to vibrate as the compressional type mode of a solid cylinder [Fig. 7(a)], as predicted by Nishino *et al.*[21]

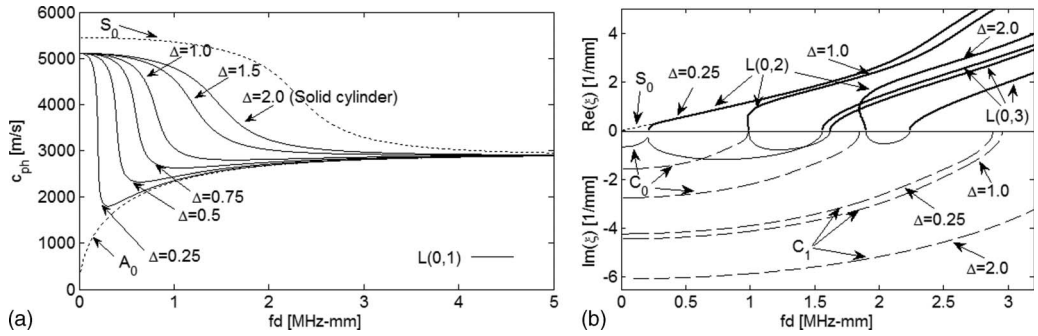


FIG. 7. The dependence of the modes on the parameter Δ. (a) Phase velocity curves for the $L(0,1)$ mode. (b) Wave number curves for $L(0,2)$ and $L(0,3)$ modes and attenuation curves for nonpropagating modes.

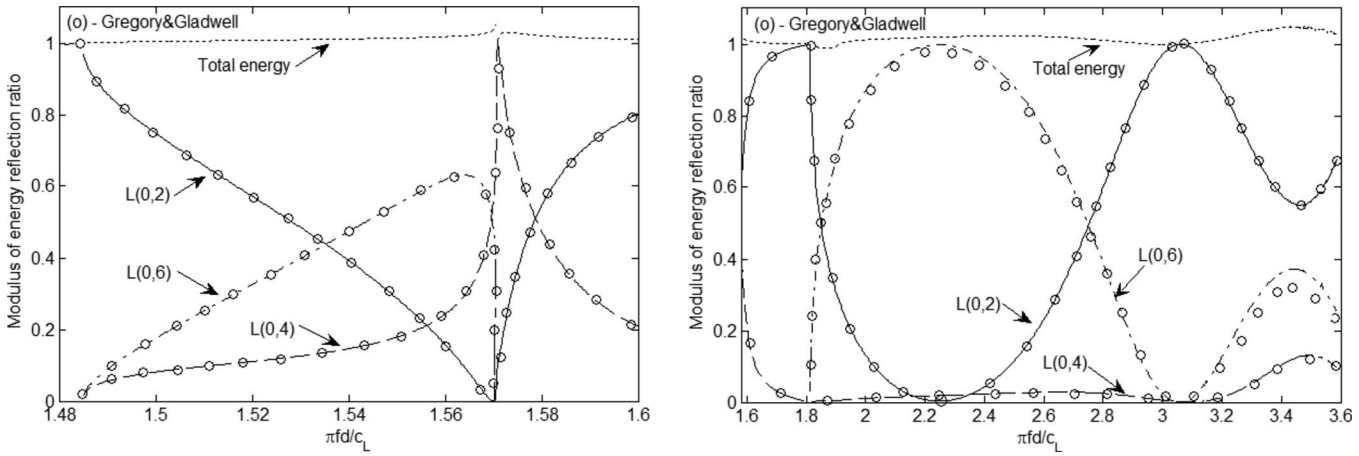


FIG. 8. Reflection of the $L(0,2)$ mode at a pipe edge. The parameters of the pipe: $a=20$ mm, $d=1$ mm, $\rho=7800$ kg/m^3, $c_L=1000$ m/s, and $c_T=577.4$ m/s.

In Fig. 7(b) the wave number curves for $L(0,2)$ and $L(0,3)$ modes and attenuation rates for a few nonpropagating modes are shown. When $\Delta=0.25$, the mode $L(0,2)$ clearly overlays the symmetric plate mode S_0 for almost all frequencies, but when the curvature radius of the pipe decreases, the shift in the cutoff frequencies of this mode toward higher frequencies and an increase in the decaying rate of nonpropagating modes C_0 and C_1 can be seen. Modes $L(0,1)$ and $L(0,4)$ are omitted here for clarity.

B. Reflection of the $L(0,2)$ mode by a free end: Validation of numerical results

From previous research with plates[3] it is well known that the edge resonance is produced by Lamb mode S_0 at a particular frequency and is due to the high amplitude standing waves raised by complex modes at the end of the plate. In the previous section it was shown that the propagation of the $L(0,2)$ mode in a thin-walled pipe is similar to Lamb mode S_0 propagation in a plate. It means that the behavior of the reflection of the $L(0,2)$ mode at the edge must be analogous. This is shown in Fig. 8, where the reflection ratio of this mode is observed as a function of frequency. A similar incident energy transformation into energies of higher order modes is seen, and was observed in the plate by Gregory and Gladwell.[30] Plate results are shown by dots, which fit very well the curves corresponding to the pipe results. Therefore it can be expected that the edge resonance in thin-walled pipes appears at the same frequency and is mainly due to the oscillation of the first complex mode C_1. This statement is confirmed in Fig. 9, where the displacement amplitudes of this mode clearly exceeds those of the incident wave and other higher order complex modes near the resonance frequency $fd=2.317$ MHz mm. Again there is a good agreement with previous results of Wilkie-Chancellier *et al.*[7] The accuracy of the numerical results was estimated by the concept of energy conservation, and the relative error was found to be less than 5%.

C. Experimental detection of the edge resonance

Due to the spread in thickness of the pipe, an effort should be made to determine the frequency of the edge resonance accurately. Initially, a resonance spectrum was measured at 0.1 mm from the upper end of the pipe in the frequency-thickness product range of 2.2–2.53 MHz mm with a step of 0.011 MHz mm. Figure 10(a) shows the variation of the amplitude for the normal surface displacement u_r depending on the fd. This untreated spectrum was filtered by means of a recursive filter in order to smooth the curve, and the edge resonance frequency was determined at the maximum of the measured magnitude of the resonance spectrum at $fd=2.435$ MHz mm. This value differs from the theoretically obtained result $fd=2.336$ MHz mm. However, the nonuniform thickness of the pipe used in the test allows for expanding the resonance frequency range to $fd=(2.324,2.565)$ MHz mm. Thereafter the spatiotemporal representation at the resonance frequency was performed, as shown in Fig. 10(b). Here it is clearly seen that after the reflection of the incident mode with the end (140 μs), the edge remains vibrating with high amplitude, which is the

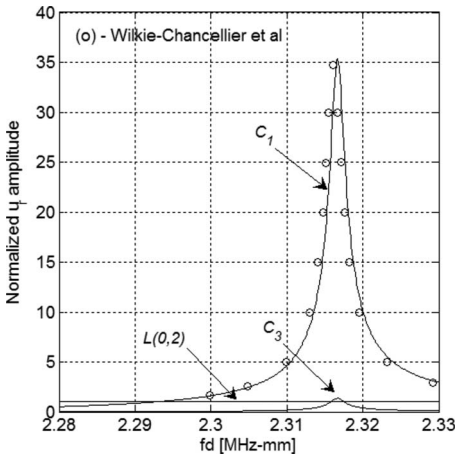


FIG. 9. Normalized displacement u_r amplitudes of reflected modes at the edge of the pipe. The parameters of the pipe: $a=40$ mm, $d=1$ mm, $\rho=7800$ kg/m^3, $c_L=5850$ m/s, and $c_T=3150$ m/s.

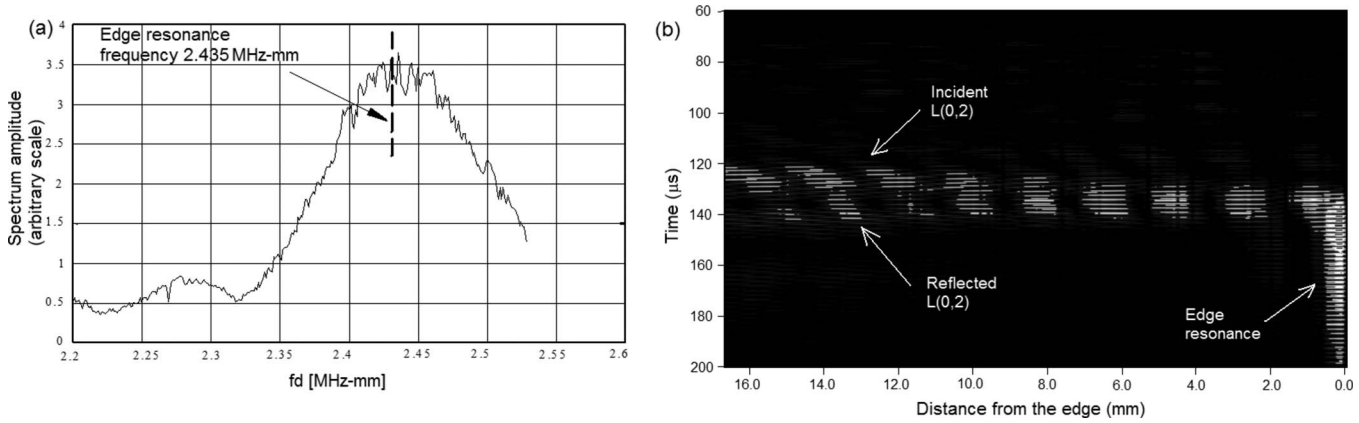


FIG. 10. (a) Unfiltered resonance spectrum of normal displacement of the $L(0,2)$ mode measured at 0.1 mm from the pipe end. (b) Contour plot of measured normal displacements in time and space domain showing the generation of edge resonance.

resonance behavior. Figure 11 shows the time domain record measured at 0.1 mm from the pipe edge. Long tail due to edge resonance can be clearly seen in this plot.

D. Influence of the curvature on the edge resonance: Mode conversion

The variation of the curvature parameter Δ strongly affects the interaction of the $L(0,2)$ mode with the end of the pipe and the generation mechanism of the edge resonance. The edge resonance can be found at a maximum displacement amplitude value of complex mode C_1 versus frequency-thickness product fd. An illustrative calculation was performed for the pipes with curvature parameters $\Delta=0.25$, 0.5, 0.75, and 1.0, as shown in Fig. 12, where the normalized radial displacement amplitude of C_1 at the edge is shown as a function of fd. A normalized amplitude is the ratio of the largest amplitude of complex mode to the largest amplitude of the $L(0,2)$ incident wave anywhere in the thickness of the pipe wall. As seen from the figure, the amplitude of the complex mode drops when the parameter Δ increases and the peaks move toward higher fd. This behavior can be explained by the increasing difference of the inner and outer radii in a thick pipe, by which the symmetry is broken. The edge resonance frequencies obtained with numerical and FE models are shown in Table II. The frequency values obtained by the FE method tend to be always smaller than those obtained using the numerical model.

To see the extent of the edge resonance along the pipe, the radial displacement of outer surface near the end for the two different pipes $\Delta=0.25$ and 0.75 is calculated by the finite element model and by the numerical approach, shown in Fig. 13. The finite element data represent the amplitudes of the displacements measured in frequency domain at resonance frequency. A rapid decrease in amplitudes of the total displacement field is seen moving away from the edge. There is a good agreement between the results using these two methods.

The curvature parameter also changes the through-thickness displacement variations of the edge. Normalized axial and radial displacements at the resonance frequencies are shown in Figs. 14(a)–14(c) for the pipes with curvature parameters $\Delta=0.25$ and 0.75, and 1.0. The axial displacement of the pipe ($\Delta=0.25$) edge in Fig. 14(a) is nearly symmetric to the midsurface of the pipe wall, which is similar to the plate case. However, when the curvature parameter Δ increases, this symmetry is broken and the inner surface of the pipe vibrates more intensively than the outer surface.

Again there is a good agreement between the numerical and finite element results, except for the thick pipe $\Delta=1.0$ in Fig. 14(c), where some discrepancy between the displacements of the two methods can be seen. This can be caused by the difference in resonance frequencies used in calculations by the two approaches. As at $\Delta=0.25$ and 0.75 the edge

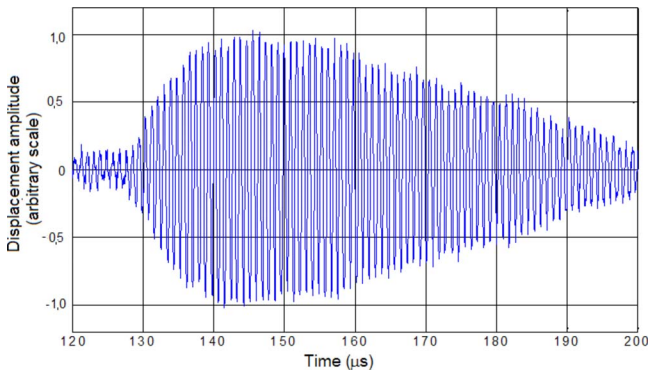


FIG. 11. (Color online) Normal surface displacement u_r time record at 0.1 mm from the pipe end at $fd=2.435$ MHz mm.

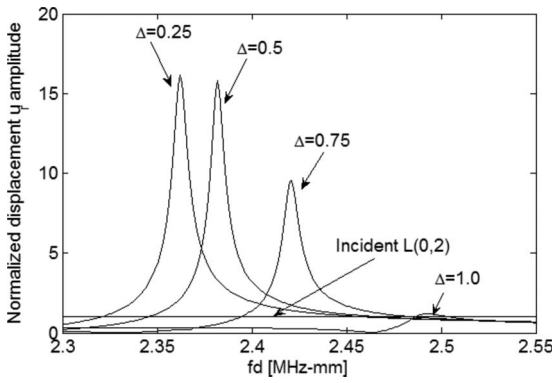


FIG. 12. Normalized displacement u_r amplitudes of reflected modes at the edge of the pipe as a function of curvature parameter Δ.

TABLE II. Predicted edge resonance frequencies by numerical and FE model.

Δ	fd_{num} (MHz mm)	fd_{FE} (MHz mm)
0.25	2.362	2.335
0.5	2.383	2.352
0.75	2.420	2.400
1.0	2.493	2.468

vibration is mainly dominated by the mode shape of the C_1 (Fig. 12). Then at $\Delta=1.0$ the displacements are comparable with the ones of the incident mode.

It is also interesting to see the energy balance at the edge of the pipe as a function of curvature parameter Δ. In Fig. 15 it is seen that when Δ is close to zero (very thin shell or plate), the energy is entirely reflected into the $L(0,2)$ mode. This is also true for big curvature parameter values ($\Delta=2$ for a solid cylinder). Between these extreme values of Δ, reflection of the $L(0,2)$ mode generates antisymmetric modes. The energy ratio of the $L(0,2)$ mode is reduced to zero at $\Delta=0.667$ (pipe thickness nearly equal to its inner radius), where the reflected energy is completely transferred into $L(0,1)$ and $L(0,3)$ modes at the resonance frequency, as seen in Fig. 16, where energy balance of modes versus the fd product is presented.

V. CONCLUSIONS

The edge resonance phenomenon in a semi-infinite pipe was investigated. The scattering of axisymmetric longitudinal wave mode $L(0,2)$ at the edge was modeled by applying the normal mode expansion technique. This method allows us to easily estimate the contribution of each wave mode in the expansion. The accuracy of the calculations can be checked by the energy conservation concept application. The edge resonance frequency of the pipe can be found by the maximum in the first complex mode C_1 spectrum. The phenomenon in the thin-walled pipe is similar to that of the plate. However, curvature influence cannot be neglected for thick pipes with a thickness to medium radius ratio of $\Delta>0.5$. In this case through-thickness displacement distribution is no more symmetric with respect to the middle surface of the pipe, and the results will increasingly differ from the plate ones. The study of the curvature effect on wave propagation showed that the edge resonance in case of $L(0,2)$

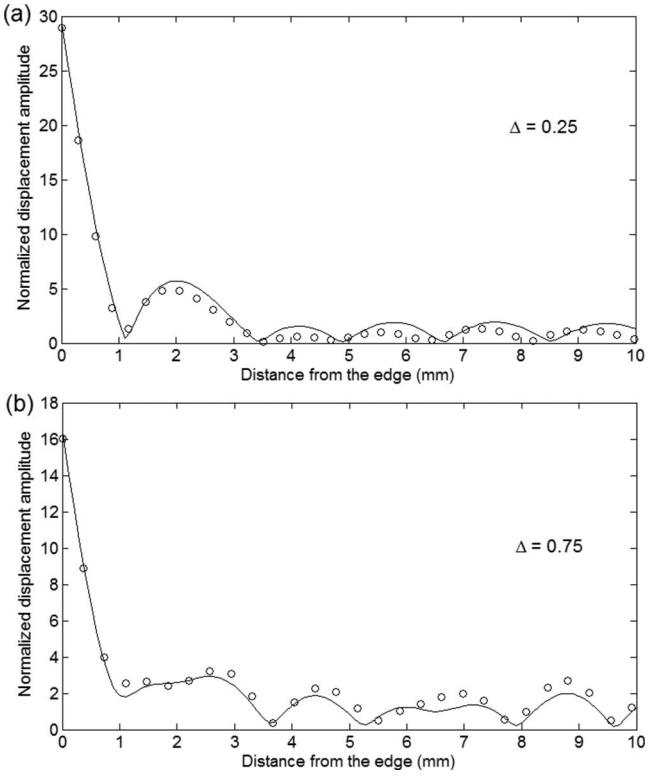


FIG. 13. The normal displacement component of the total displacement field at the outer surface measured as the function of distance near the edge at resonance frequency. (—) numerical model; (○○○) FE predictions. (a) $\Delta=0.25$; (b) $\Delta=0.75$. The results have been normalized by the outer surface radial displacement amplitude of the incident mode.

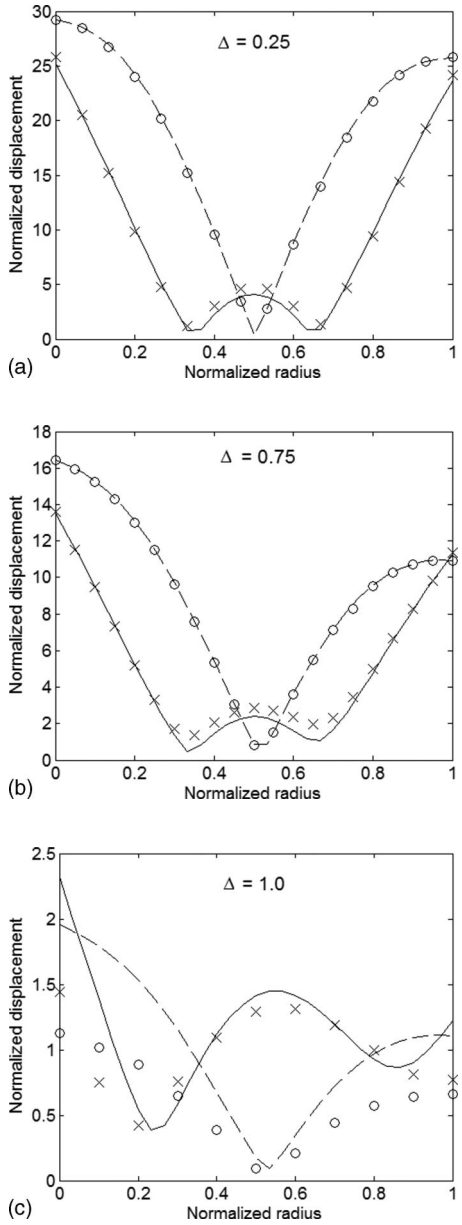


FIG. 14. Through-thickness mode shapes of the edge at resonance frequency for different pipes: (a) $\Delta=0.25$, (b) $\Delta=0.75$, and (c) $\Delta=1.0$. Numerical predictions u_r (dashed line), u_z (solid line); FE predictions u_r (○○○), u_z (×××). The extremities of the normalized radius scale represent the inner and outer radii of the pipe.

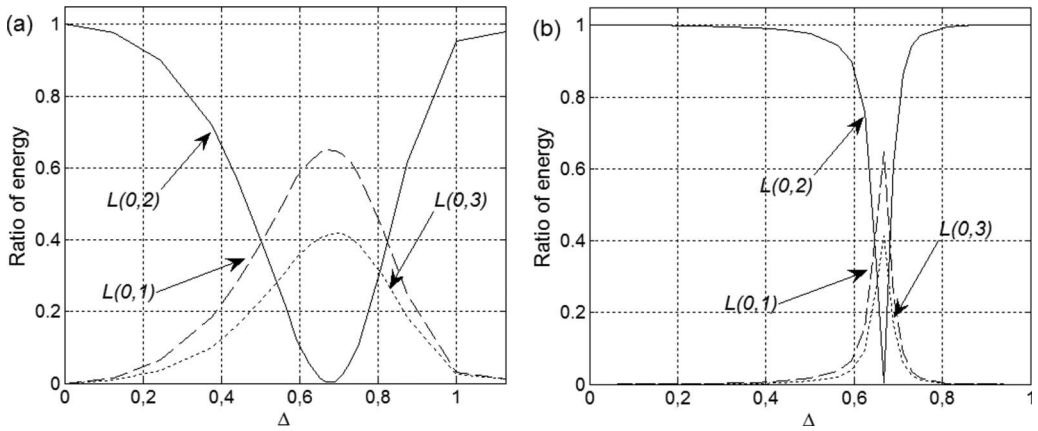


FIG. 15. Energy balance of the $L(0,2)$ mode as a function of curvature parameter Δ. (a) The energies for the modes have been obtained at resonance frequencies. (b) All energies have been calculated at $fd=2.405$ MHz mm.

weakens when curvature radius decreases, and the resonance frequency shifts to higher values. Interestingly at $\Delta=0.667$ the $L(0,2)$ wave is completely converted into antisymmetric $L(0,1)$ and $L(0,3)$ waves at the resonance frequency. The edge resonance of the pipe was experimentally detected by observing the displacement field near the edge caused by incident $L(0,2)$ mode. Both FE and experimental studies confirmed the existence of the edge resonance in pipes and were in a good agreement with the theory.

ACKNOWLEDGMENTS

This research was supported by Estonian Science Foundation Grant No. 6169 and by the Estonian Ministry of Education and Science Grant No. SF0140072s08. We would like to thank Professor Peter Cawley and Professor Mike Lowe from the NDT Group at Imperial College NDT Lab for their assistance.

APPENDIX A: DISPERSION EQUATION

For a detailed analysis, the reader is referred to Refs. 18 and 31. The dispersion frequency equation for axisymmetric longitudinal modes in a pipe is

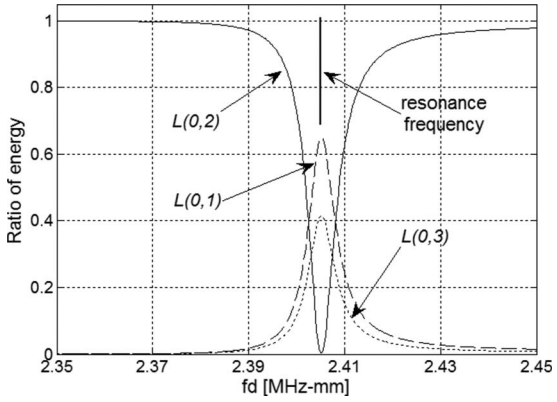


FIG. 16. Energy balance of the $L(0,2)$ mode as a function of frequency-thickness fd for the pipe of $\Delta=0.667$.

$$GC=\begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix}=0, \tag{A1}$$

where

$$G_{11}=-(\beta^2-\xi^2)a^2J_0(\alpha a)+2\alpha a J_1(\alpha a),$$

$$G_{12}=2\xi\beta a^2J_0(\beta a)-2\xi a J_1(\beta a),$$

$$G_{13}=-(\beta^2-\xi^2)a^2N_0(\alpha a)+2\alpha a N_1(\alpha a),$$

$$G_{14}=2\xi\beta a^2N_0(\beta a)-2\xi a N_1(\beta a),$$

$$G_{21}=2\xi\alpha a^2J_1(\alpha a),$$

$$G_{22}=(\beta^2-\xi^2)a^2J_1(\beta a),$$

$$G_{23}=2\xi\alpha a^2N_1(\alpha a),$$

$$G_{24}=(\beta^2-\xi^2)a^2N_1(\beta a). \tag{A2}$$

Here

$$\alpha^2=\omega^2/c_L^2-\xi^2,\quad \beta^2=\omega^2/c_T^2-\xi^2, \tag{A3}$$

where c_L and c_T are velocities of compressional and shear waves of the medium, $\omega=2\pi f$ is the ring frequency, ξ is the axial wave number, and J and N are Bessel functions. The remaining two rows of the system [Eq. (A1)] are obtained from the first two by the substitution of b (outer radius) for a (inner radius). To obtain nontrivial eigensolutions ξ, the characteristic equation should be

$$\mathrm{Det}[G]=0. \tag{A4}$$

Here a robust root finding routine has been developed, which is based on finding the minima in Eq. (A4) with the Newton–Raphson method. However, some difficulties were experienced while tracing complex roots when using Bessel functions with the complex argument in the solution. Numerically stable solutions were obtained after replacing Bessel functions with Hankel functions, which were calcu-

lated with the routines developed by Thompson and Barnett.[24]

APPENDIX B: DISPLACEMENTS AND STRESSES

Using notations

$$f(r) = C_1 J_0(\alpha r) + C_3 N_0(\alpha r), \tag{B1}$$

$$g(r) = C_2 J_1(\beta r) + C_4 N_1(\beta r), \tag{B2}$$

the displacements and stresses are

$$u_r = (\partial f/\partial r + \xi g) e^{i(\xi z - \omega t)},$$

$$u_z = i(\xi f + g/r + \partial g/\partial r) e^{i(\xi z - \omega t)}, \tag{B3}$$

$$\sigma_{rr} = [-\lambda(\alpha^2 + \xi^2) f + 2\mu(\partial^2 f/\partial r^2 + \xi \partial g/\partial r)] e^{i(\xi z - \omega t)},$$

$$\sigma_{rz} = i\mu[2\xi \partial f/\partial r + (\xi^2 - \beta^2) g] e^{i(\xi z - \omega t)},$$

$$\sigma_{zz} = [-\lambda(\alpha^2 + \xi^2) f - 2\mu\xi(\xi f + g/r + \partial g/\partial r)] e^{i(\xi z - \omega t)}, \tag{B4}$$

where λ and μ are Lamé's constants.

APPENDIX C: REFLECTION COEFFICIENTS

For a detailed analysis, the reader is referred to Refs. 22 and 23. In a normal mode superposition technique an acoustic field can be developed as an expansion of vibration modes of the structure. Thus, boundary conditions of the inhomogeneity can be written as a sum of vibration mode field elements. In the case of a stress-free edge of the pipe the normal and shear stresses σ_{zz} and σ_{rz} must vanish,

$$\sigma_{zz}(r) = \sum_{j=0}^{\infty} a_j \sigma_{zzj}(r) = 0,$$

$$\sigma_{rz}(r) = \sum_{j=0}^{\infty} a_j \sigma_{rzj}(r) = 0, \tag{C1}$$

where a_j represents the amplitude coefficient for the jth wave mode. This system can be solved approximately by collocation or least squares method accounting for a finite number of equations in [Eq. (B1)]. This is achieved by setting σ_{zz} and σ_{rz} equal to zero at a fixed number k of points along the edge of the pipe, thus forming a linear equation system. The energy carried by the jth reflected wave is expressed as

$$\Psi_j = a_j \bar{a}_j \psi_j, \tag{C2}$$

where $\bar{a}_j$ denotes the complex conjugate. In this expression, ψ_j is the acoustic Poynting vector flow of the jth wave mode,

$$\psi_j = -\tfrac{1}{2} \mathrm{Re}\left[-i\omega \int \int (\bar{u}_{rj} \sigma_{rzj} + \bar{u}_{zj} \sigma_{zzj}) d\theta dr \right]. \tag{C3}$$

The propagating wave energy reflection coefficient R_j can be calculated by dividing the energy of the reflected wave Ψ_j by the incident wave energy Ψ_0,

$$R_j = \frac{\Psi_j}{\Psi_0}. \tag{C4}$$

In the case of nonpropagating and inhomogeneous wave modes, the Poynting vector flow ψ_j is zero and all the energy is retained in propagating modes. Therefore the energy conservation concept can be applied to check the computational accuracy,

$$\Psi_0 = \sum_j \Psi_j \text{ or } \sum_j R_j = 1. \tag{C5}$$

[1] A. Demma, P. Cawley, M. J. S. Lowe, A. G. Roosenbrand, and B. Pavlakovic, "The reflection of guided waves from notches in pipes: A guide for interpreting corrosion measurements," NDT & E Int. **37**, 167–180 (2004).
[2] E. A. G. Shaw, "On the resonant vibrations of thick barium titanate disks," J. Acoust. Soc. Am. **28**, 38–50 (1956).
[3] P. J. Torvik, "Reflection of wave trains in semi-infinite plates," J. Acoust. Soc. Am. **41**, 346–353 (1967).
[4] B. A. Auld and M. T. Tsao, "A variational analysis of edge resonance in a semi-infinite plate," IEEE Trans. Sonics Ultrason. **SU-24**, 317–326 (1977).
[5] V. T. Grinchenko and N. S. Gorodetskaya, "Excitation of the edge mode in an elastic half strip," Int. Appl. Mech. **34**, 115–122 (1998).
[6] E. Le Clezio, M. V. Predoi, M. Castaings, B. Hosten, and M. Rousseau, "Numerical predictions and experiments on the free-plate edge mode," Ultrasonics **41**, 25–40 (2003).
[7] N. Wilkie-Chancellier, H. Duflo, A. Tinel, and J. Duclos, "Numerical description of the edge mode at the beveled extremity of a plate," J. Acoust. Soc. Am. **117**, 194–199 (2005).
[8] V. Pagneux, "Revisiting the edge resonance for Lamb waves in a semi-infinite plate," J. Acoust. Soc. Am. **120**, 649–656 (2006).
[9] V. Zernov, A. V. Pichugin, and J. Kaplunov, "Eigenvalue of a semi-infinite elastic strip," Proc. R. Soc. London, Ser. A **246**, 1255–1270 (2006).
[10] J. Oliver, "Elastic wave dispersion in a cylindrical rod by a wide-band short duration pulse technique," J. Acoust. Soc. Am. **29**, 189–194 (1957).
[11] H. D. McNiven, "Extensional waves in a semi-infinite elastic rod," J. Acoust. Soc. Am. **33**, 23–27 (1960).
[12] J. Zemanek, "An experimental and theoretical investigation of elastic wave propagation in a cylinder," J. Acoust. Soc. Am. **51**, 265–282 (1972).
[13] R. D. Gregory and I. Gladwell, "Axisymmetric waves in a semi-infinite elastic rod," Q. J. Mech. Appl. Math. **42**, 327–337 (1989).
[14] M. de Billy, "End resonance in infinite immersed rods of different cross sections," J. Acoust. Soc. Am. **100**, 92–97 (1996).
[15] J. Tasi, "Reflection of extensional waves at the end of a thin cylindrical shell," J. Acoust. Soc. Am. **44**, 291–292 (1968).
[16] J. Kaplunov, L. Y. Kossovich, and M. V. Wilde, "Free localized vibrations of a semi-infinite cylindrical shell," J. Acoust. Soc. Am. **107**, 1383–1393 (2000).
[17] V. T. Grinchenko and G. L. Komissarova, "Features of the dynamic deformation of a hollow cylinder," Int. Appl. Mech. **22**, 393–397 (1986).
[18] D. C. Gazis, "Three-dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation," J. Acoust. Soc. Am. **31**, 568–573 (1959).
[19] A. H. Meitzler, "Mode coupling occurring in the propagation of elastic pulses in wires," J. Acoust. Soc. Am. **33**, 435–445 (1961).
[20] B. Pavlakovic, M. J. S. Lowe, D. Allayne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1997), Vol. **16**, pp. 185–192.
[21] H. Nishino, S. Takashina, F. Uchida, M. Takemoto, and K. Ono, "Modal analysis of hallow cylindrical guided waves and applications," Jpn. J. Appl. Phys., Part 1 **40**, 364–370 (2001).
[22] B. Morvan, N. Wilkie-Chancellier, H. Duflo, A. Tinel, and J. Duclos, "Lamb wave reflection at the free edge of the plate," J. Acoust. Soc. Am. **113**, 1417–1425 (2003).
[23] M. Castaings, E. Le Clezio, and B. Hosten, "Modal decomposition method for modelling the interaction of Lamb waves with cracks," J. Acoust. Soc. Am. **112**, 2567–2582 (2002).
[24] I. J. Thompson and A. R. Barnett, "COULCC: A continued-fraction algorithm for Coulomb functions of complex order with complex arguments," Comput. Phys. Commun. **36**, 363–372 (1985).

[25]ABAQUS 6.5, Analysis User's Manual, Abaqus, 2004.
[26]D. N. Alleyne, M. J. S. Lowe, and P. Cawley, "The reflection of guided waves from circumferential notches in pipes," J. Appl. Mech. **65**, 635–641 (1998).
[27]M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "The mode conversion of a guided wave by a part-circumferential notch in a pipe," J. Appl. Mech. **65**, 649–656 (1998).
[28]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).
[29]B. Pavlakovic, D. Alleyne, and M. J. S. Lowe, "Simulation of Lamb wave propagation using pure mode excitation," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1998), Vol. **17**, pp. 1003–1010.
[30]R. D. Gregory and I. Gladwell, "The reflection of a symmetric Rayleigh-Lamb wave at the fixed or free edge of a plate," J. Elast. **13**, 185–206 (1983).
[31]B. PavIakovic, "Leaky guided ultrasonic waves in NDT," Ph.D. thesis, University of London, 1998.

Active damping control unit using a small scale proof mass electrodynamic actuator

Cristóbal González Díaz,[a] Christoph Paulitsch,[b] and Paolo Gardonio[c]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 9 July 2007; revised 30 April 2008; accepted 23 May 2008)

This paper presents a study on the design and use of a small scale proof mass electrodynamic actuator, with a low mounting resonance frequency, for velocity feedback control on a thin rectangular panel. A stability-performance formula is derived, which can be effectively used to assess the down scaling effects on the stability and control performance of the feedback loop. The design and tests of a velocity feedback loop with a prototype small scale proof mass actuator are also presented. When a feedback control having a gain margin of about 6 dB is implemented, so that there is little control spillover effect around the fundamental resonance of the actuator, reductions of vibration between 5 dB and 10 dB in the frequency band between 80 Hz and 250 Hz have been measured at the control position. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945167]

PACS number(s): 43.40.Vn, 43.50.Ki, 43.40.At [KAC] Pages: 886–897

I. INTRODUCTION

This paper presents a study on the design and practical implementation of a velocity feedback control loop on a thin rectangular panel with a small scale proof mass electrodynamic actuator. Particular emphasis is given to the scaling effects of reducing the actuator size on stability and control performance when direct velocity feedback is implemented. The aim of the scaling study is to provide general guidelines for the design of the prototype small scale proof mass actuator.

Sound radiation and transmission by thin, lightly damped, partitions are relevant problems both in land and air transportation vehicles. For instance, in order to improve the fuel consumption efficiency, the new designs of aircraft fuselage and automobile bodywork involve stiff and lightweight panels, which therefore efficiently radiate noise generated by external acoustic sources (i.e., jet noise or reciprocating engine noise), by aerodynamic sources (i.e., turbulent boundary layer pressure fields on aircraft skins or on car bodyworks), and by structure borne paths (i.e., engine induced vibrations or road induced vibrations).[1–3] Passive treatments, such as stiffening, mass, or damping treatments, can be used to reduce this problem[4,5] although, in order to be effective at low audio frequencies, they tend to be bulky and introduce extra weight that contrasts with the fuel consumption requirement. Recent research work has shown that decentralized active vibration control with point force actuators provides an efficient solution to the low frequency sound transmission problem.[6]

At low frequency, the sound radiation by a lightly damped panel is characterized by well separated resonances of the low order modes of the panel.[7,8] Thus, steady state broad band sound radiation can be effectively reduced by tackling the low frequency resonances of the panel itself[6,8] without the need to involve control arrangements with complicated sensor and actuator pairs that operate on the sound radiation modes of the structure.[9–11] In general, the resonant response of structures is governed by damping, which can be effectively controlled by active systems that produce damping actions.[7] This can be achieved with simple direct velocity feedback control systems.[12]

If ideal point force actuators and velocity sensors are used, then collocated velocity feedback is bound to be unconditionally stable.[13,14] In practice, force actuators are constructed with an electrodynamic actuator that reacts off a proof mass. This type of actuator produces a constant force above the fundamental resonance frequency of the resiliently mounted mass.[12] Thus, they must be designed with a relatively low fundamental resonance frequency, which must be well below the first resonance frequency of the structure to be controlled. Also, Elliott *et al.*[15] have shown that in order to guarantee large control gains, this actuator fundamental resonance must be well damped.

Theoretical work has shown that, in order to effectively produce damping on panels with point actuators, several velocity feedback loops should be implemented.[6] The control bandwidth tends to rise with the number of control units.[10] Thus, it is necessary to design small scale control systems which withstand the necessary stability requirements for the implementation of direct velocity feedback loops and produce the largest possible damping effect.

This paper introduces the stability and performance study of a small scale prototype electrodynamic proof mass actuator for the implementation of a direct velocity feedback loop on a thin panel structure. In Sec. II, a mobility/impedance formulation is presented which provides a simple "stability-performance" formula that can be used to assess simultaneously the stability and control performance of the feedback control loop with the proof mass actuator. In Sec.

[a]Electronic mail: cgd@isvr.soton.ac.uk
[b]Electronic mail: cpaulits@gmx.de
[c]Electronic mail: pg@isvr.soton.ac.uk

TABLE I. Geometry and physical parameters for the clamped aluminium panel.

Parameter	Value
Dimensions	$l_x \times l_y = 414 \times 314$ mm^2
Thickness	$h = 1$ mm
Mass density	$\rho = 2720$ kg/m^3
Young's modulus	$E = 7.1 \times 10^{10}$ N/m^2
Poisson ratio	$\nu = 0.33$
Damping loss factor	$\eta = 0.02$
Coordinates of primary force excitation	$x_p, y_p = 341,246$ mm
Coordinates of control point	$x_c, y_c = 109,75$ mm
Mass of the force transducer for the primary excitation	$M_s = 30$ g

TABLE II. Geometry and physical parameters for the actuator.

Parameter	Value
Proof mass diameter	24 mm
Proof mass height	12 mm
Magnet diameter	18 mm
Magnet height	9.3 mm
Base disk diameter	38 mm
Base disk thickness	1 mm
Housing and base disk mass	$M_b = 8$ g
Proof mass	$M_a = 22$ g
Suspension system stiffness	$K_a = 347.4$ N/m
Suspension system damping	$C_a = 3.3$ N/m s^{-1}
Fundamental natural frequency	$f_a = 20$ Hz
Vioce coil coefficient	$\psi = 2.6$ N/A

III, the principal downscaling and design issues of the proof mass actuator are discussed in view of the stability requirements and control performance properties of the feedback loop. In particular, the scaling laws are derived for (a) the fundamental natural frequency ω_a, (b) the static displacement δ_a, (c) the current in the coil windings i_a, (d) the generation of electrodynamic force f_a and transmitted force f_c, (e) the stroke of the suspended mass Δw, (f) the maximum gain that guarantees stability $g_{\max}$, and (f) the "control performance ratio" R_k. The design and experimental tests of a prototype velocity feedback control unit with a small scale proof mass actuator are then presented in Sec. IV. In particular, the stability and control performance are discussed with reference to the Nyquist criterion applied to the locus of the open loop sensor-actuator Frequency Response Function (FRF).

II. DIRECT VELOCITY FEEDBACK CONTROL USING A PROOF MASS ACTUATOR

Before entering into the details of the downscaling study, the principal characteristics of force actuation with a proof mass electrodynamic actuator mounted on a rectangular panel are considered. This analysis has been carried out with reference to the panel and prototype actuator considered in Sec. IV whose geometrical and physical properties are summarized in Tables I and II. The stability and control performance of a negative velocity feedback loop using this type of actuator are examined with a simple stability-performance formula which derives the reduction of vibration at the control position for the maximum gain of the feedback control loop that guarantees stability. The scaling laws of the principal mechanical and electrodynamic components are then derived and used to assess the downscaling effect on the stability and control performance of the control unit.

A. Actuation mechanism

As schematically shown in Fig. 1(a), the control system considered in this study uses a coil-magnet electrodynamic linear motor. The coil is fixed to the base of the actuator and the magnet is suspended on springs so that it provides the inertial reaction necessary to generate a point force f_c on the structure where the actuator is fixed.[12]

Figure 1(b) shows the equivalent electromechanical schematic that has been used to model the response of this actuator when it is fixed on a clamped rectangular aluminum panel. The model takes into account the inertial effect of the proof mass and stiffness-damping effects of the suspension system. In order to keep the formulation simple, the inertial effect of the base and coil masses is instead neglected. Assuming time harmonic vibratory motion of the form $\exp(j\omega t)$, where ω is the circular frequency and $j = \sqrt{-1}$, the fully coupled response of the plate and actuator system has been derived by considering the following mobility and impedance equations:

$$\dot{w}_c = Y_{cc} f_c + Y_{cp} f_p, \tag{1}$$

$$\dot{w}_m = Y_a f_m, \tag{2}$$

$$f_c = -Z_a \dot{w}_c + Z_a \dot{w}_m - f_a \tag{3}$$

and

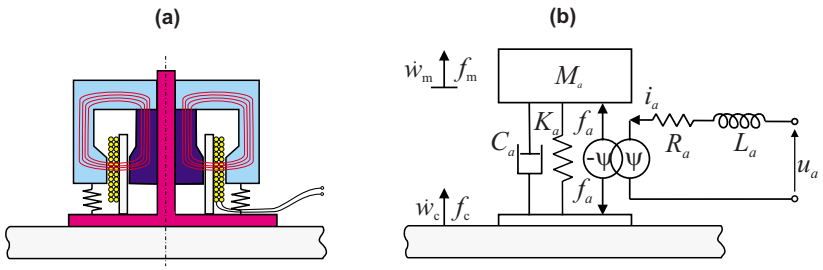


FIG. 1. (Color online) Proof mass electrodynamic force actuator: (a) sketch; (b) electromechanical schematic. As shown in (a), the proof mass is formed by a magnetic core cylinder and an outer ferromagnetic ring.

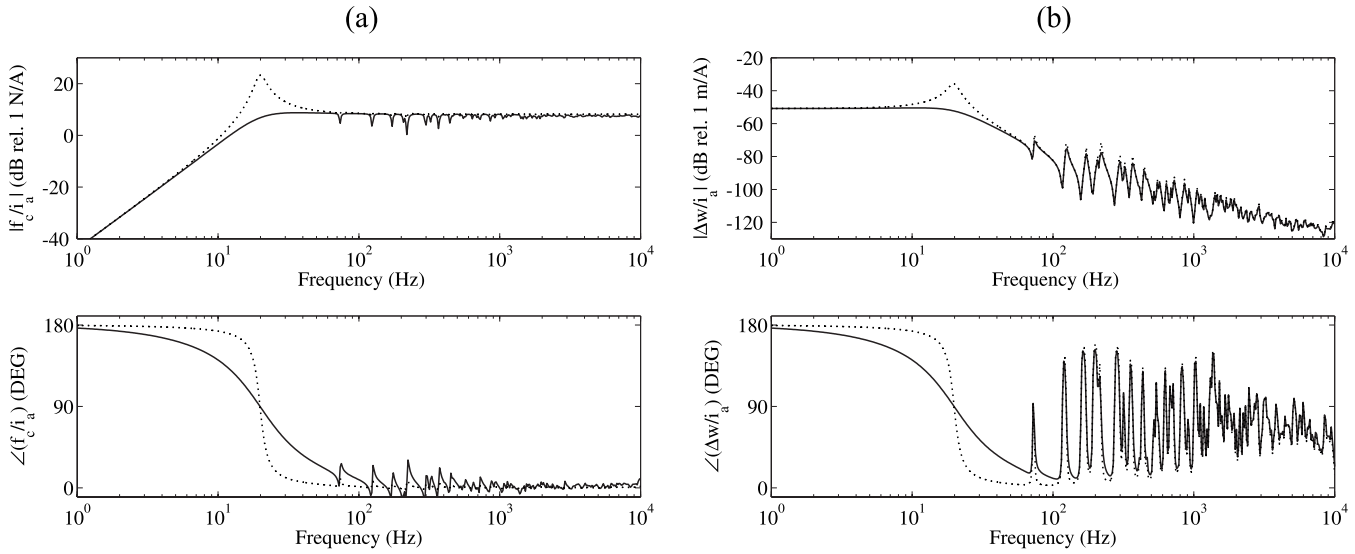


FIG. 2. (a) Simulated force transmitted to the plate per unit driving current. (b) Simulated stroke per unit driving current. Thick lines heavily damped actuator (C_a=3.3 N/m s^{-1}), dotted lines lightly damped actuator (C_a=0.5 N/m s^{-1}).

$$f_m = Z_a \dot{w}_c - Z_a \dot{w}_m + f_a, \tag{4}$$

where $\dot{w}_c(\omega)$, $\dot{w}_m(\omega)$, $f_c(\omega)$, and $f_m(\omega)$, are, respectively, the complex velocities and forces at the base and proof mass components of the actuator and $f_p(\omega)$ is the complex primary force excitation acting on the plate. $Y_{cc}(\omega)$ and $Y_{cp}(\omega)$ are the plate mobility functions, respectively, at the point where the actuator is attached and between the point where the actuator is attached and the location of the primary force. The two mobility functions have been derived in terms of the following modal summations:[16]

$$Y_{cc} = j\omega \sum_{n=1}^{N} \frac{[\phi_n(x_c,y_c)]^2}{M_p[\omega_n^2(1+j\eta)-\omega^2]}, \tag{5}$$

$$Y_{cp} = j\omega \sum_{n=1}^{N} \frac{\phi_n(x_c,y_c)\varphi_n(x_p,y_p)}{M_p[\omega_n^2(1+j\eta)-\omega^2]}, \tag{6}$$

where M_p is the mass of the plate, η is the loss factor, and ω_n and $\phi_n(x,y)$ are, respectively, the nth natural frequency and nth natural mode of the plate at position (x,y), which have been taken from Ref. 16 for a clamped plate. Finally, $Z_a(\omega)$ and $Y_a(\omega)$ are the impedance and mobility functions for the spring-dashpot and proof mass components of the actuator:[16]

$$Z_a = \frac{K_a}{j\omega} + C_a, \tag{7a}$$

$$Y_a = \frac{1}{j\omega M_a}, \tag{7b}$$

where K_a, C_a, and M_a are, respectively, the stiffness, viscous damping coefficient, and mass of the three components of the actuator.

Assuming the actuator is driven by current, i_a, so that

$$f_a = -\psi i_a, \tag{8}$$

where ψ is the voice coil factor of the actuator;[12] using Eqs. (1)–(4), the force transmitted to the base f_c and the stroke of the suspended mass with reference to the base of the actuator $\Delta w = w_m - w_c$ are given by

$$f_c = \frac{\psi}{1 + Z_a(Y_a + Y_{cc})} i_a, \tag{9}$$

$$\Delta w = -\frac{1}{j\omega} \frac{(Y_a + Y_{cc})\psi}{1 + Z_a(Y_a + Y_{cc})} i_a. \tag{10}$$

The plot in Fig. 2(a) shows the spectrum of the transmitted force f_c per unit driving current i_a. Considering first the case where the actuator is lightly damped (dotted line), at frequencies below the fundamental resonance frequency of the actuator, the transmitted force f_c is out of phase with the driving current signal and monotonically rises from zero up to a maximum value at the fundamental resonance of the actuator at about 20 Hz. At higher frequencies, the transmitted force f_c is in phase with the driving current signal and its amplitude levels down to a constant value which is approximately equal to the reactive forces f_a generated by the coil-magnet linear motor.[12] The narrow band troughs of the amplitude are due to the low impedance effect produced by the plate at resonance frequencies. When the actuator is heavily damped (thick line), the actuator resonance peak is smoothed down and can hardly be seen. Also, the transition from out of phase to in phase force actuation is stretched out over a wider frequency band. In conclusion, if negative velocity feedback is implemented, the desired damping action is produced only above the fundamental resonance of the actuator. In contrast, below this frequency, negative damping is generated, which tends to destabilize the control loop. This observation already offers a key indication about the fundamental issue of this type of actuator, that is, the actuator should be designed with the smallest possible fundamental resonance frequency in order to ensure a constant force excitation in a wider range of low audio frequencies where the active damping effect is mostly desired.

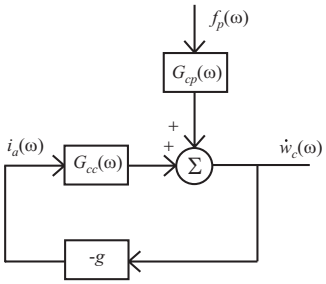


FIG. 3. Block diagram of the velocity feedback control loop using a current driven electrodynamic proof mass actuator.

The plot in Fig. 2(b) shows the spectrum of the stroke Δw per unit driving current i_a. Considering first the case where the actuator is lightly damped (dotted line), at frequencies below the fundamental resonance frequency of the actuator, the stroke is out of phase with the driving current signal. According to Eq. (10), for $\omega=0$, the stroke per unit current is given by $\Delta w=-(\psi/K_a)i_a$. The amplitude of the stroke tends to remain constant up to frequencies close to the fundamental resonance of the actuator at about 20 Hz where it becomes maximum. At higher frequencies, the stroke is in phase with the driving current signal and its amplitude rolls off monotonically with frequency. Between 100 Hz and 1 kHz, the amplitude of the spectrum of the stroke is characterized by a sequence of resonance peaks due to the low order natural modes of the panel. As the frequency rises, the modal overlap in the plate increases so that the amplitude spectrum of the stroke becomes smoother above 1 kHz. As found for the control force f_c, when the actuator is heavily damped, the actuator resonance peak is smoothed down and the phase transition from +180° to 0° is stretched out over a wider frequency band. In conclusion, the maximum stroke produced by the actuator is at the fundamental resonance of the actuator at about 20 Hz.

Substituting Eq. (10) into Eq. (9), the transmitted force f_c can be expressed in terms of the stroke of the proof mass

$$f_c = -\frac{j\omega}{(Y_a + Y_{cc})}\Delta w. \tag{11}$$

This expression highlights that the transmitted force f_c depends on the stroke of the proof mass via the impedance $Z_{pa}=1/(Y_a+Y_{cc})$ offered by the proof mass and the plate at the control point. Thus, for a given maximum current i_a that can be fed to the actuator without damaging the coil and for a maximum stroke Δw that the suspended mass can withstand, without hitting the end stops of the actuator, the force transmitted f_c to the base of the actuator can be enhanced by increasing the proof mass so that the plate-actuator impedance Z_{pa} also rises.

B. Stability of a direct velocity feedback loop

Assuming the error sensor for the feedback loop is an ideal velocity sensor located at the base of the actuator, in which case it measures exactly $\dot{w}_c$, the response of the panel measured by the error sensor can be modeled in terms of the classic disturbance rejection feedback block diagram shown in Fig. 3, where $G_{cc}(\omega)$ and $G_{cp}(\omega)$ are the fully coupled FRFs between the error sensor velocity $\dot{w}_c$ and either the control current i_a or primary force excitation f_p, which can be derived from Eqs. (1)–(4):

$$G_{cc}(\omega) = \frac{\dot{w}_c}{i_c} = \frac{Y_{cc}\psi}{1 + Z_a(Y_a + Y_{cc})}, \tag{12}$$

$$G_{cp}(\omega) = \frac{\dot{w}_c}{f_p} = \frac{(1 + Z_a Y_a)Y_{cp}}{1 + Z_a(Y_a + Y_{cc})}. \tag{13}$$

Figure 4 shows the Bode and Nyquist plots of the open loop sensor-actuator FRF $gG_{cc}(\omega)$ assuming $g=1$. The Bode plot shows that the modulus of $G_{cc}(\omega)$ is characterized by a heavily damped resonance at about 20 Hz, which is due to the actuator fundamental natural mode, and then a sequence of resonance peaks, which are due to the low order natural modes of the plate. The phase of $G_{cc}(\omega)$ starts from +270° at low frequency, drops to +90° beyond the resonance of the actuator, and then it alternates between +90° and −90° for the resonances of the plate. As a result, the Nyquist plot shows that the loci of $G_{cc}(\omega)$ is characterized by one circle in the left hand side quadrants, which is due to the resonance of the actuator, and many circles in the right hand side quadrants, which are due to the resonances of the plate. Therefore, using the Nyquist stability criterion,[17] the control system is found to be only conditionally stable since, for relatively high control gains, the circle on the left hand side due to the fundamental resonance of the actuator can enclose the Nyquist instability point $-1+j0$. Also, for control gains that ensure stability, the circle in the left hand side quadrants indicates that control spillover is bound to occur around the fundamental resonance frequency of the actuator.

In summary, the stability and control performance of a velocity feedback control loop with a proof mass actuator are heavily affected by the presence of the fundamental resonance of the actuator. In order to guarantee a stable feedback control loop with high control gains, it is necessary to reduce the amplitude of the actuator resonance so that the left hand side circle in the Nyquist plot would be small. Nevertheless, when large control gains are implemented, undesired control spillover effect takes place.

In order to reduce the amplitude of the first resonance peak of $G_{cc}(\omega)$, the fundamental natural frequency of the actuator should be kept as low as possible. This can be achieved by designing the proof mass suspension system with very soft springs. However, with a too soft suspension system, the static displacement of the proof mass becomes too big so that practical problems, such as nonlinearity due to the proof mass striking the end stops of the actuator, may disrupt the stability of the feedback control loop.[18] The amplitude of the fundamental resonance of the proof mass actuator can also be lowered by increasing the damping effect in the actuator. However, this approach is also affected by practical problems, which again may involve nonlinearity.

C. Control performance

As discussed above, the velocity feedback control system considered in this paper is devised to produce active damping, which efficiently reduces the response around the

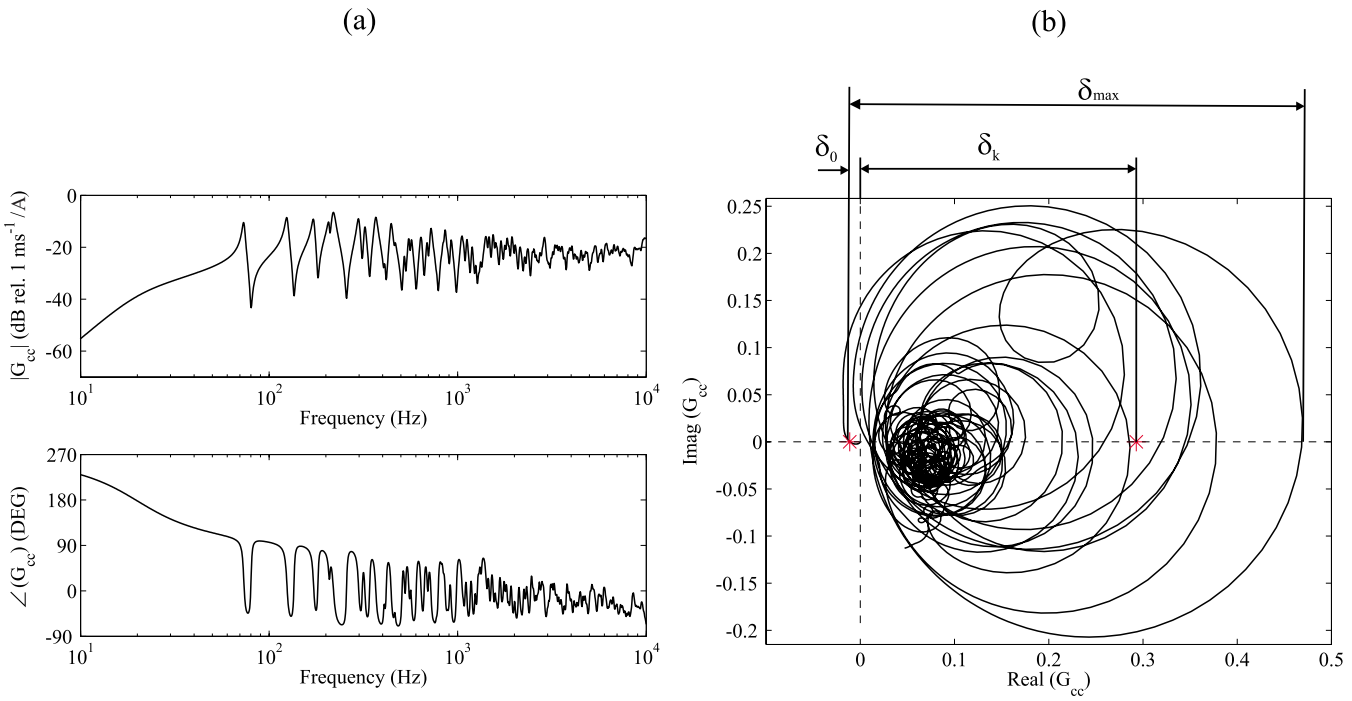


FIG. 4. (Color online) Bode (a) and Nyquist (b) plots of the simulated open loop sensor-actuator FRF $gG_{cc}(\omega)$, assuming $g=1$ when a proportional feedback loop is used for current control.

resonance frequencies of low order modes of the panel. Thus, the effectiveness of the control system can be evaluated by looking to the reduction of vibration that can be achieved at each resonance frequency of the panel rather than over the whole frequency band. The active damping effect produced by each control unit can be assessed by considering the reduction of vibration at the control position, although it should be noted that this does not directly correspond to a mean reduction of vibration over the panel surface. In fact, as discussed in Ref. 19 for too high control gains, the control units produce a pinning effect at the control position which simply rearranges the spatial vibration of the panel and does not inject damping to the structure.

According to the block diagram in Fig. 3, the responses at the control position per unit primary excitation with and without feedback control are, respectively, given by

$$\frac{\dot{w}_{c/c}}{f_p} = \frac{G_{cp}}{1 + gG_{cc}}, \tag{14}$$

$$\frac{\dot{w}_{c/nc}}{f_p} = G_{cp}. \tag{15}$$

The maximum reduction of vibration at the control position at the kth resonance frequency ω_k is therefore defined as

$$\rho_k = \frac{|\dot{w}_{c/c}/f_p(\omega_k)|}{|\dot{w}_{c/nc}/f_p(\omega_k)|} = \frac{1}{|1 + g_{\max}G_{cc}(\omega_k)|}, \tag{16}$$

where $g_{\max}$ is the maximum feedback control gain that guarantees stability which, approximately, can be taken as the reciprocal of the real part of the open loop sensor-actuator FRF, G_{cc}, at the fundamental natural frequency of the actuator $\omega_a = \sqrt{K_a/M_a}$:

$$g_{\max} \approx -\frac{1}{\mathrm{Re}\{G_{cc}(\omega_a)\}}. \tag{17}$$

The response at the resonance frequencies of the low order natural modes of the plate can also be approximated by the real parts of the open loop sensor-actuator FRF, G_{cc}, at the resonance frequencies, i.e., $G_{cc}(\omega_k) \approx \mathrm{Re}\{G_{cc}(\omega_k)\}$, so that the ratio ρ_k can be expressed as

$$\rho_k \approx \frac{1}{1 + \delta_{k0}}, \tag{18}$$

where $\delta_{k0} = \delta_k/\delta_0$ with $\delta_k = \mathrm{Re}\{G_{cc}(\omega_k)\}$ and $\delta_0 = -\mathrm{Re}\{G_{cc}(\omega_a)\}$, as shown in Fig. 4(b). Normally, the control performance ratio ρ_k is expressed in decibels with the following formula:

$$R_k = 20\log_{10}\left(\frac{1}{\rho_k}\right) \approx 20\log_{10}(1 + \delta_{k0}). \tag{19}$$

The stability and control performance effects are often analyzed separately. This simple expression contains both information on the stability and control performance of the feedback system. The stability-performance formulas in Eqs. (18) and (19) suggest that in order to maximize the control performance, the ratio $\delta_{k0} = \delta_k/\delta_0$, that is, the ratio between $\mathrm{Re}\{G_{cc}(\omega_k)\}$ and $-\mathrm{Re}\{G_{cc}(\omega_a)\}$, should be maximized.

The ratio $\delta_{k0} = \delta_k/\delta_0$ can be derived in a simplified form by making some assumptions about the driving point mobility function of the plate structure $Y_{cc}(\omega)$. In fact, assuming $\omega = \omega_a < \omega_1$, where ω_1 is the first natural frequency of the plate, the mobility function for $Y_{cc}(\omega)$ given in Eq. (5) can be approximated by

$$Y_{cc}(\omega_n) \approx \frac{j\omega_a}{K_p}, \tag{20}$$

where $1/K_p = \Sigma_{n=1}^{N}[\phi_n(x_c,y_c)]^2/M_p\omega_n^2$. Thus, using Eq. (12), the value of δ_0 is found to be

$$\delta_0 = -\mathrm{Re}\{G_{cc}(\omega_a)\} \approx \frac{\omega_a \psi}{2\zeta_a K_p - C_a\omega_a}. \tag{21}$$

Assuming now $\omega=\omega_k>\omega_a$, the mobility function for $Y_{cc}(\omega)$ in Eq. (5) can be approximated by

$$Y_{cc}(\omega_k) \approx \frac{1}{C_{p,k}}, \tag{22}$$

where $1/C_{p,k}=[\phi_k(x_c,y_c)]^2/M_p\omega_k\eta$. Thus, using Eq. (12), the value of δ_k is found to be

$$\delta_k = \mathrm{Re}\{G_{cc}(\omega_k)\} \approx \frac{\psi}{C_a + C_{p,k}}. \tag{23}$$

In conclusion, assuming $\omega=\omega_k>\omega_a$, where ω_k is the kth natural frequency of the plate, the maximum reduction of vibration at the control position given by Eq. (18) is found to be

$$\rho_k \approx \frac{\omega_a(C_{p,k}+C_a)}{\omega_a C_{p,k} + 2\zeta_a K_p}. \tag{24}$$

This simple expression indicates that the control performance of a velocity feedback control loop with a proof mass electrodynamic actuator for the well separated resonance frequencies of low order modes of the structure rises as the fundamental resonance frequency of the actuator decreases.

III. DOWNSCALING OF AN ELECTRODYNAMIC PROOF MASS ACTUATOR

Normally the downscaling study of a control system is carried out by assessing how the force generated by an actuator varies as the size of the actuator is scaled down. Although at first sight this looks as the right approach, it is also important to analyze how the stability and active control, i.e., active damping, effects vary with the downscaling of the actuator. In this way, it is possible to get an indication whether for decentralized control it would be convenient to use few, large scale, control units or many, small scale, control units over the surface of a panel. In the following sections, the downscaling laws for the principal components of the actuator are first revised. The downscaling laws for the actuator driving current, control force, and stroke are then considered. Finally, the stability-performance analysis presented in the previous section is used to assess how the maximum control performance varies with the downscaling of the actuator. As described by Madou,[20] the scaling laws are expressed with a $[L^n]$ notation, where n identifies the power of the linear dimension L. For those quantities that remain unchanged with scaling, a $[L^0]$ scaling law is assigned. For instance, the physical (mobility function and natural frequencies) and geometrical (dimensions) properties of the panel will have a scaling law $[L^0]$.

A. Scaling laws of the mechanical components of the actuator

According to the schematic shown in Fig. 1(b) and the analytical formulation for the response of the system presented in Sec. II, the principal mechanical components of the actuator are the proof mass, the suspension system, and, although it does not correspond to a self-contained mechanical component, the viscous damper which describes the damping effect produced by the squeeze film of lubricant between the guiding stinger and the central hole in the proof mass [see Fig. 1(a)]. The scaling laws for the mass, stiffness, and damping effects of these three components can be derived by inspection of the formulas for these quantities.

Despite the cross section of the magnet proof mass is rather involved, it can be readily shown that its scaling law is

$$M_a \propto [L^3].$$

In order to minimize the deformation stress effect in the mounting spring,[21] a spring with ring shape has been chosen for the suspension of the mass in the prototype system studied in this paper. According to Young and Budynas,[22] the stiffness of a ring to a diametric load is given by

$$K_a = \frac{EI}{R^3}\left(\frac{4\pi}{k_1\pi^2 - 8k_2^2}\right), \tag{25}$$

where E is Young's modulus of the material of the spring, R is the radius of the ring, I is the area moment of the cross section of the ring (for rectangular section of dimensions $b \times h$, it is $I=bh^3/12$), and k_1, k_2 are the dimensionless constants.[22] As a result, the scaling law for the suspension system is given by

$$K_a \propto [L^1].$$

Finally, the coefficient for the damping effect produced by the squeeze film of lubricant between the guiding stinger and the central hole in the proof mass results directly proportional to the linear dimension;[23] thus,

$$C_a \propto [L^1].$$

In conclusion, as graphically summarized in Fig. 5(a), as the sizes of the actuator are scaled down the stiffness and damping coefficient falls down with power 1 of L while the mass effectively falls down with power 3 of L.

B. Electrodynamic actuation force scaling laws

The scaling laws of the actuation force produced by a coil-magnet actuator have been presented by Trimmer[24] who has considered three cases: (a) constant current density, J_a, in the coil; (b) constant heat flow, ΔQ, per unit surface area of the windings in the coil, and (c) constant temperature difference, ΔT, between the windings of the coil and the surrounding environment. For these three cases, the scaling of current density, J_a, in the windings of the coil with reference to dimension is given by the following:

(a) constant J_a $J_a \propto [L^0]$,
(b) constant ΔQ $J_a \propto [L^{-0.5}]$, and
(c) constant ΔT $J_a \propto [L^{-1}]$.

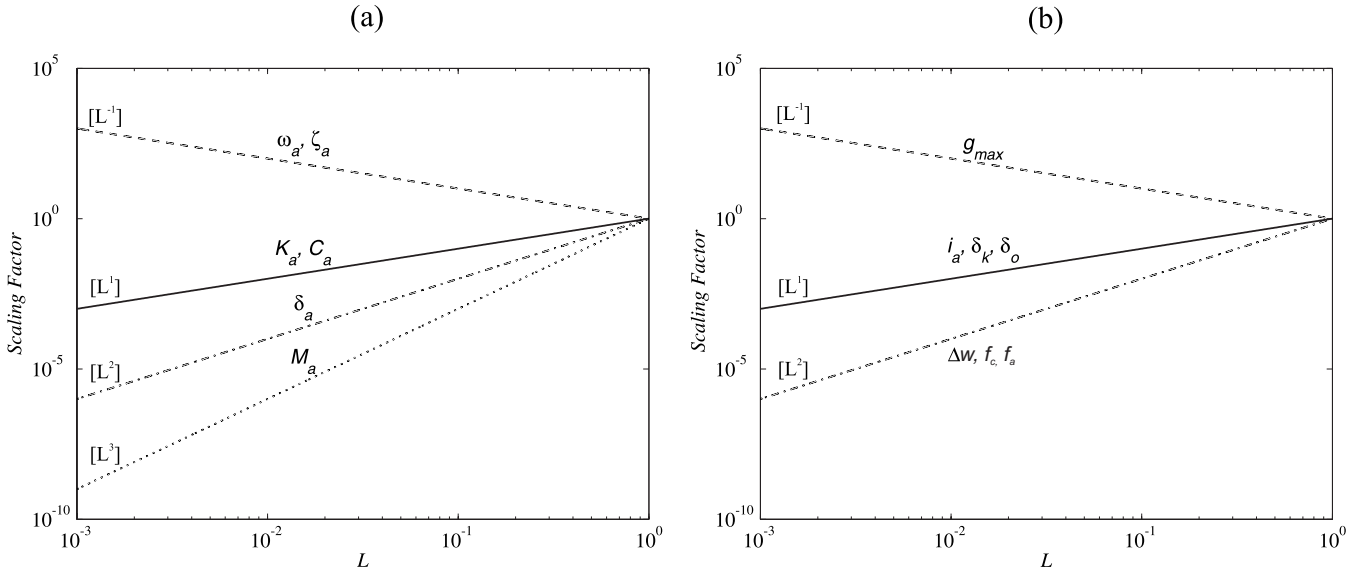


FIG. 5. Scaling laws. (a) Mass M_a, damping factor C_a, stiffness K_a, fundamental natural frequency ω_a, static displacement δ_a, and damping ratio ζ_a. (b) Driving current i_a, electrodynamic actuation force f_a, force transmitted to the base of the actuator f_c, stroke of the suspended mass Δw, δ_0, and δ_k functions, and maximum control gain $g_{\max}$.

In practice, the most relevant case, which is considered in the remaining part of the paper, is that for constant temperature difference since, normally, it is high temperature that causes coil windings to fail. As shown by the block diagram in Fig. 3, the feedback loop considered in this study drives the actuator with a current signal. The current in the winding of the actuator coil is given by $i_a=J_aA$, where A is the cross sectional area of the winding. Thus, the scaling law of the current that can be fed to the coil of the actuator with reference to dimension is given by

$$|i_a| \propto [L^1].$$

Assuming the proof mass is made of a permanent magnet cylindrical core with an outer ferromagnetic ring [see Fig. 1(a)] that, because of its cross sectional shape, generates a constant magnetic field B across the windings of the coil, the actuation force f_a is given by Eq. (8) with the voice coil coefficient given by $\psi=Bl$, where l is the total length of the windings. Thus,

$$f_a = -Bli_a. \tag{26}$$

As a result, considering constant ΔT between windings and environment and assuming that B scales with $[L^0]$, the scaling law of the actuation force is given by

$$|f_a| \propto [L^2].$$

In summary, as shown in Fig. 5(b), as the size of the actuator is scaled down the current that can be fed to drive a coil-magnet motor i_a falls down with power 1 of L and the force generated by a coil-magnet linear motor f_a falls down with power 2 of L.

C. Scaling laws of the static and dynamic characteristics of the proof mass actuator

The principal static and dynamic characteristics of the proof mass actuator are given by its fundamental natural frequency, static displacement, and damping ratio, which are given by the following expressions:

$$\omega_a = \sqrt{\frac{K_a}{M_a}}, \tag{27a}$$

$$\delta_a = \frac{M_a g}{K_a}, \tag{27b}$$

$$\zeta_a = \frac{C_a}{2\sqrt{K_a M_a}}. \tag{27c}$$

Thus, the scaling laws for these three functions are given by

scaling of ω_a: $\omega_a \propto [L^{-1}]$,

scaling of δ_a: $\delta_a \propto [L^2]$,

scaling of ζ_a: $\zeta_a \propto [L^{-1}]$.

As shown in Fig. 5(a), these three expressions indicate that as the size of the actuator is scaled down, the fundamental natural frequency and the damping ratio tend to rise with power 1 of L, while the static displacement falls down with power 2 of L. The first effect is undesirable since, as we have seen in Sec. II, it is of critical importance to design the actuator with the smallest possible fundamental natural frequency. In contrast, the fact that the static displacement falls down with the square of downscaling is a positive effect. Finally, the increase of the damping ratio in proportion to the downscaling should also be regarded as a positive effect since, according to the discussion presented in Sec. II, the closer is the damping to the critical damping factor, the

smaller is likely to be the amplitude at the fundamental resonance of the actuator and thus the higher should be the maximum control gain for a stable feedback control loop.

The last two fundamental characteristics of the proof mass actuator to be assessed are the stroke of the suspended mass $\Delta w = w_m - w_c$ and the force it can transmit to the base f_c. To ensure a correct operation of the actuator, the stroke of the suspended mass should be kept within the linear deflection range of the axial springs; thus, it should fall down with power 1 of L as the size of the actuator is scaled down, i.e., $|\Delta w| \propto [L^1]$. Thus, recalling the direct relation between the control force and the stroke of the suspended mass given in Eq. (11), also the control force should fall down with power 1 of L as the size of the actuator is scaled down, i.e., $|f_c| \propto [L^1]$. However, Eqs. (9) and (10) indicate that the stroke of the suspended mass and the force transmitted to the base depend on the current driving the actuator i_a, which was found to fall down with power 1 of L as the actuator is scaled down. Thus, the downscaling of the stroke of the suspended mass and control force should also be analyzed with respect to the downscaling of the current driving the actuator. As shown in Fig. 2(b), the maximum stroke of the actuator occurs at its fundamental resonance frequency, i.e., for $\omega = \omega_a$, in which case Eq. (10) reduces to

$$\Delta w = \frac{(\omega_a/K_a + jY_{cc})}{C_a\omega_a Y_{cc} - j(2\zeta_a\omega_a + K_a Y_{cc})}\psi i_a, \tag{28}$$

so that, considering constant ΔT between windings and environment, the scaling law for the stroke is given by

$$\text{Scaling of } |\Delta w|\text{: } |\Delta w| \propto [L^2].$$

The actuator is designed to implement the velocity feedback loop above its fundamental resonance frequency, i.e., for $\omega > \omega_a$, in which case Eq. (9) reduces to

$$f_c = \frac{\psi}{1 + Z_a Y_{cc}} i_a, \tag{29}$$

Thus, considering constant ΔT between windings and environment, the scaling law for the control force is given by

$$\text{Scaling of } |f_c|\text{: } |f_c| \propto [L^2].$$

Hence, it is the current that can be fed to the actuator rather than the linear deflection range of the axial springs that determines the downscaling laws for the stroke of the suspended mass and the force transmitted to the base of the actuator. As shown in Fig. 5(b), as the size of the actuator scales down, the stroke of the proof mass falls down with power 2 of L. Also, above the fundamental resonance frequency of the proof mass actuator, as the size of the actuator scales down, the force transmitted to the base of the actuator $|f_c|$ falls down with power 2 of L.

D. Scaling laws for the stability and performance of the velocity feedback loop

As discussed in Sec. II C, assuming the downscaling is limited to values such that the fundamental resonance frequency of the proof mass actuator is below the first resonance of the panel, the downscaling effect on the maximum control performance of the velocity feedback loop using the proof mass actuator can be assessed with Eq. (19): $R_k = 20\log_{10}(1+\delta_{k0})$. The scaling of the ratio $\delta_{k0} = \delta_k/\delta_0$ can be readily derived by substituting in Eqs. (21) and (23) the scaling laws for the mechanical components and the dynamic characteristics of the actuator, which gives

$$\text{Scaling of } \delta_0 = -\text{Re}\{G_{cc}(\omega_a)\} \quad \delta_0 \propto [L^1].$$

$$\text{Scaling of } \delta_k = \text{Re}\{G_{cc}(\omega_k)\} \quad \delta_k \propto [L^1].$$

Thus, as shown in Fig. 5(b), both terms fall down with power 1 of L as the size of the actuator is scaled down. As a result, the ratio $\delta_{k0} = \delta_k/\delta_0$, and consequently the control performance ratio R_k, should not change with the downscaling of the actuator. However, the control performance ratio derived with Eq. (19) refers to the maximum feedback control gain that guarantees a stable feedback control loop, which according to Eq. (17) is given by $g_{\max} = 1/\delta_0$. Thus, as shown in Fig. 5(b), it implies that the maximum control gain rises with power of 1 of L as the actuator size is scaled down:

$$\text{Scaling of } g_{\max} = 1/\delta_0 \quad g_{\max} \propto [L^1].$$

As a result, the current i_a feedback to the actuator should also rise with power 1 of L as the actuator size is scaled down. However, as highlighted in Sec. III B, the maximum current i_a that can be fed to a coil-magnet actuator falls down with power 1 of L as the actuator size is scaled down. Thus, the effective scaling law of the control performance should be assessed assuming $g_{\max} \propto [L^1]$, and thus $\delta_0 \propto [L^{-1}]$, in which case the scaling law for the control performance index is given by

$$\text{Scaling of } R_k = 20\log_{10}(1+\delta_{k0}) \quad R_k \propto [L^2].$$

Thus, as shown in Fig. 6, the control performance of a velocity feedback loop using a proof mass electrodynamic actuator falls down with power 2 of L as the size of the actuator is scaled down. It should be emphasized that this loss of control performance is purely due to the limitation of current that can be fed to the coil-magnet actuator as the size is

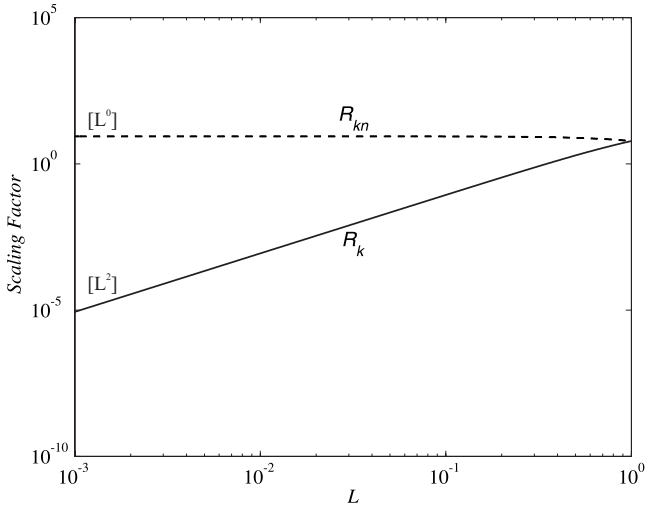


FIG. 6. Scaling laws for the control performance ratio considering one control units R_k and n decentralized control units R_{kn}.

scaled down. In other words, as the control actuator is scaled down, the feedback control gain that can be implemented in practice is much lower than the maximum value that would guarantee stability.

The number of control units that can be fitted per unit surface of the plate structure grows with power 2 or L as the size of the units is scaled down, i.e., $n \propto [L^{-2}]$; thus, the overall control performance that would be produced by n feedback control loops R_{kn} should remain constant. However, as highlighted above, the feedback control loops would be used increasing gain margins which guarantees more robust feedback loops. Bauman and Elliott[25] have highlighted that the control performance of multiple feedback loops tends to degrade as the number of control units rises because of instability issues generated by cross-talking effects between neighbor actuators. Thus, if small scale control units that operate with large gain margins were to be used, this problem could be less pronounced and thus better control performance can be obtained with dense arrays of small scale control units rather than with fewer large scale control units.

It is important to emphasize that these speculations are based on the assumption that the fundamental resonance frequency of the actuator is lower than the first resonance of the panel which, as discussed in Sec. III C, rises with power 2 of L. Thus, it is likely that there is an optimal trade-off between the number, size, mechanical properties, and gain margin of the control units that would produce the best control performance effect.

IV. DESIGN AND TESTING OF A PROTOTYPE SMALL SCALE PROOF MASS ACTUATOR FOR VELOCITY FEEDBACK CONTROL

This section introduces the design and testing of a prototype small scale electrodynamic proof mass actuator that has been used to implement a velocity feedback control loop on a clamped rectangular panel whose properties are summarized in Table I. The panel is excited by a point force produced by a shaker located underneath the panel at $x_p = 341$ mm, $y_p = 246$ mm. The actuator is positioned on the top of the panel at $x_c = 109$ mm, $y_c = 75$ mm. In order to implement a velocity feedback control loop, a small accelerometer sensor is located underneath the panel at the center of the actuator footprint. The velocity feedback loop is implemented with an analog controller which is composed by an integrator to get the error velocity signal, a dc decoupling high pass filter with corner frequency at about 10 Hz, and a current amplifier.

The stability analysis of the feedback control loop has been carried out following the methodology presented in Sec. II B. The control performance has been assessed by plotting the spectrum of the response at the control position per unit primary force excitation in a frequency range between 0 and 1 kHz.

A. Prototype small scale proof mass actuator

Following the guidelines of the scaling study presented in Sec. III, the miniaturization of the prototype actuator has been carried out up to a limit such that the fundamental natural frequency of the suspended mass remains below the first resonance frequency of the plate under control. The first resonance of the rectangular panel considered in this study is at about 73 Hz. Thus, it has been decided to design the proof mass and suspension spring system such that the fundamental resonance of the actuator is around 20 Hz. A trial and error approach has been used to identify the dimensions of the components of the actuator. In addition to the physical restraints derived from the scaling study, the design of the prototype actuator has also been constrained by fabrication limitations due to the manufacturing machine tools available in the laboratory.

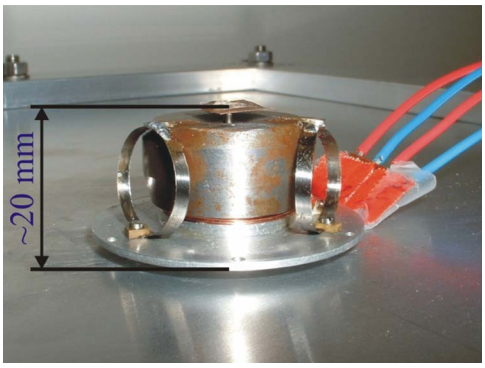


FIG. 7. (Color online) Photo of the small scale electrodynamic proof mass actuator.

The small scale prototype electrodynamic proof mass actuator designed and built for this study is shown in Fig. 7. The physical and geometrical properties of the actuator are summarized in Table II. As schematically shown in Fig. 1(a), this actuator is composed of a base disk with a cylindrical former on which the coil is wound. The proof mass is formed by a magnetic core cylinder and an outer ferromagnetic ring. The proof mass is mounted on three springs and a vertical bushing, which forces the magnet to oscillate in the axial direction.

As shown in Fig. 7, the three springs are made of small circular rings which guarantee a relatively larger stiffness in the transverse direction than in the axial direction. In this way, the fundamental axial natural frequency of the proof mass actuator can be kept rather low with a good transverse guiding which prevents nonlinear effects due to the stick slip friction on the axial bushing. As shown in Fig. 1(a), the cross section of the magnet is shaped in such a way as to have a magnetic circuit that generates a field oriented in the direction orthogonal to the coil winding. In this way, a current flow through the coil produces the reactive axial force between the coil and the magnet which is given by Eq. (8). The details of the design of the circular springs and coil-magnet transducers are presented by Paulitsch.[21]

B. Open loop stability analysis

The Bode plot in Fig. 8 shows the measured open loop sensor-actuator FRF. The plot shows that there is nearly no peak of the fundamental resonance of the actuator at about 20 Hz. As discussed in Sec. II B, this is due to the fact that the fundamental resonance of the actuator is well below the

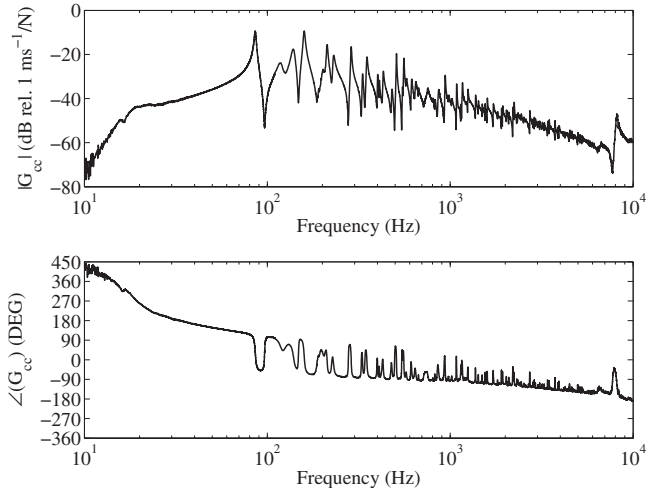


FIG. 8. Measured Bode plot of the open loop frequency response function between the error sensor signal and the input signal to the analog controller of the feedback loop.

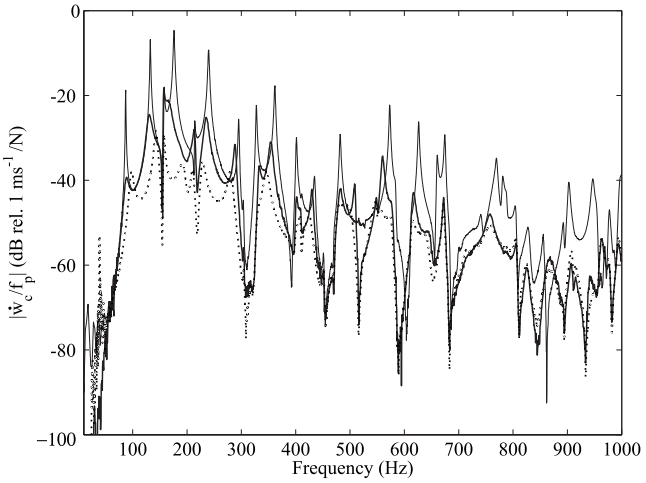


FIG. 10. Measured velocity at the error sensor position per unit primary force without actuator (faint line), with actuator and no control (thick line), with actuator when a gain margin of 6 dB (dotted line) is implemented.

first resonance of the panel and to the fact that there is a rather high internal mechanical damping effect in the actuator. Also, a high pass filter used to implement dc decoupling has been used in the feedback loop, which contributes to lower the amplitude of the actuator fundamental resonance. Between 60 Hz and 1 kHz, the FRF is characterized by a sequence of lightly damped plate resonances. At higher frequencies, there is a constant amplitude roll-off with frequency of the FRF, which is due to the inertial effect of the base disk and coil components of the actuator. At about 6 kHz, there is a deep trough followed by a sharp resonance peak. A finite element analysis of the base disk and cylindrical former has shown a resonance frequency in a similar frequency range. At higher frequencies, two other resonances are visible, which are due to the flexible modes of the circular springs.

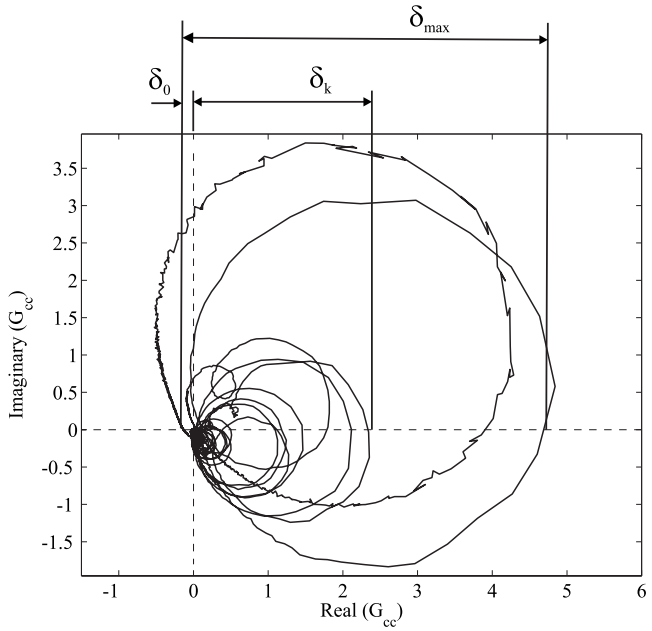


FIG. 9. Measured Nyquist plot of the open loop frequency response function between the error sensor signal and the input signal to the analog controller of the feedback loop.

The Nyquist plot in Fig. 9 shows the measured open loop sensor-actuator FRF. The plot shows a small semicircle on the left hand side which is due to the heavily damped fundamental resonance of the actuator and the high pass filter for the dc decoupling. There are then a number of circles on the right hand side which are due to the plate resonances of the low order modes of the plate. The phase lag effect generated by the inductive effect of the coil gradually drifts the loci toward the imaginary axis in the bottom right quadrant so that the large circle for the base disk and cylindrical former resonance is oriented along the imaginary negative axis. At higher frequencies, the locus enters the left hand side quadrants but, because of the roll-off effect produced by the inertia of the base disk and coil components of the actuator, the amplitude is relatively small compared to the low frequency part of the locus located on the positive real quadrants. In conclusion, the control system is bound to be conditionally stable with a maximum control gain that guarantees stability $g_{\max}=4.5$.

As discussed in Sec. II C, the Nyquist plot in Fig. 9 can also be used to estimate the maximum reduction of vibration at the control position that can be achieved at a certain resonance frequency of the panel. For the kth resonance frequency, the maximum reduction of vibration in decibels is given by $R_k=20\log_{10}|1+\delta_k/\delta_0|$, so that for the largest resonance circle at about 180 Hz the maximum reduction is predicted to be about 20 dB.

C. Closed loop control performance analysis

The plot in Fig. 10 shows the spectrum of velocity measured by the error sensor per unit excitation force. The faint line in the plot shows that the response at the error sensor when the actuator is not mounted on the panel is characterized by well separated sharp resonance peaks in nearly the whole frequency range considered. When the actuator is mounted on the plate, the resonance peaks are slightly

moved down in frequency and, more importantly, are rounded off by the passive damping effect produced by the actuator (thick line). At frequencies above the fundamental resonance frequency of the actuator, the actuator proof mass acts as an inertial reference so that, if the mechanical and electromechanical damping (due to the back emf in the coil) in the actuator are subcritical, additional passive damping is injected to the plate.[26] The constant roll-off is due to the inertial effect of the base disk and coil components of the actuator. When the feedback loop is closed with a gain margin of about 6 dB that guarantees stability and low control spillover effect around the fundamental resonance frequency of the proof mass actuator, as shown by the dotted line, between 80 and 250 Hz, the vibration at the error sensor goes down by 5–10 dB. However, at very low frequency, there is some control spillover effect which, as discussed in Sec. II A, is due to the low frequency smooth −180° phase transition of the actuation force per unit driving current signal.

V. CONCLUDING REMARKS

This paper has introduced the stability and performance study of a small scale prototype proof mass actuator for the implementation of a direct velocity feedback loop on a thin panel structure. A stability-performance formula that can be used to assess simultaneously the stability and control performance of the feedback control loop with the proof mass actuator has been derived from a mobility/impedance formulation. The principal downscaling and design issues of the proof mass actuator have then been analyzed in view of the stability requirements and control performance properties of the feedback loop. The principal outcome of this study has highlighted that the downscaling of a proof mass actuator using an electrodynamic linear motor produces both positive and negative effects. Among the positive effects are the reduction of the static displacement δ_a and stroke Δw and the increment of the damping ratio ζ_a, which scale, respectively, with $[L^2]$, $[L^2]$, and $[L^{-1}]$. Alternatively, among the negative effects are the rise of the fundamental natural frequency ω_a and decrement of the current driving the actuator i_a and the control f_a and transmitted f_c forces by the actuator, which scale, respectively, with $[L^{-1}]$, $[L^1]$ and $[L^2]$, $[L^2]$. Assuming the downscaling is limited to values such that the fundamental resonance frequency of the proof mass actuator is below the first resonance of the rectangular panel, the reduction of vibration for the maximum control gain that guarantees a stable feedback control loop remains constant. However, since the maximum current that can be fed to the actuator scales with $[L^1]$, also the control gain must be scaled with $[L^1]$ so that the control performance also falls down with power 2 of L as the size of the actuator is scaled down. This limitation is, however, mitigated by the fact that the number of control units that can be fitted per unit surface of the panel increases with power 2 of L as the actuators are scaled down so that the overall control performance with multiple decentralized control units would remain constant.

In the second part of the paper, the design of a small scale prototype actuator with a fundamental natural frequency $\omega_a = 20$ Hz well below the first resonance of the panel at about 73 Hz has been introduced. The stability and control performance of one direct velocity feedback control unit with such an actuator mounted on a thin rectangular panel which is clamped on a rigid frame have been assessed experimentally. When the feedback control loop is closed with a gain margin of 6 dB, so that there is little control spillover effect around the fundamental resonance of the actuator, reductions of vibration between 5 and 10 dB in the frequency band between 80 and 250 Hz have been measured at the control position.

ACKNOWLEDGMENTS

The work done by C.G.D. and C.P. for this project was supported by the "Early Stage Training site Marie Curie" program for the "European Doctorate in Sound and Vibration Studies" (EDSVS), which is funded by the European Commission. The prototype actuator studied in this paper was built in collaboration with Rene' Boonen at Katholieke Universiteit Leuven.

[1] D. J. Thompson and J. Dixon, "Vehicle noise," in *Advanced Applications in Acoustics, Noise and Vibration*, edited by F. J. Fahy and J. G. Walker (E & FN Spon, London, 2004), Chap. 6, pp. 236–291.
[2] J. S. Mixson and J. S. Wilby, "Interior noise," in *Aeroacoustics of Flight Vehicles, Theory and Practice*, edited by H. H. Hubbard (NASA Langley Research Center Hampton, Virginia, 1995), Chap. 16, pp. 271–335.
[3] P. Gardonio, "Review of active techniques for aerospace vibro-acoustic control," J. Aircr. **39**, 206–214 (2002).
[4] F. J. Fahy, "Fundamentals of noise and vibration control," in *Fundamentals of Noise and Vibration*, edited by F. J. Fahy and J. G. Walker (E & FN Spon, London, 1998), Chap. 5, pp. 255–309.
[5] M. J. Brennan and N. S. Ferguson, "Vibration Control," in *Advanced Applications in Acoustics, Noise and Vibration*, edited by F. J. Fahy and J. G. Walker (E & FN Spon, London, 2004), Chap. 12, pp. 530–580.
[6] S. J. Elliott, P. Gardonio, T. C. Sors, and M. J. Brennan, "Active vibroacoustic control with multiple local feedback loops," J. Acoust. Soc. Am. **111**, 908–915 (2002).
[7] F. J. Fahy and P. Gardonio, *Sound and Structural Vibration: Radiation, Transmission and Response*, 2nd ed. (Academic, London, 2007).
[8] E. Bianchi, P. Gardonio, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. Part III: Control system implementation," J. Sound Vib. **274**, 215–232 (2004).
[9] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, New York, 1996).
[10] P. Gardonio and S. J. Elliott, "Smart panels for active structural acoustic control," Smart Mater. Struct. **13**, 1314–1336 (2004).
[11] R. L. Clark, W. R. Saunders, and G. P. Gibbs, *Adaptive Structures*, 1st ed. (Wiley, New York, 1998).
[12] A. Preumont, *Vibration Control of Active Structures*, 2nd ed. (Kluwer, London, 2002).
[13] M. J. Balas, "Direct velocity feedback control of large space structures," J. Guid. Control **2**, 252–253 (1979).
[14] V. Jayachandran and J. Q. Sun, "Unconditional stability domains of structural control systems using dual actuator-sensor pairs," J. Sound Vib. **208**, 159–166 (1997).
[15] S. J. Elliott, M. Serrand, and P. Gardonio, "Feedback stability limits for active isolation systems with reactive and inertial actuators," J. Vibr. Acoust. **123**, 250–261 (2001).
[16] P. Gardonio and M. J. Brennan, "Mobility and impedance methods in structural dynamics," in *Advanced Applications in Acoustics, Noise and Vibration*, edited by F. J. Fahy and J. Walker (E & FN Spon, London, 2004), Chap. 9, pp. 387–388.
[17] L. Meirovitch, *Dynamics and Control of Structures* (Wiley, New York, 1990).
[18] O. N. Baumann and S. J. Elliott, "Destabilization of velocity feedback controllers with stroke limited inertial actuators," J. Acoust. Soc. Am. **121**, 211–217 (2007).
[19] P. Gardonio and S. J. Elliott, "Modal response of a beam with a sensor-actuator pair for the implementation of velocity feedback control," J.

Sound Vib. **284**, 1–22 (2005).
[20]M. J. Madou, *Fundamentals of Microfabrication: The science of Miniaturization*, 1st ed. (CRC, Boca Raton, FL, 1997).
[21]C. Paulitsch, "Vibration control with electrodynamic actuators," Ph.D. thesis, ISVR, University of Southampton, 2005.
[22]W. C. Young and R. G. Budynas, *Roark's Formulas for Stress and Strain*, 6th ed. (McGraw-Hill, New York, 1989).
[23]J. Peirs, "Design of micromechatronic systems: scale laws, technologies, and medical applications," Ph.D. thesis, Katholieke Universiteit Leuven, 2001.
[24]W. S. N. Trimmer, "Microrobots and micromechanical systems," Sens. Actuators **19**, 267–287 (1989).
[25]O. N. Baumann and S. J. Elliott, "The stability of decentralized multichannel velocity feedback controllers using inertial actuators," J. Acoust. Soc. Am. **121**, 188–196 (2007).
[26]C. Paulitsch, P. Gardonio, and S. J. Elliott, "Active vibration damping using an inertial, electrodynamic actuator (DETC2005-84632)," ASME J. Vibr. Acoust. **129**, 39–47 (2007).

Smart panel with active damping units. Implementation of decentralized control

Cristóbal González Díaz,[a] Christoph Paulitsch,[b] and Paolo Gardonio[c]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 9 July 2007; revised 1 May 2008; accepted 23 May 2008)

This paper contains the second part of a study on a smart panel with five decentralized velocity feedback control units using proof mass electrodynamic actuators [González Díaz *et al.*, J. Acoust. Soc. Am. **124**, 886 (2008)]. The implementation of five decentralized control loops is analyzed, both theoretically and experimentally. The stability properties of the five decentralized control units have been assessed with the generalized Nyquist criterion by plotting the loci of the eigenvalues of the fully populated matrix of frequency response functions between the five error signals and five input signals to the amplifiers driving the actuators. The control performance properties have been assessed in terms of the spatially averaged response of the panel measured with a scanning laser vibrometer and the total sound power radiated measured in an anechoic room. The two analyses have shown that reductions of up to 10 dB in both vibration response and sound radiation are measured at low audio frequencies, below about 250 Hz.
 [DOI: 10.1121/1.2945168]

PACS number(s): 43.40.Vn, 43.50.Ki, 43.40.At [KAC] Pages: 898–910

I. INTRODUCTION

This paper presents the second part of a study on the practical implementation of decentralized velocity feedback control on a thin aluminum rectangular panel with small scale proof mass electrodynamic actuators. In Ref. 1, the principal problems related to the design of a stable and effective velocity feedback loop using a small scale actuator have been presented. In particular, a down scaling study with reference to a "stability-performance" formula[1] has been introduced. The downscaling of the actuator produces contrasting results such as the desirable increment of internal damping and reduction of static deflection and the undesired increase of the actuator fundamental resonance frequency or reduction of the current that can be fed to the coil of the actuator so that the control force is also brought down. Thus, as the size of the actuator is scaled down, the feedback loop should be closed with increasingly smaller control gains, i.e., higher gain margins, so that the feedback loop becomes more stable and the control performance falls down. However, as the size of the actuator is reduced, an increasingly dense array of decentralized control units could be fitted to the panel, which should produce the same control performance but with more robust feedback control loops. Thus, the downscaling of the proof mass actuator should be carried out in such a way that (a) the fundamental resonance frequency of the actuator is kept below the first resonance of the panel, (b) a reasonable number of control units (with respect to geometrical, weight, and cost constraints) is required to obtain the desired control effect, and (c) the feedback loops are closed with large gain margins, so that the stability of the decentralized control system is robust to changes of the panel and actuator dynamic response. Based on these observations, a small scale prototype actuator has been designed, built, and studied. In particular, its stability and control performance have been analyzed experimentally.[1]

In this paper, the implementation on a thin rectangular panel of five decentralized feedback control units using five of these prototype small scale proof mass actuators is studied in detail. The passive and active effects of the five control units are analyzed both theoretically and experimentally. In particular, the stability of the control system is assessed since, although each control unit has been designed with a wide range of stable control gains, the stability of the set of five decentralized control units is constrained by the "cross talking" effects between the proof mass actuators.[2]

The stability and control performance of the smart panel with five decentralized control units are first analyzed theoretically in Sec. II using a mobility/impedance model. Section III briefly describes the prototype smart panel built for this study. The stability and control performance at the control locations of the five decentralized control units are then assessed experimentally in Sec. IV. Finally, in Sec. V the global control performance produced by the five decentralized control units is assessed in terms of the spatially averaged response of the panel, which has been measured with a scanning laser vibrometer, and in terms of the sound power radiated measured in an anechoic room.

II. STABILITY AND CONTROL PERFORMANCE THEORETICAL ANALYSIS

The smart panel considered in this paper is made of five decentralized velocity feedback control units, which, as shown in Fig. 1, are arranged along the diagonals of the

[a]Electronic mail: cgd@isvr.soton.ac.uk
[b]Electronic mail: cpaulits@gmx.de
[c]Electronic mail: pg@isvr.soton.ac.uk

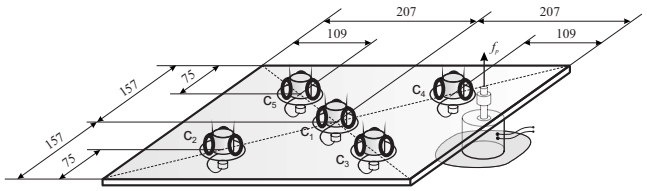


FIG. 1. Smart panel with five decentralized velocity feedback control units using proof mass electrodynamic actuators. The panel is excited by a shaker acting on the top left corner of the panel. Dimensions are in mm.

panel and at its center. The thin rectangular panel is clamped on a rigid frame placed at the top side of a rectangular box made with thick Plexiglas walls. The panel is excited by a shaker acting on the top right hand side corner or a speaker located inside the box. The details of the dimensions and material properties of the panel and control actuators are given in Tables I and II of Ref. 1.

A. Plate-actuator coupled model

The stability and control performance of the smart panel with five decentralized control units have been predicted with a mathematical model, which is based on a mobility/impedance formulation. In order to accurately predict the response of the prototype smart panel built for this study, the dynamic effects produced by all three mechanical components of the actuator have been taken into account. Thus, according to the sketch shown in Fig. 2, the effects of the housing and base disk mass have been accounted for, together with the dynamic effect of the proof mass and spring-dashpot suspension system. Also, the inertial effect of the mass of the force gauge and the moving part of the primary shaker has been taken into account.

The steady state response of the panel has been derived assuming the primary force disturbance to be harmonic, with time dependence of the form $\mathrm{Re}\{\exp(j\omega t)\}$ where ω is the circular frequency and $j=\sqrt{-1}$. The mechanical and electrical functions in the model have therefore been taken to be the real part of anticlockwise rotating complex vectors, i.e., phasor, given in the form $X(\omega)e^{j\omega t}$, where $X(\omega)$ is the phasor at $t=0$.

Considering the notation shown in Fig. 2, the phasors of the complex transverse velocities at the control positions, $\dot{w}_{cr}(\omega)$, and at the base of the actuators and primary shaker positions, $\dot{w}_{br}(\omega)$, have been grouped, respectively, into two column vectors $\dot{\mathbf{w}}_c(\omega)=\lfloor \dot{w}_{c1}(\omega)\cdots\dot{w}_{c5}(\omega)\rfloor^T$ and $\dot{\mathbf{w}}_b(\omega)=\lfloor \dot{w}_{b1}(\omega)\cdots\dot{w}_{b6}(\omega)\rfloor^T$. The flexural vibration at these error and base positions can be expressed by mobility matrix expressions in terms of the complex primary force, $f_p(\omega)$, and

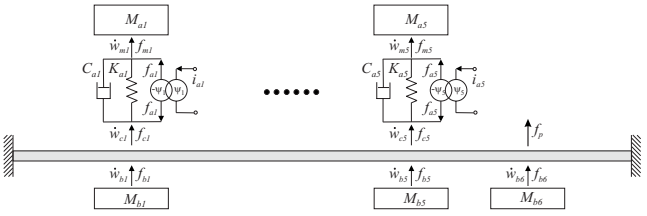


FIG. 2. Schematic with the notation for the velocity and force functions used in the mobility impedance model.

the column vectors $\mathbf{f}_c(\omega)=\lfloor f_{c1}(\omega)\cdots f_{c5}(\omega)\rfloor^T$ and $\mathbf{f}_b(\omega)=\lfloor f_{b1}(\omega)\cdots f_{b6}(\omega)\rfloor^T$ with the complex secondary excitations generated by the proof mass actuators, $f_{cs}(\omega)$, and the complex excitations generated by the inertia effects of the case and base masses of the actuators and primary shaker, $f_{bs}(\omega)$; so that

$$\dot{\mathbf{w}}_c = \mathbf{Y}_{cc}\mathbf{f}_c + \mathbf{Y}_{cb}\mathbf{f}_b + \mathbf{Y}_{cp}f_p, \tag{1}$$

$$\dot{\mathbf{w}}_b = \mathbf{Y}_{bc}\mathbf{f}_c + \mathbf{Y}_{bb}\mathbf{f}_b + \mathbf{Y}_{bp}f_p. \tag{2}$$

The elements in the various mobility matrices have been derived with the following modal expansions:

$$Y^{r,s}(\omega) = j\omega\sum_{n=1}^{N}\frac{\phi_n(x_r,y_r)\phi_n(x_s,y_s)}{M_p[\omega_n^2(1+j\eta)-\omega^2]}. \tag{3}$$

In this equation $M_p=\rho h l_x l_y$ is the mass of the plate, ρ is the density of the material of the plate, l_x, l_y, and h are, respectively, the dimensions and the thickness of the plate, and η is the loss factor. Also, ω_n and $\phi_n(x,y)$ are, respectively, the nth natural frequency and nth natural mode of the plate, which have been taken from Ref. 3 for a clamped panel.

Considering the lumped models of the actuators shown in Fig. 2, the velocities of the actuator proof masses, M_{ai}, and the forces generated by the base masses, M_{bi}, can also be expressed in terms of the following mobility and impedance matrix relations:

$$\dot{\mathbf{w}}_m = \mathbf{Y}_{mm}\mathbf{f}_m, \tag{4}$$

$$\mathbf{f}_b = -\mathbf{Z}_b\dot{\mathbf{w}}_b, \tag{5}$$

where $\dot{\mathbf{w}}_m(\omega)=\lfloor \dot{w}_{m1}(\omega)\cdots\dot{w}_{m5}(\omega)\rfloor^T$ and $\mathbf{f}_m(\omega)=\lfloor f_{m1}(\omega)\cdots f_{m5}(\omega)\rfloor^T$ are the vectors, respectively, with the complex velocities and forces acting on the proof masses of the actuators; $\mathbf{Y}_{mm}$ is a diagonal matrix with the mobilities of the actuators proof masses; $Y_{m,i}(\omega)=1/j\omega M_{ai}$; and $\mathbf{Z}_b$ is a diagonal matrix with the impedances of the actuator base and case masses $Z_{b,i}(\omega)=j\omega M_{bi}$.

Substituting Eq. (5) into Eq. (2), the vector with the velocities of the base masses is given by

$$\dot{\mathbf{w}}_b = \mathbf{Q}_{bc}\mathbf{f}_c + \mathbf{Q}_{bp}f_p, \tag{6}$$

where $\mathbf{Q}_{bc}=(\mathbf{I}+\mathbf{Y}_{bb}\mathbf{Z}_b)^{-1}\mathbf{Y}_{bc}$ and $\mathbf{Q}_{bp}=(\mathbf{I}+\mathbf{Y}_{bb}\mathbf{Z}_b)^{-1}\mathbf{Y}_{bp}$ and $\mathbf{I}$ is a 6×6 identity matrix. Also, substituting this equation into Eq. (5), and then into Eq. (1), the vector with the velocities at the control positions is obtained as follows:

$$\dot{\mathbf{w}}_c = \mathbf{Q}_{cc}\mathbf{f}_c + \mathbf{Q}_{cp}f_p, \tag{7}$$

where $\mathbf{Q}_{cc}=\mathbf{Y}_{cc}-\mathbf{Y}_{cb}\mathbf{Z}_b\mathbf{Q}_{bc}$ and $\mathbf{Q}_{cp}=\mathbf{Y}_{cp}-\mathbf{Y}_{cb}\mathbf{Z}_b\mathbf{Q}_{bp}$. Equations (7) and (4) can be compiled in one mobility equation as follows:

$$\dot{\mathbf{w}} = \mathbf{Y}_c\mathbf{f} + \mathbf{Y}_p f_p, \tag{8}$$

where

$$\dot{\mathbf{w}} = \begin{Bmatrix} \dot{\mathbf{w}}_c \\ \dot{\mathbf{w}}_m \end{Bmatrix}, \tag{9a}$$

$$\mathbf{Y}_c = \begin{bmatrix} \mathbf{Q}_{cc} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_m \end{bmatrix}, \tag{9b}$$

$$\mathbf{f} = \begin{Bmatrix} \mathbf{f}_c \\ \mathbf{f}_m \end{Bmatrix}, \tag{9c}$$

$$\mathbf{Y}_p = \begin{Bmatrix} \mathbf{Q}_{cp} \\ \bar{\mathbf{0}} \end{Bmatrix}, \tag{9d}$$

where $\mathbf{0}$ is a 5×5 matrix of zeros and $\bar{\mathbf{0}}$ is a 5×1 vector of zeros. The vector with the forces transmitted to the plate and to the actuator proof masses can be expressed in terms of the following impedance matrix expression:

$$\mathbf{f} = -\mathbf{Z}_c\dot{\mathbf{w}} + \boldsymbol{\Psi}_a\mathbf{I}_a, \tag{10}$$

with

$$\mathbf{Z}_c = \begin{bmatrix} \mathbf{Z}_{kc} & -\mathbf{Z}_{kc} \\ -\mathbf{Z}_{kc} & \mathbf{Z}_{kc} \end{bmatrix} \text{ and } \boldsymbol{\Psi}_a = \begin{bmatrix} +\boldsymbol{\Psi}_{fa} \\ -\boldsymbol{\Psi}_{fa} \end{bmatrix}, \tag{11}$$

where the elements in the diagonal matrix $\mathbf{Z}_{kc}$ are given by the impedance of the spring-damper mounting system, $Z_{kc,i}(\omega)=K_{a,i}/j\omega+C_{a,i}$, and the elements in the diagonal matrix $\boldsymbol{\Psi}_{fa}$ are given by the voice coil coefficients of the actuators, $\Psi_{fa,i}=\psi_i$. The column vector $\mathbf{I}_a(\omega)=\lfloor i_{a1}(\omega)\cdots i_{a5}(\omega)\rfloor^T$ contains the current control signals driving the five proof mass actuators. At this point, substituting Eq. (10) into Eq. (8), the vector with the velocities both at the base and proof mass positions of each actuator is found as follows:

$$\dot{\mathbf{w}} = \mathbf{T}_a\mathbf{I}_a + \mathbf{T}_p f_p, \tag{12}$$

where the matrices $\mathbf{T}_a$ and $\mathbf{T}_p$ are given by $\mathbf{T}_a=(\mathbf{I}+\mathbf{Y}_c\mathbf{Z}_c)^{-1}\mathbf{Y}_c\boldsymbol{\Psi}_a$ and $\mathbf{T}_p=(\mathbf{I}+\mathbf{Y}_c\mathbf{Z}_c)^{-1}\mathbf{Y}_p$. The velocities at the five control positions can be extracted from Eq. (12) by premultiplying it by $\mathbf{U}=[\mathbf{I}\ \mathbf{0}]$, where $\mathbf{I}$ is a 5×5 identity matrix and $\mathbf{0}$ is a 5×5 matrix of zeros, that is,

$$\dot{\mathbf{w}}_c = \mathbf{G}_{ca}\mathbf{I}_a + \mathbf{G}_{cp}f_p, \tag{13}$$

where $\mathbf{G}_{ca}=\mathbf{U}\mathbf{T}_{ca}$ and $\mathbf{G}_{cp}=\mathbf{U}\mathbf{T}_{cp}$.

When the five feedback control loops are closed independently with the same constant feedback gains, the control current signals are given by

$$\mathbf{I}_a = -\mathbf{H}\dot{\mathbf{w}}_c, \tag{14}$$

where $\mathbf{H}$ is a diagonal matrix with the five control gains; $g_1,\ldots,g_5$. The response when the five decentralized feedback control loops are closed can be described with the multiple input multiple output (MIMO) rejection feedback block diagram shown in Fig. 3 so that

$$\dot{\mathbf{w}}_c = (\mathbf{I}+\mathbf{G}_{ca}\mathbf{H})^{-1}\mathbf{G}_{cp}f_p. \tag{15}$$

Thus, substituting Eq. (15) into Eq. (14) and then Eq. (14) into Eq. (12), the vector with the velocities both at the base and proof mass positions of each actuator can be found. This vector can then be substituted into Eq. (10) to get the vector with the forces at the base and proof mass positions of each actuator so that the vector with the forces at the five control positions can be expressed as follows:

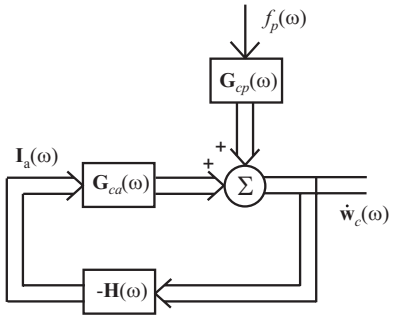


FIG. 3. Block diagram of the multichannel feedback control system implemented on the plate.

$$\mathbf{f}_c = -\mathbf{F}_{cp}f_p, \tag{16}$$

where $\mathbf{F}_{cp}=\mathbf{U}[(\boldsymbol{\Psi}_a-\mathbf{Z}_c\mathbf{T}_a)\mathbf{H}(\mathbf{I}+\mathbf{G}_{ca}\mathbf{H})^{-1}\mathbf{G}_{cp}+\mathbf{Z}_c\mathbf{T}_p]$. Also, the vector with the forces generated by the inertia effects of the case and base masses of the actuators and by the inertia effect of the force gauge and moving parts of the primary shaker is obtained by substituting Eq. (16) into Eq. (6) and then Eq. (6) into Eq. (5) as follows:

$$\mathbf{f}_b = \mathbf{F}_{bp}f_p, \tag{17}$$

where $\mathbf{F}_{bp}=\mathbf{Z}_b(\mathbf{Q}_{bc}\mathbf{F}_{cb}-\mathbf{Q}_{bp})f_p$.

The overall vibration of the plate can be assessed in terms of its total kinetic energy which, for the plate considered in this paper, is given by the following formula:[4]

$$T(\omega) = \frac{1}{4}\int_A \rho h|\dot{w}(x,y,\omega)|^2 dA, \tag{18}$$

where $\dot{w}(x,y,\omega)$ is the complex transverse velocity over the plate surface. The flexural vibration on a generic point of the panel due to the primary force, f_p, and the vectors with the control and base excitations, $\mathbf{f}_c$ and $\mathbf{f}_b$, generated by the proof mass actuators can be expressed with the following matrix relation:

$$\dot{w}(x,y) = \boldsymbol{\phi}\mathbf{A}_c\mathbf{f}_c + \boldsymbol{\phi}\mathbf{A}_b\mathbf{f}_b + \boldsymbol{\phi}\mathbf{a}_p f_p, \tag{19}$$

where $\boldsymbol{\phi}=\lfloor\phi_1(x,y)\cdots\phi_R(x,y)\rfloor$ is a row vector with the amplitudes of the modes at the generic point (x,y), $\mathbf{A}_c$ and $\mathbf{A}_b$ are the matrices with the complex modal excitations functions generated by the control and base forces, respectively,

$$A_{c_{r,s}} = \frac{\phi_r(x_{cs},y_{cs})}{\mathbf{M}_p[\omega_r^2(1+j\eta)-\omega^2]}, \tag{20a}$$

$$A_{b_{r,s}} = \frac{\phi_r(x_{bs},y_{bs})}{\Lambda[\omega_r^2(1+j\eta)-\omega^2]}, \tag{20b}$$

and $\mathbf{a}_p$ is a column vector with the complex modal excitation functions generated by the primary force excitation

$$a_{p_r} = \frac{\phi_r(x_p,y_p)}{\mathbf{M}_p[\omega_r^2(1+j\eta)-\omega^2]}. \tag{21}$$

Using Eqs. (15) and (16), the kinetic energy of the plate is found to be

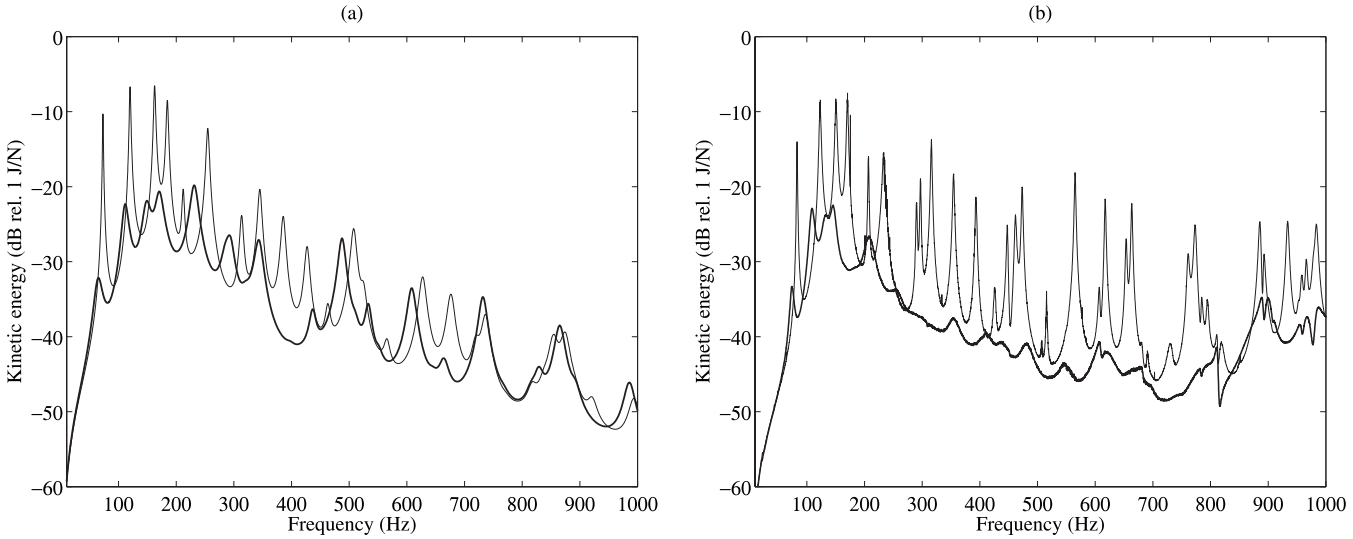


FIG. 4. Kinetic energy per unit primary force of the panel with no control units (faint line) and with five control units (thick line). (a) Theoretical prediction; (b) experimental measurement with a scanning laser vibrometer.

$$T(\omega) = \tfrac{1}{4} M_p f_p^* [\mathbf{a}_p + \mathbf{A}_c \mathbf{F}_{cp} + \mathbf{A}_b \mathbf{F}_{bp}]^H \times [\mathbf{a}_p + \mathbf{A}_c \mathbf{F}_{cp} + \mathbf{A}_b \mathbf{F}_{bp}] f_p. \quad (22)$$

This equation can be used to derive the response of the plate for various control gains of the five decentralized feedback control loops.

B. Passive effects of the actuators

Before analyzing the control performances produced by the five feedback control loops, it is interesting to consider the passive effects produced by the actuators. Paulitsch *et al.*[5] have shown that, for moderate levels of internal damping in the actuator, above the fundamental resonance of the actuator, the proof mass vibration goes progressively to zero, so that the mass acts as a sky hook reference. As a result, the internal damping of the actuator becomes a sort of sky hook damping effect on the panel, which should damp down the response of the panel at the low frequency resonances. Also, the masses of the base disk, and coil system of each actuator are directly attached to the plate. Thus, they produce a substantial mass effect, which tends to reduce the response of the structure as the frequency increases.

The primary shaker is attached to the plate via a force gauge, which also introduces a mass effect on the plate together with the moving parts of the shaker.

The passive effects produced by all these components can be readily analyzed with the model introduced in the previous section. The plot (a) in Fig. 4 shows the kinetic energy of the panel when there are no control units and when the five decentralized control units are mounted on the panel. The plot (b) shows the same type of result, where the kinetic energy has been estimated from the measured response of the panel over a grid of points taken with a laser vibrometer. When the five control actuators are mounted on the panel, the response tends to fall down more rapidly with frequency than in the case with no actuators because of the inertial effects of the base disk and coil system of each actuator. Also, the resonance peaks are smoothed down by the internal damping effect of the actuators transmitted to the plate. Comparing the simulation results shown in Fig. 4(a) and the experimental results shown in Fig. 4(b), a good agreement is noted for both cases with and without actuators. In the case with actuators, there are more discrepancies between the measured and experimental results, which depend on the fact that the components of the actuators have been modeled in terms of lumped parameter elements although their size and shapes certainly have an effect on the response of the panel also at low audio frequency. Moreover, the model does not take into account the rotational inertia effects of the actuator masses, which are likely to be rather important, particularly at higher audio frequencies. Nevertheless, the model has certainly captured the most important features that characterize the response of the smart panel with five decentralized control units and can be used profitably to predict the response of the control system and its performances.

C. Stability of the five channel feedback control system

The control performance of the smart panel strongly depends on the stability of the five decentralized control units, González Diaz *et al.*[1] have shown that even for the single control unit the feedback loop is only conditionally stable. The problem arises from the low frequency dynamics of the actuator, which is characterized by a resonance that introduces a −180° phase shift of the control force that is the principal cause of the instability in the feedback loop. Although the five control units implement decentralized control loops, their stability has to be reassessed since, as discussed by Baumann and Elliott[2] the stability of each loop is influenced by the vibration generated by neighboring actuators.

The stability of multiple channel feedback control loops can be analyzed with the generalized Nyquist stability criterion, which states that, assuming that both the plant and controller are individually stable, a multichannel feedback system is bound to be stable provided the locus of $\det[\mathbf{I} + \mathbf{G}_{ca}\mathbf{H}] = 0$ does not encircle the instability point $(0, j0)$ as ω

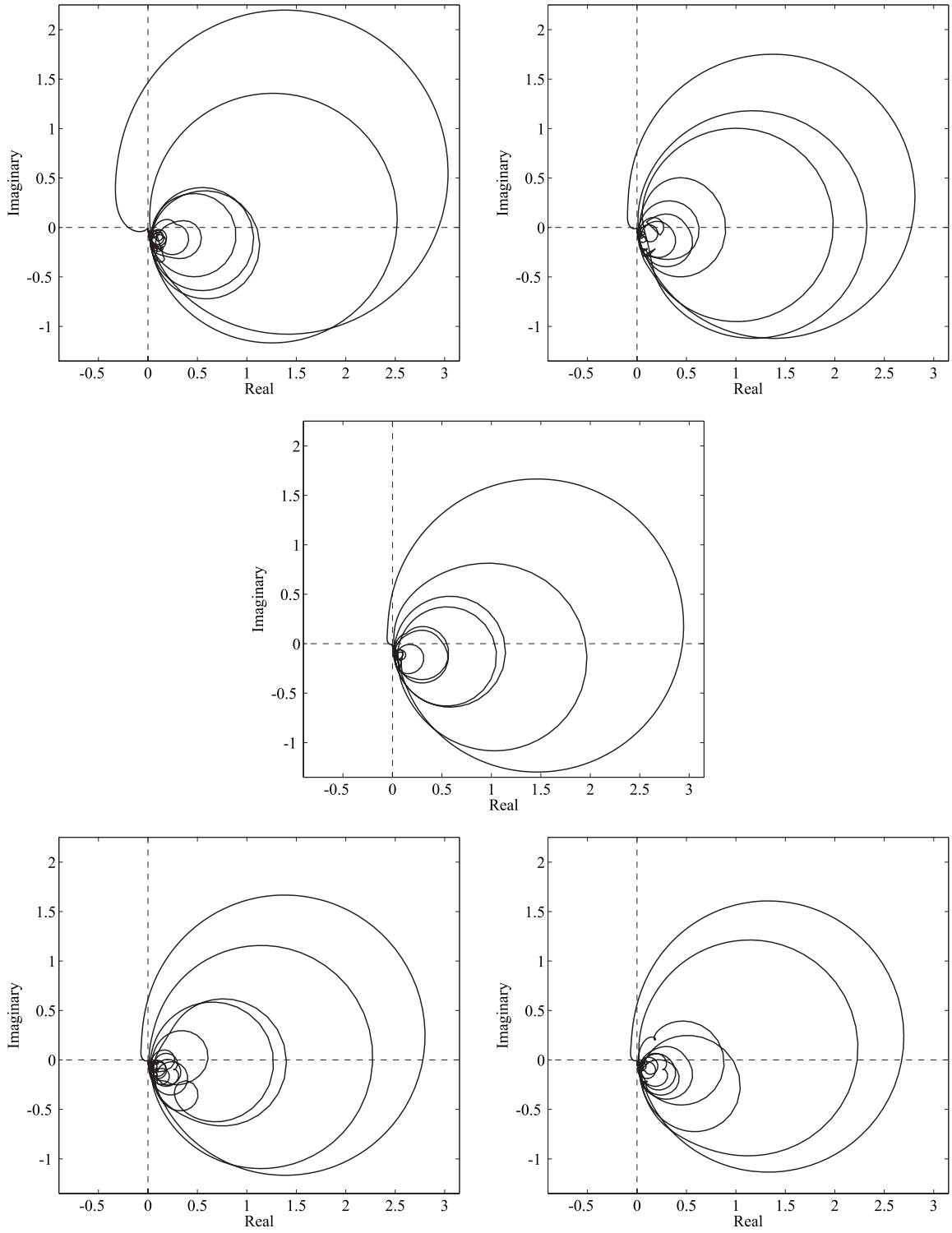


FIG. 5. Loci of the eigenvalues of the 5×5 matrix of sensor-actuator FRFs $\mathbf{G}_{ca}\mathbf{H}$ simulated between 5 Hz and 50 kHz.

varies from $-\infty$ to $+\infty$.[6] Thus, for the case of decentralized control with the same control gains for all feedback loops, so that $\mathbf{H}$ is a diagonal matrix, the stability of the control loop can be assessed by considering the fully populated matrix of frequency response functions (FRFs) between the five control velocities and the five input current signals to the controller driving each actuator. Moreover, the determinant of a matrix is the product of its eigenvalues;[6] that is, $\det[\mathbf{I}+\mathbf{G}_{ca}\mathbf{H}]=(1+g\lambda_1)(1+g\lambda_2)\cdots(1+g\lambda_5)$, where $\lambda_i(\omega)$ is the ith eigenvalue of $\mathbf{G}_{ca}\mathbf{H}$. Thus, the stability analysis of the five channel control system can be implemented with reference to the polar plots of the five eigenvalues of $\mathbf{G}_{ca}\mathbf{H}$. In this case, in order to ensure the system is stable, the five loci should not encircle the instability point $(-1,j0)$ as ω varies from $-\infty$ to $+\infty$.

The five plots in Fig. 5 show the loci of the five eigenvalues of the simulated matrix of sensor-actuator FRFs $\mathbf{G}_{ca}\mathbf{H}$. They are all characterized by a low frequency loop

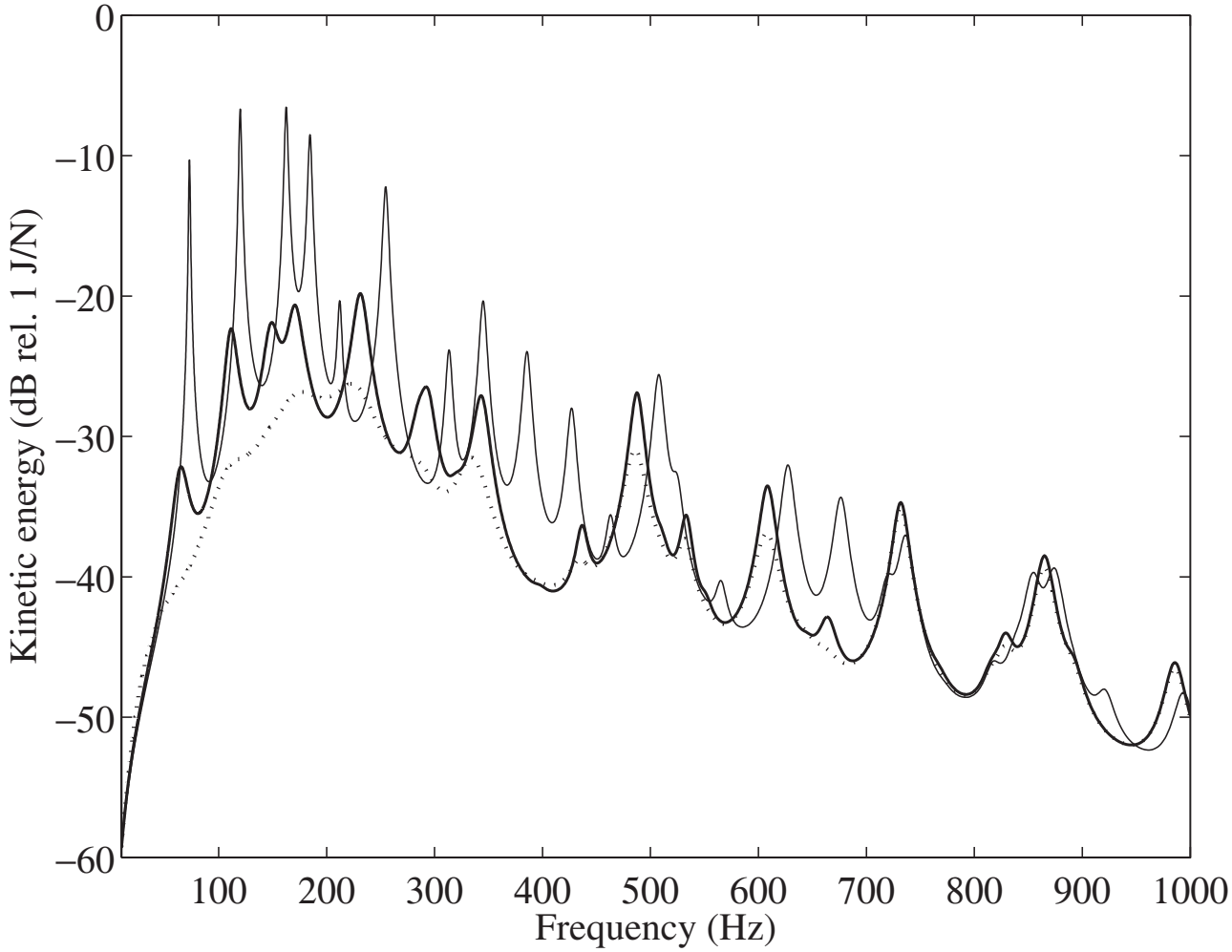


FIG. 6. Simulated kinetic energy per unit primary force of the panel with no control units (faint line), with five control units when the feedback loops are left open (thick line), and when the five feedback loops are closed with the maximun control gain that guarantees stability (dotted line).

that starts in the left hand side quadrants and then enters into the right hand side quadrants. There are then a series of smaller circles located across the real positive axis. The overall picture resulting from the analysis of the eigenvalues confirms that the set of five decentralized feedback control loops is only conditionally stable. The maximum control gain that guarantees stability is $g_{\max}=4.3$. The stability limit is dictated by the low frequency part of the eigenvalues, which is characterized by the low frequency dynamics of the actuators that produce positive, rather than negative, velocity feedback effects. As discussed by Baumann and Elliott,[2] this positive feedback effect is also the principal cause of cross talking between actuators, which enhances the instability problems with multiple feedback loops.

D. Control performances

The performance of the five channel control system has been assessed with reference to the maximum control gain $g_{\max}=4.3$, which guarantees stability. The plot in Fig. 6 shows the simulated kinetic energy of the panel and five control units when there are no actuators on the panel (faint line), when there are five actuators on the panel (thick line), and when the five actuators implement the maximum control gain that guarantees stability (dotted line). This plot suggests that the five control units can produce reductions between 3 and 10 dB at low frequencies up to 250 Hz. This result cannot be considered outstanding on its own. However, if both passive and active reductions of the vibration are considered, then, comparing the kinetic energy of the panel when there are no actuators on the panel (faint line) and the kinetic energy of the panel when the five actuators are implementing feedback loops (dotted line), it can be concluded that reductions between 10 and 20 dB or more of the low frequency resonances of the panel are generated. It is important to highlight that both the passive and active effects produced by the control units occur at low frequencies where, in fact, the spectrum of the response is maximum. Thus, the control effect can be interpreted as a reduction of the maximum level of the kinetic energy spectrum from −8 to −27 dB, which is produced by the passive effects of the actuators (from −8 to −20 dB) and by the active effects of the control loops (from −20 to −27 dB).

Therefore, it can be concluded that the implementation of five decentralized velocity feedback control loops using proof mass actuators is limited by instability issues generated by the low frequency dynamics of the actuators. Also, the

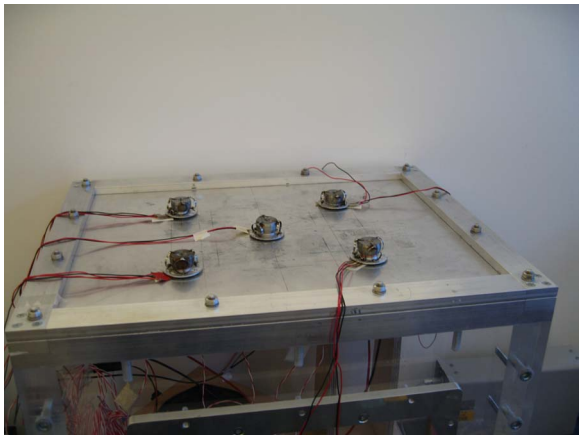


FIG. 7. (Color online) Photograph of the box test rig and smart panel with five proof mass actuators for the implementation of decentralized velocity feedback loops.

performance of the control system is a combination of passive and active effects. In fact, there is a rather effective passive reduction of the vibration due to the internal damping of the actuator which, if it is not high enough to lock the proof mass to the housing of the actuator, it is then exerted to the plate since the proof mass acts as a sky hook reference at frequencies above the fundamental resonance of the actuator. In addition to this passive effect, there is an additional active effect that further brings down the response of the panel.

These results indicate that the use of inertial control actuators may represent a rather good control mean, particularly for low frequency disturbances that, normally, requires bulky and heavy passive treatments to be controlled. The active arrangement proposed in this paper may provide a much more convenient approach, particularly when large numbers of small scale control units are to be used.

III. PROTOTYPE SMART PANEL WITH FIVE DECENTRALIZED CONTROL UNITS

Considering the theoretical analysis of the control performance of five decentralized control units, presented in Sec. II of this paper, a prototype panel has been built with five decentralized control units.

As shown in Fig. 7, the panel is mounted over a rigid aluminum frame, which has been fixed on the top open side of a rectangular box with thick Perspex walls. With this arrangement, the panel can be excited both acoustically, with the loudspeaker placed in the cavity, or mechanically, by the shaker that is mounted on a wooden stand located in the background right hand side corner of the box. The thick walls of the box are made of a light material so that its first resonance frequency occurs at a relatively higher frequency than the frequency range considered in the measurement. In this way, the flanking sound radiation through the sidewalls of the box is relatively lower than that through the top smart panel.[7] As a result, it has been possible to assess the effective sound radiation by the panel with and without control.

As shown in Fig. 7, five actuators have been built and mounted on the aluminum panel. For each actuator, an accelerometer sensor has been fixed on the other side of the panel in correspondence to the axis of the actuator itself. In order to get the desired negative velocity feedback control loops for each control unit, the output signal from the accelerometer has been inverted, integrated, and amplified with an analog control system and then fed back to the actuator. Additionally, in order to reduce the dc electrical coupling a high-pass filter with a cut-off frequency of about 10 Hz is used in the feedback loop.

IV. STABILITY AND CONTROL PERFORMANCE TESTS

The operation of the smart panel has been tested in two stages. First, the stability of the control system has been evaluated by measuring the 5×5 fully populated matrix of FRFs between the outputs of the five error sensors, passed through analog integrators to get the velocity signals, and the five input signals to the analog control amplifiers driving the proof mass actuators. The maximum stable gain has then been derived from the five plots with the loci of the 5×5 sensor-actuator matrix of FRFs.

Second, the implementation and performance of the five decentralized control units have been assessed by plotting the maximum reductions generated at the control positions when the maximum control gains are implemented.

A. Stability analysis

Figure 8 shows the loci of the eigenvalues of the 5×5 matrix of sensor-actuator FRFs measured between 5 Hz and 28 kHz. The loci are slightly irregular because they have been calculated from the measured data over a particularly large frequency range. The plots show similar characteristics to those of the loci predicted from the simulated matrix of sensor-actuator open loop FRFs. The low frequency portions of the loci are characterized by rather big loops, which start in the left hand side quadrants and then drift towards the right hand side quadrants. These loops appear slightly rotated in the clockwise direction. This is the result of the dc decoupling implemented in the controllers with a high-pass filter with cut-off frequency of about 10 Hz. As predicted in the simulation study, there are then a certain number of smaller loops, which are approximately located across the real positive axis.

B. Control performance

The control performance of the smart panel has been tested with gain margins of 7.6 dB that guarantees stability and little control spillover effects at low frequency. The passive and active effects produced at the five error sensors per unit primary force excitation are plotted in Fig. 9. As discussed in Sec. II B, the passive damping generated by the actuators on the panel effectively smooth out the resonance peaks. Also, the base and coil components of the actuators produce an inertial effect, which further reduces the response of the panel as the frequency rises. The active control performances vary according to the location of the control units. The control unit N.4, which is closely located to the primary excitation, is the one that shows smaller vibration reductions. This is probably due to the primary excitation near field,

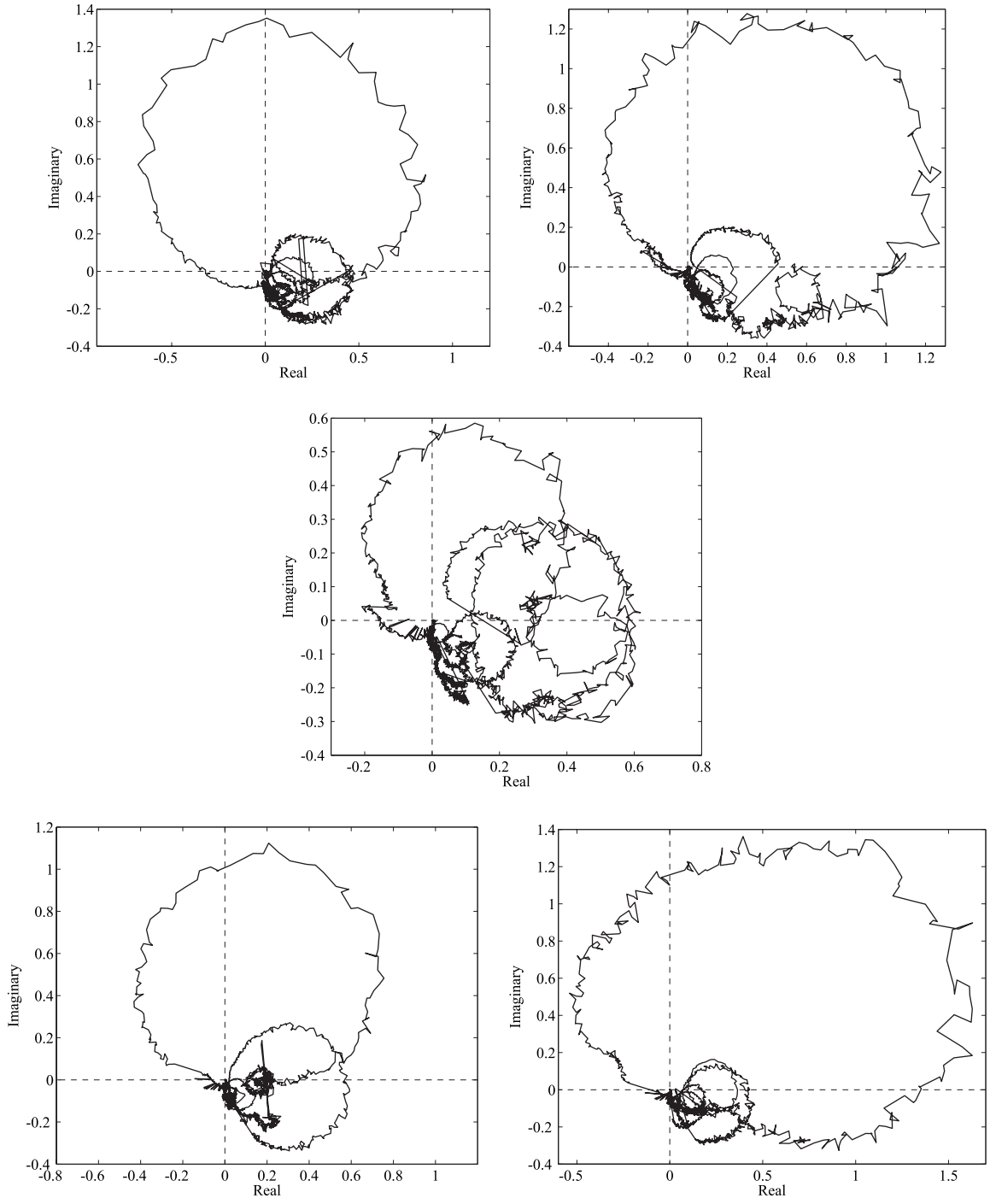


FIG. 8. Loci of the eigenvalues of the 5×5 matrix of sensor-actuator FRFs measured between 5 Hz and 28 kHz.

which cannot be effectively controlled by active damping. In contrast, the two control units located further away from the primary excitation, i.e., control units N.2 and N.5, show relatively higher performances, possibly because the near field effect of the primary excitation to the panel tends to be relatively lower in those points and the active damping action efficiently reduces the steady state response at the control position.

Considering the control units N.2, N.3, and N.5 that produce the largest control effects, reductions between 20 and 30 dB up to 550 Hz are measured by the error sensors. In contrast, the control units N.1 and N.4 produce relatively small reductions between 5 and 10 dB up to 550 Hz. It is important to note that this uneven performance result does not represent *a priori* a bad design of the control system. In fact, it is a prerogative of this design to implement a robust control scheme that does not set the control actuators to minimize an optimal function, but instead implement active damping over the panel regardless of the positions of the actuators. Thus, it is not a surprise that some control units will be more efficient than the others depending on their position and the type and location of the primary excitation. Nevertheless, when a large enough number of systems are used a good control effect should be produced over the de-

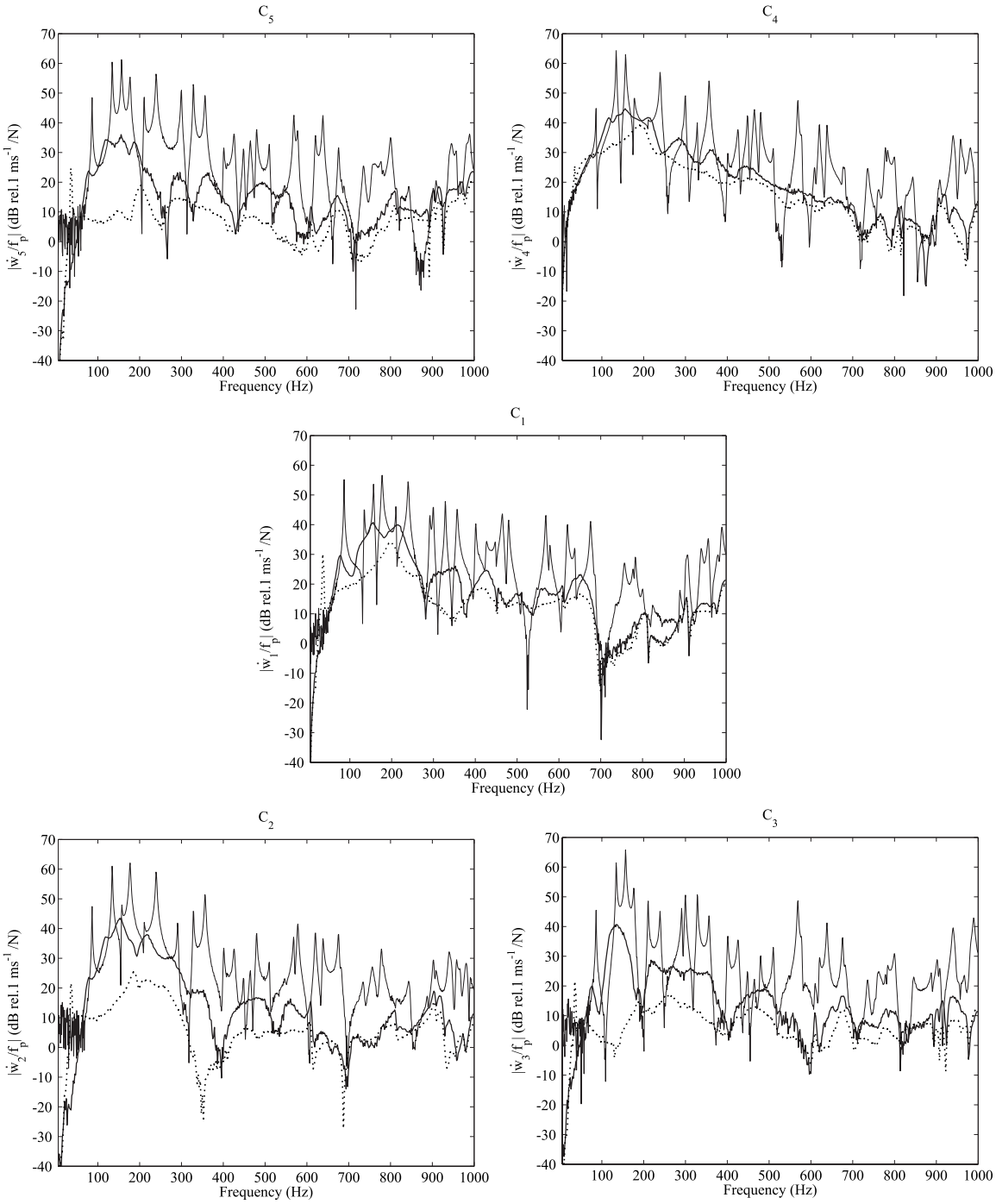


FIG. 9. Measured velocities at the five error sensors per unit force excitation to the plate with no actuators (faint lines), with actuators and no control (thick lines), and with actuators implementing the maximum control gains that guarantee stability (dotted lines).

sired control frequency band.

V. GLOBAL CONTROL EFFECTS PRODUCED BY THE SMART PANEL

The global control performance produced by the smart panel with the five decentralized control units implementing feedback loops with gain margins of 7.6 dB has been assessed in two ways. First, with the kinetic energy of the panel derived from the spatially averaged response of the panel measured with a scanning laser vibrometer. Second, with the total sound power radiated derived from the measured sound pressure in nine positions around the box in an anechoic room, according to the standard procedure described by the ISO 3744 guidelines. These measurements have been taken for two cases where the panel is excited either by the shaker force actuator, which is also located within the cavity, or the acoustic field generated in the box by a loudspeaker source.

The measurements taken with the laser vibrometer show the control performance produced by the five control units on the response of the panel. Also, they provide an indication about the near field sound radiation generated by the panel.[4]

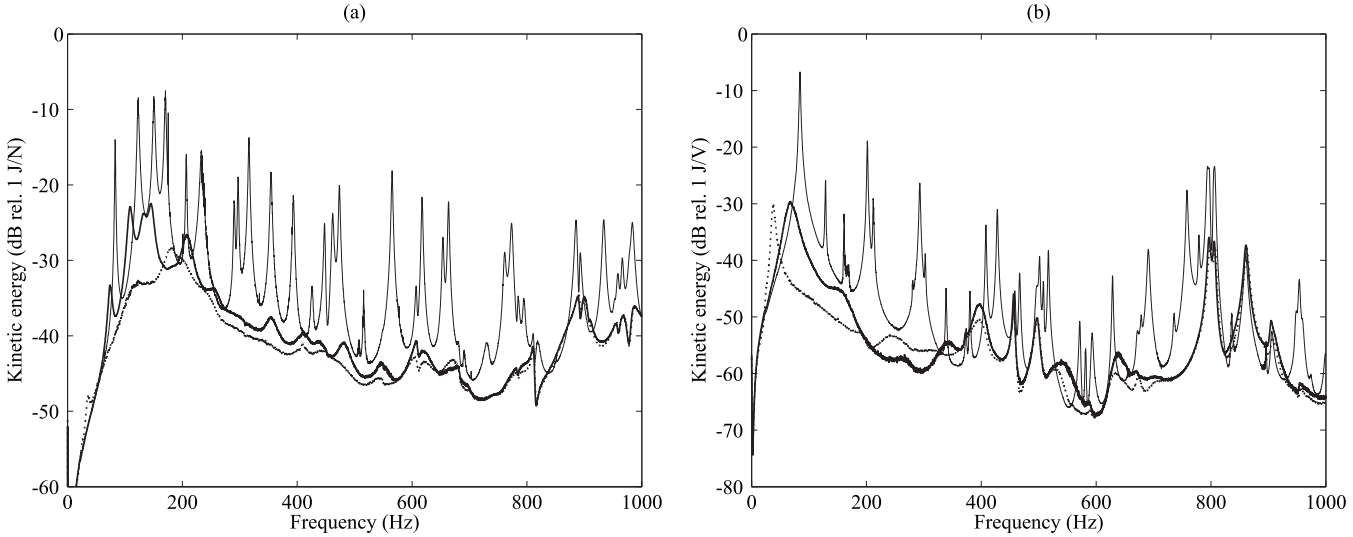


FIG. 10. Measured narrowband spectra of kinetic energy of the panel derived from the spatially averaged response of the panel measured with a scanning laser vibrometer when the panel is excited by the shaker (a) and by the loudspeaker in the cavity (b). Faint line: response of the panel with no control units; thick line: response of the panel with the five control units; dotted line: response of the panel with the five control units implementing decentralized velocity feedback control.

The measurement taken in the anechoic room gives instead an indication of the far field mean sound radiation produced by the panel.

The two types of primary excitations have been chosen in such a way as to assess the control performance produced by the five control units when most of the structural modes are efficiently excited by the shaker point force excitation on the panel or only few volumetric structural modes are efficiently excited by the acoustic field in the cavity generated by the loudspeaker.

A. Kinetic energy of the panel measured with laser vibrometer

The two plots in Fig. 10 show the narrowband spectra of the kinetic energy of the panel derived from the spatially averaged response of the panel measured with a scanning laser vibrometer. The response of the panel when there are no actuators mounted on the panel (faint line) is characterized by very lightly damped resonances in both cases where the panel is excited by the shaker or loudspeaker primary sources. As seen in Sec. II B, when the five control actuators are mounted on the panel, they produce a rather high passive damping and mass effects on the panel, which efficiently reduces the response of the panel (thick lines). The passive effects produced by the five control units generate very high reductions of the response comprised in a range between 10 and 20 dB. Figure 10(b) shows a number of peaks, for example, those at 500, 820, and 860 Hz, which are little affected by the presence of the five actuators. These resonances are not structural resonances; they are instead acoustic resonances of the cavity underneath the panel, which are efficiently excited by the loudspeaker source placed in the cavity.

The dotted lines in the two plots of Fig. 10 show the response of the panel in the two primary excitation cases when the five control units are used to implement decentralized velocity feedback control loops. Both plots show that the active system further reduces the response of the panel by about 3–10 dB at the low frequencies below about 250 Hz.

Although the active control effect may look not so significant compared to the passive effect introduced by the control system, it is important to emphasize that the active control reduction of vibration occurs at low frequency where the response of the panel is relatively larger than that at mid and high audio frequencies. The passive reduction of vibration at low frequency with passive means is a challenging problem that often cannot be solved unless bulky and heavy passive treatments are applied on the structure. The combined passive and active effects produced by five control units effectively cover the whole frequency range up to 1 kHz. This is clearly shown in the third octave plots of Fig. 11 where the significance of the additional reduction of vibration introduced by the feedback loops on top of the passive effects can be assessed by comparing the levels of the center and right bars. For instance, considering the smart panel excited by the shaker, the insertion of the five control actuators brings down the maximum response of the panel from 74 to 62 dB. When the active control system is also turned on, the maximum response of the panel is further reduced to 58 dB. Similarly, for the system excited by the loudspeaker, the maximum response of the panel falls down from 73 to 58 dB when the five control units are mounted on the panel and then to 49 dB when the control units are turned on.

Figure 12 shows the response of the panel at the 123.4 and 84.4 Hz resonance frequencies when the panel is excited, respectively, by the shaker and loudspeaker primary sources (note the change of scale between the top and center bottom plots). Figures 12(a) and 12(b) show the efficacy of both passive and active effects on the panel. The passive action of the five control actuators effectively reduces the response of the panel, which is controlled by the (2,1) mode at 123.4 Hz and by the (1,1) mode at 84.4 Hz. Then, the active action of the five control units further reduces the

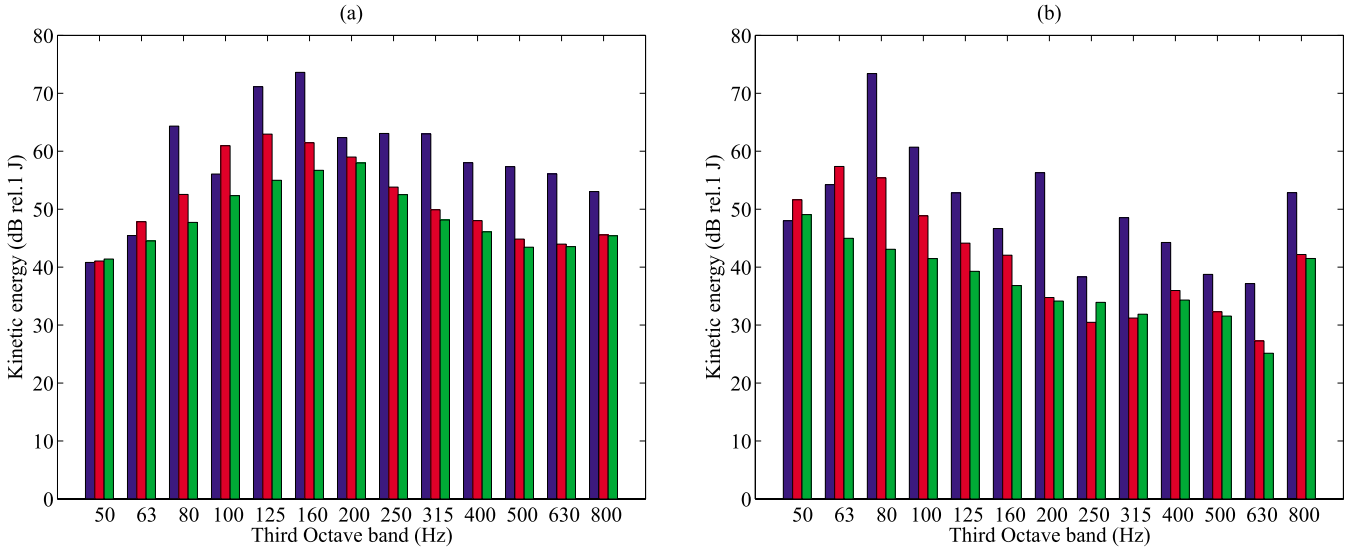


FIG. 11. (Color online) Measured third octave band spectra of the kinetic energy of the panel derived from the spatially averaged response of the panel measured with a scanning laser vibrometer when the panel is excited by the shaker (a) and by the loudspeaker in the cavity (b). Left bar: response of the panel with no control units; center bar: response of the panel with the five control units; right bar: response of the panel with the five control units implementing decentralized velocity feedback control.

responses in such a way as the patterns of the (2,1) and (1,1) modes are nearly completely erased so that the responses are controlled by residual modes and show a rather uniform patterns.

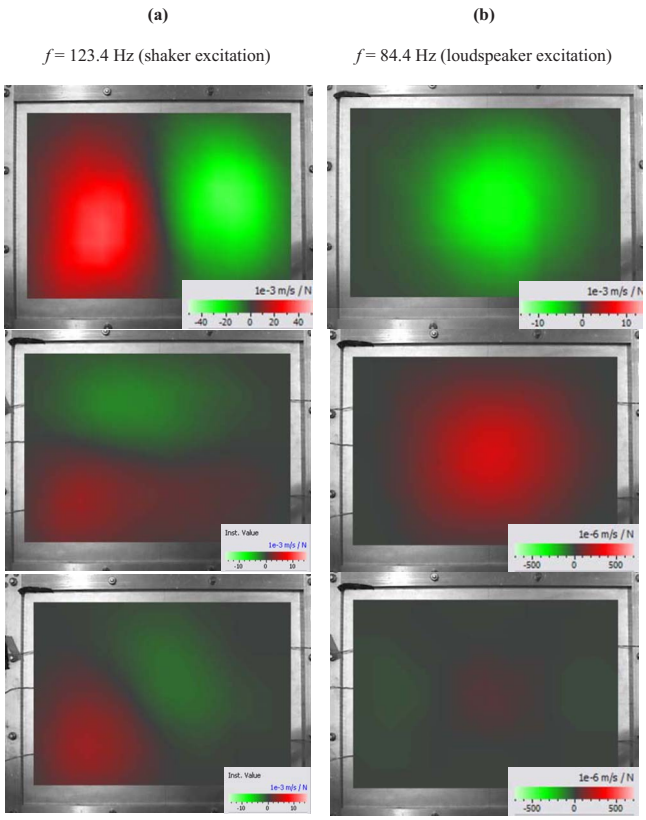


FIG. 12. (Color online) Response of the smart panel (a) excited by the shaker at the 123.4 Hz resonance frequency, which is controlled by the (2,1) mode of the panel, and (b) excited by the loudspeaker at the 84.4 Hz resonance frequency, which is controlled by the (1,1) mode of the panel. Top plots: response of the panels with no control actuators. Center plots: response of the panels with the five control units. Bottom plots: response of the panels when the five units implement decentralized velocity feedback control loops.

B. Total sound power radiated measured in anechoic room

The two plots in Fig. 13 show the total sound power radiated derived from the measured sound pressure in nine positions around the box in an anechoic room when the panel is excited by the shaker (a) and the loudspeaker in the cavity (b). As found for the response of the panel, the total sound power radiated by the panel without actuators (faint lines) is characterized by very lightly damped resonances in both cases where the panel is excited by the shaker or loudspeaker primary sources. In contrast to the spectrum found for the response of the panel, there are far less resonance peaks. This is due to the fact that, for frequencies below acoustic coincidence,[4] the sound radiation of even resonant modes is much lower than that of odd modes. This phenomenon is even more important when the panel is excited by the acoustic field in the cavity, which efficiently couples only with a selected set of modes of the panel. When the five actuators are mounted on the panel, the spectrum of the sound radiations (thick lines) is smoothed and lowered, respectively, by the passive damping and mass effects of the actuators. As a result of the sound radiation filtering effect and damping-mass actuator effects, the sound power radiated by the panel with the five actuators is characterized by very few heavily damped peaks at low frequencies. Also, the mean level is brought down by about 20–30 dB. It is interesting to note that the sound radiated when the panel is excited by the acoustic field in the cavity generated by the loudspeaker [Fig. 13(b)]: above 400 Hz, there are some sharp resonances, which are due to the acoustic modes in the cavity and thus cannot be damped by the actuators.

When the five decentralized feedback control loops are implemented on the panel excited by the shaker, a reduction of the total sound power radiation between 5 and 20 dB is measured in the frequency range up to 400 Hz. In the second case, where the panel is excited by the acoustic field in the cavity generated by the primary loudspeaker, when the five

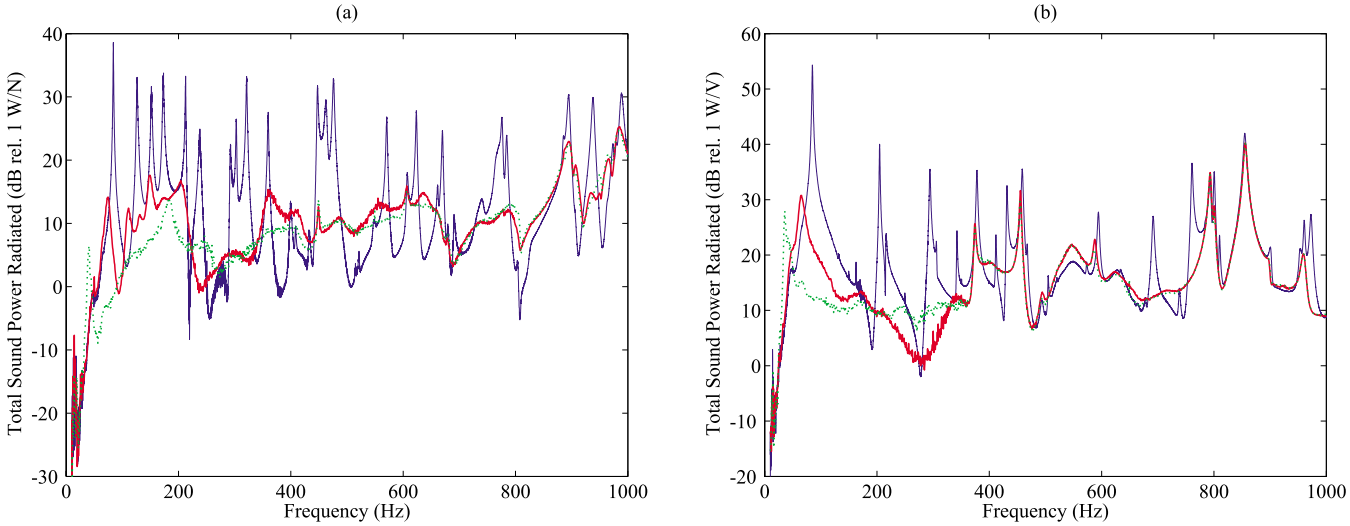


FIG. 13. (Color online) Measured narrowband spectra of the total sound power radiated derived from the measured sound pressure in nine positions around the box in an anechoic room when the panel is excited by the shaker (a) and the loudspeaker in the cavity (b). Faint line: response of the panel with no control units; thick line: response of the panel with the five control units; dotted line: response of the panel with the five control units implementing decentralized velocity feedback control.

decentralized feedback control loops are implemented, sound reductions are measured only up to 200 Hz, although a large reduction of about 16 dB is measured in correspondence to the resonance peak at 60 Hz of the (1,1) mode. This result is obtained at the expenses of a relatively large spillover effect at the fundamental resonance frequency of the five actuators.

The third octave band plots in Fig. 14 show that the passive effects of the five proof mass actuators produce maximum reductions of the sound power radiated up to about 10 and 11 dB, respectively, for the panel excited by the shaker and acoustic field generated by the loudspeaker in the cavity. When the five control units are activated, the maximum sound radiation is further brought down by another 6 and 10 dB in the two primary excitation cases. This result indicates that the reduction of vibration produced by the five decentralized control units reflects a correspondent reduction of the low frequency sound radiation.

VI. CONCLUDING REMARKS

This paper has presented a study on the implementation of decentralized velocity feedback control on a panel using proof mass electrodynamic actuators. In this paper, the passive and active effects of a set of five control units have been assessed both theoretically and experimentally.

A fully coupled model of the panel with the five electrodynamic actuators has been introduced and used to single out the principal dynamic effects produced by the actuators and

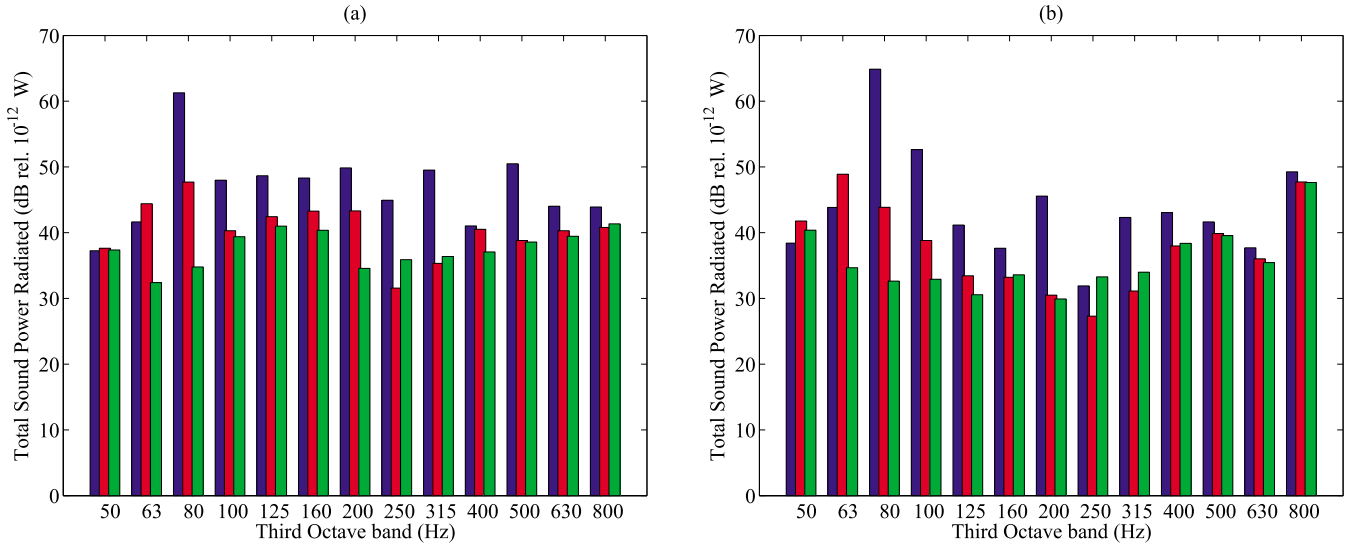


FIG. 14. (Color online) Measured third octave band spectra total sound power radiated derived from the measured sound pressure in nine positions around the box in an anechoic room when the panel is excited by the shaker (a) and the loudspeaker in the cavity (b). Left bar: sound radiated by the panel with no control units; center bar: sound radiated by the panel with the five control units; right bar: sound radiated by the panel with the five control units implementing decentralized velocity feedback control.

feedback control loops. It has been shown that the five actuators produce passive damping and mass effects, which effectively reduce the lightly damped response of the panel by 10–20 dB at resonance frequencies. Also, it has been shown that the stability of the five decentralized control units is affected by cross talking effects between the proof mass actuators. When the five control loops implement the maximum control gain that guarantees stability, the response of the panel is further reduced by 3–10 dB at low frequencies below 250 Hz.

The experimental study has been carried out in two stages. First, the stability and control performance at the error positions have been assessed. The stability analysis has confirmed the issue due to the cross talking between the actuators highlighted in the theoretical study. Nevertheless, the implementation of five decentralized feedback loops with gain margins of 7.6 dB that guarantees stability and little control spillover effect at the fundamental resonance frequency of the actuator has produced reductions of vibration at the control positions between 10 and 20 dB up to 550 Hz.

In the second stage of the experimental study, the global control performance produced by the smart panel with the five decentralized control units has been assessed with reference to the kinetic energy of the panel derived from the spatially averaged response of the panel measured with a scanning laser vibrometer and the total sound power radiated derived from the measured sound pressure in nine positions around the box in an anechoic room. Also in this case the two experiments have confirmed the theoretical predictions. The passive effects of the actuators efficiently bring down the lightly damped resonant response of the panel by 10–20 dB in the frequency range considered up to 1 kHz. The implementation of the five decentralized control units further reduces the response by another 3–8 dB at frequencies below 250 Hz. This behavior also reflects the measured total sound power radiated, although in this case the spectrum of the sound radiated is characterized by fewer peaks due to the efficiently radiating modes of the panel. As a result, the net reductions of the sound radiation are relatively smaller than those measured for the vibration response and they reach maximum values between 3 and 10 dB up to 250 Hz. Nevertheless, considering third octave plots, the system clearly produces an overall (due to both passive and active effects) reduction of the maximum response and reduction of maximum sound radiation of about 20 dB. When broadband noise is controlled, it is of great importance to bring down the maximum level, which usually lies at low audio frequencies, where, normally, passive systems are not that effective while, in contrast, the system presented in this study shows promising results.

ACKNOWLEDGMENTS

The work done by Cristóbal González Díaz and Christoph Paulitsch for this project was supported by the "Early Stage Training Site Marie Curie" programme for the "European Doctorate in Sound and Vibration Studies" which is funded by the European Commission. The prototypes of the actuators used in this study were built in collaboration with Rene'Boonen at Katholieke Universiteit Leuven.

[1]C. González Díaz, C. Paulitsch, and P. Gardonio, "Active damping control unit using a small scale proof mass electrodynamic actuator," J. Acoust. Soc. Am. **124**. 886 (2008).
[2]O. N. Baumann and S. J. Elliott, "The stability of decentralized multichannel velocity feedback controllers using inertial actuators," J. Acoust. Soc. Am. **121**, 188–196 (2007).
[3]P. Gardonio and M. J. Brennan, "Mobility and impedance methods in structural dynamics," in *Advanced Applications in Acoustics, Noise and Vibration*, edited by F. J. Fahy and J. Walker (E & FN Spon, London, 2004), Chap. 9, pp. 387–388.
[4]F. J. Fahy and P. Gardonio, *Sound and Structural Vibration: Radiation, Transmission and Response*, 2nd ed. (Academic, London, 2006).
[5]C. Paulitsch, P. Gardonio, and S. J. Elliott, "Active vibration damping using an inertial, electrodynamic actuator," ASME J. Vibr. Acoust. **129**, 39–47 (2007).
[6]S. J. Elliott, *Signal Processing for Active Control*, 1st ed. (Academic, London, 2001).
[7]E. Bianchi, P. Gardonio, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. Part III: Control system implementation," J. Sound Vib. **274**, 215–232 (2004).

Nondestructive characterization of musical pillars of Mahamandapam of Vitthala Temple at Hampi, India

Anish Kumar,[a] T. Jayakumar, C. Babu Rao, Govind K. Sharma, K. V. Rajkumar, and Baldev Raj
Indira Gandhi Centre for Atomic Research, Kalpakkam-603102, Tamil Nadu, India

P. Arundhati
ISVSA Project Scientist, Indian National Science Academy, New Delhi-110002, India

(Received 20 August 2007; revised 11 April 2008; accepted 27 May 2008)

This paper presents the first scientific investigation on the musical pillars of the Vitthala Temple at Hampi, India. The solid stone columns in these pillars produce audible sound, when struck with a finger. Systematic investigations on the acoustic characteristics of the musical pillars of mahamandapam (great stage) of the Vitthala Temple have been carried out. The 11 most popular pillars that produce sounds of specific musical instruments are considered for the investigations. The sound produced from these 11 most popular musical pillars was recorded systematically and different nondestructive testing techniques such as low frequency ultrasonic testing, impact echo testing, and *in situ* metallography were employed on the musical columns of these pillars. The peak frequencies in the amplitude spectrum of the sound produced from various columns in these pillars are correlated with the dimensional measurements and ultrasonic velocity determined using impact echo technique. The peak frequencies obtained experimentally have been found to have excellent correlation with the calculated flexural frequencies based on the dimensional measurements and ultrasonic velocities of the columns.
 [DOI: 10.1121/1.2945170]

PACS number(s): 43.40.At, 43.40.Cw, 43.75.Zz [JGM] Pages: 911–917

I. INTRODUCTION

The Vitthala Temple is the most magnificent of the religious edifices at Hampi, one of the world heritage sites in India. The Vitthala Temple stands on the southern bank of the Tungabhadra River and it portrays the high watermark of perfection of the Vijayanagara Kingdom style. The existence of the temple is traced at least to 1422–1461.[1] The temple stands in a large rectangular enclosure (164 m × 94.5 m). One of the unique features of the Vitthala Temple is the musical pillars on the mahamandapam (great stage) of the temple. The mahamandapam of the temple is reported to contain 56 pillars, each 3.6 m high, 40 of which are regularly disposed to form an aisle.[1] The remaining 16 form a rectangular court in the center. Each pillar is a massive composite sculptural unit measuring as much as 1.5 m across and has a group of monolithic sculptures. All the structures in the temple, including the musical pillars, are made of granite stone. The solid stone columns in these pillars produce audible sound, when struck with a finger. The musical columns in the pillars are of different size, shape, length, and width, which make them produce sounds of different musical instruments. It is said that the Bahamani invaders burnt the temple in the 15th century, damaging many pillars. These musical pillars have been a great attraction and surprise for tourists from all over the world. However, no systematic acoustic analysis of these pillars has been reported so far.

The present study is the first attempt to scientifically investigate the acoustic properties of the musical columns in the pillars of the mahamandapam of the Vitthala Temple and correlate them with the dimensional and nondestructive measurements.

II. EXPERIMENTAL DETAILS

The present study is concentrated on the 11 most popular pillars, which are believed to produce sounds of specific musical instruments. The details of these pillars and their locations are given in Table I and Fig. 1, respectively.

A. Dimensional measurements

Details of the locations of the pillars and the musical columns in the pillars were systematically recorded and all the pillars were photographed using a digital camera from various directions to record the details, to assist in analyzing the musical notes from different pillars. Figure 2(a) shows a photograph of the mahamandapam taken from southwest direction showing the pillars and the musical columns in the pillars. The close-up view of one of the musical columns toward the east in pillar 2 (Table I) is shown in Fig. 2(b).

Dimensional measurements were carried out on each musical column of the 11 pillars shown in Table I. The pillars were found to have larger dimension at the bottommost portion and it decreased almost continuously to the minimum at the uppermost portion. The diameter of the columns was measured at the lowest (maximum) and uppermost portions (minimum) by measuring the circumference with 0.5 mm ac-

[a]Electronic mail: anish@igcar.gov.in

TABLE I. Details of the 11 pillars examined in depth.

Pillar number as per Fig. 1	Musical instruments they are correlated to	Number of musical columns in the pillar	Total height range (mm)	Diameter range (mm)	Ultrasonic velocity range (m/s)	Frequency range of the sound produced (Hz)		Average time for decay of amplitude to 1/20 of the peak in autocorrelation function (ms)
						Lower	Higher	
2 (east side)	Saptaswara (seven notes)	8	960–1180	75.6–99.5	5300–5500	243–396	687–1000	226
2 (north side)	⋯	6+1 (Cut)	960–1140	89–104	5250–5430	296–372	738–967	130
3	Panchtala (five tones)	9	870–950	62.5–100.5	4600–5300	442–588	1046–1280	215
4	Jaltarang (water instrument)	7	880–920	90.0–130	5400–5500	495–780	–	266
5	Mridanga (percussion instrument)	5	910–985	98–106.5	5300–5500	444–457	1148	178 (114–284)
11	Ghanta (bell)	6	910–985	102.7–111.2	5400–5740	533–656	953	500
14	Ghatam (earthern pot, (percussion instrument)	12	1220–1250	102.7–122.5	5300–5600	305–454	774	300
16	Veena (sting instrument)	4	830–980	100.3–106.6	5200–5500	745–821	⋯	250
24	Tabla (percussion instrument)	10	1115–1120	97.3–107.0	5200–5600	296–407	⋯	188
3A	Damaru (small percussion instrument)	9	1170–1220	97.1–106.6	5200–5700	273–375	739–786	258
5A	Kerala Mridangam (percussion instrument)	5	1080–1130	115–122	5500–5600	338–412	783–901	295
20A	Shankha (shell)	5	940–1065	95.5–1127.5	5200–5450	399–471	983–1060	332
15A	Damaged				4200	680–760		91

curacy. The total length of the columns was also recorded with the same accuracy. The shape of the columns was also recorded.

B. Nondestructive testing

Different nondestructive testing techniques, i.e., low frequency ultrasonic testing, impact echo testing, and *in situ* metallography were employed on the musical columns. Low frequency ultrasonic testing was carried out to assess the internal structure of the pillar. Impact echo testing was carried out to measure the impact wave velocity in the columns and *in situ* metallography was carried out to identify the material of the column and presence of microcracks, if any.

1. Low frequency ultrasonic testing

A microprocessor-based ultrasonic flaw detector (Panametrics-NDT EPOCH-IV, Olympus NDT Inc., MA, USA) along with a pair of ultrasonic transducers of 500 kHz was used in through transmission mode to inspect the musical columns. One of the transducers was used to generate the sound waves. The transducer was coupled to the musical columns using grease as couplant. The time of travel of ultrasonic waves through the diameter of the columns was determined by observing the arrival of the ultrasonic waves in the receiver transducer, placed on diametrically opposite sides.

2. Impact echo testing

Impact echo technique involves introducing a transient stress pulse into a test object by mechanical impact and monitoring the surface displacements caused by the arrival of reflections of the pulse from internal defects and external boundaries.[2,3] The pulse consists of compression (P) and shear (S) waves that propagate into the object along spherical wave fronts, and a Rayleigh (R) wave that propagates along the surface. The bulk waves are reflected by internal defects and boundaries and the reflected waves propagate back to the surface. At the top surface, the waves are reflected again and they propagate into the test object. Thus a transient resonance condition is set up by multiple reflections of waves between the top surface and internal flaws or external boundaries. The frequency of P-wave arrivals at the transducer is determined by transforming the time domain signal into the frequency domain using the fast Fourier transform technique. The frequencies associated with the peaks in the resulting amplitude spectrum represent the dominant frequencies in the waveform. The specimen thickness or the depth of a flaw,

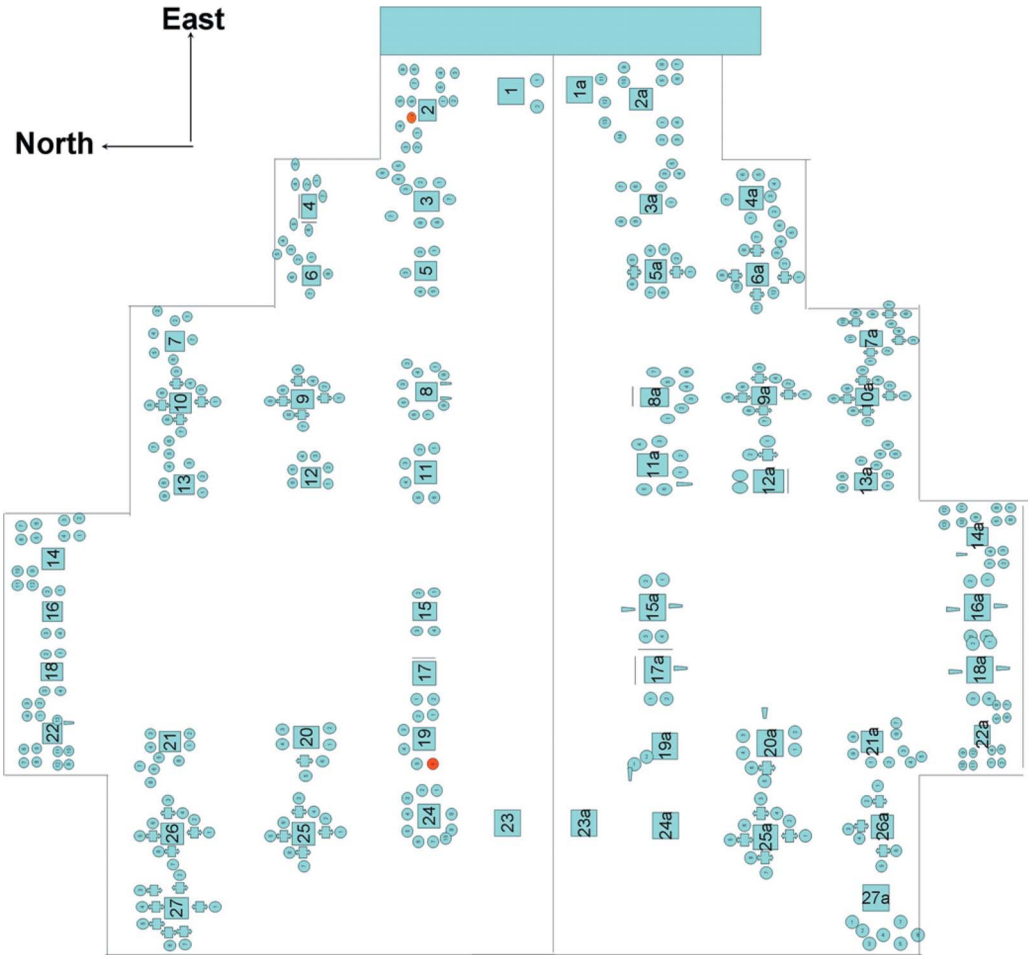


FIG. 1. (Color online) Plan of the mahamandapam indicating the pillars and the musical columns in them.

where waves are reflected between two materials/air interfaces, can be calculated using a simple formula which relates the depth, d, to the frequency of wave reflections, f, and to the P-wave speed, C_P:

$$d = \beta C_P / 2f, \tag{1}$$

where β is a structure factor depending upon the shape of the specimen. For cylindrical structures, β is equal to 0.92.[3]

An in-house developed impact echo system with a commercial receiver was used for measurement of impact wave velocities in the musical columns. Impact wave velocity measurements were carried out on all the musical columns of the 11 pillars mentioned earlier.

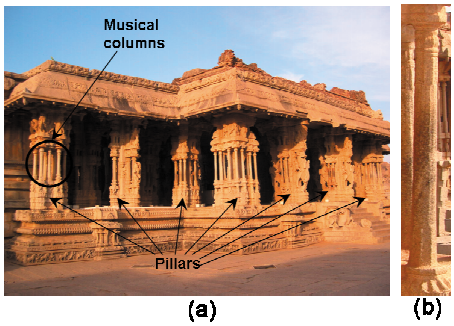


FIG. 2. (Color online) Photograph of (a) the mahamandapam taken from the southwest direction showing the pillars and the musical columns in the pillars and (b) one of the musical columns toward the east in pillar 2 (see Table I).

3. In situ metallography

In situ metallographic studies were carried out for identification of the type of material and to study the presence of microcracks/pores, if any, in the musical columns. It was carried out at two locations, i.e., the broken portions of two of the musical columns, one each in pillars 2 and 19. An *in situ* grinding machine was used for polishing the broken surface of the musical columns under continuous flow of water. A replica of the polished surface was taken using a replica tape to observe the features under an optical microscope.

C. Analysis of the sound waves from the musical columns

The sound produced by striking at the center of the musical columns with thumb was recorded in a laptop computer using a capacitor-type computer microphone for each of the musical columns in the pillars. Schematic of the experimental setup for recording the sound waves is shown in Fig. 3. Specific software was developed in LABVIEW with features for recording the sound waves and for online and offline analysis of the sound pattern in time and frequency domains.

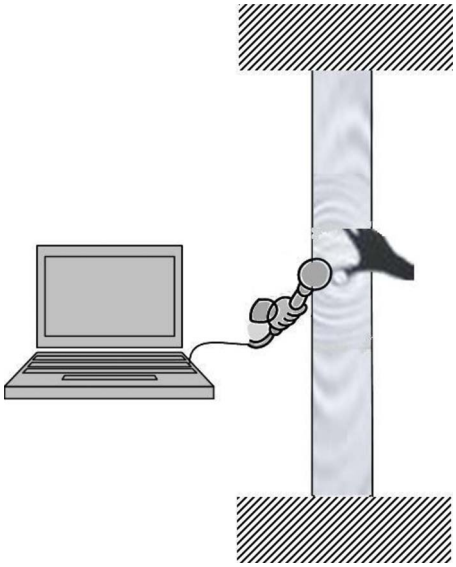


FIG. 3. (Color online) Schematic of recording of the sound produced by tapping on the musical columns with the thumb. Sound is recorded in a laptop computer using a capacitor microphone.

The sound waves were recorded at 22 kHz digitization with 16 bit resolution. A typical time domain signal, the corresponding amplitude spectrum, and the autocorrelation function are shown in Figs. 4(a)–4(c), respectively. Peak frequencies [as in Fig. 4(b)] and amplitude decay rate of the sound produced were analyzed for each of the musical columns. Half the width of the autocorrelation function at 1/20 of the peak amplitude has been used as a parameter related to the decay rate of the sound produced [Fig. 4(c)]. This parameter provides the time required for the decay of the amplitude to 1/20th of the peak amplitude and hence represents the attenuation/damping characteristics of the sound produced in the pillar. A higher value of this parameter indicates lower rate of attenuation/damping in the pillar and hence better coupling of the sound waves. The width of the autocorrelation in place of that of the sound signals is used to avoid error arising from the noise during the recording. Prior to recording the sound waves produced by the musical columns, the quality of the performance of the microphone was verified by recording and analyzing the sound waves produced by using a signal generator and a standard speaker.

III. RESULTS AND DISCUSSION

A. Detailed plan of the mahamandapam

Figure 1 shows the plan of the mahamandapam of the Vitthala Temple showing the placement of the pillars and musical columns in them. The mahamandapam is reported to consist of 56 main pillars.[1] However, only 54 main pillars could be observed during the detailed study, as shown in Fig. 1. These pillars are symmetrically placed on the north and south sides of the mahamandapam. In this paper, the pillars are referred to with numbers as shown in Fig. 1. Out of the 54 pillars, 4 pillars (1, 23, 23A, and 24A) are totally damaged and do not have any musical column now. The remaining pillars consist of different numbers of musical columns in the range of 4–15 numbers in each pillar.

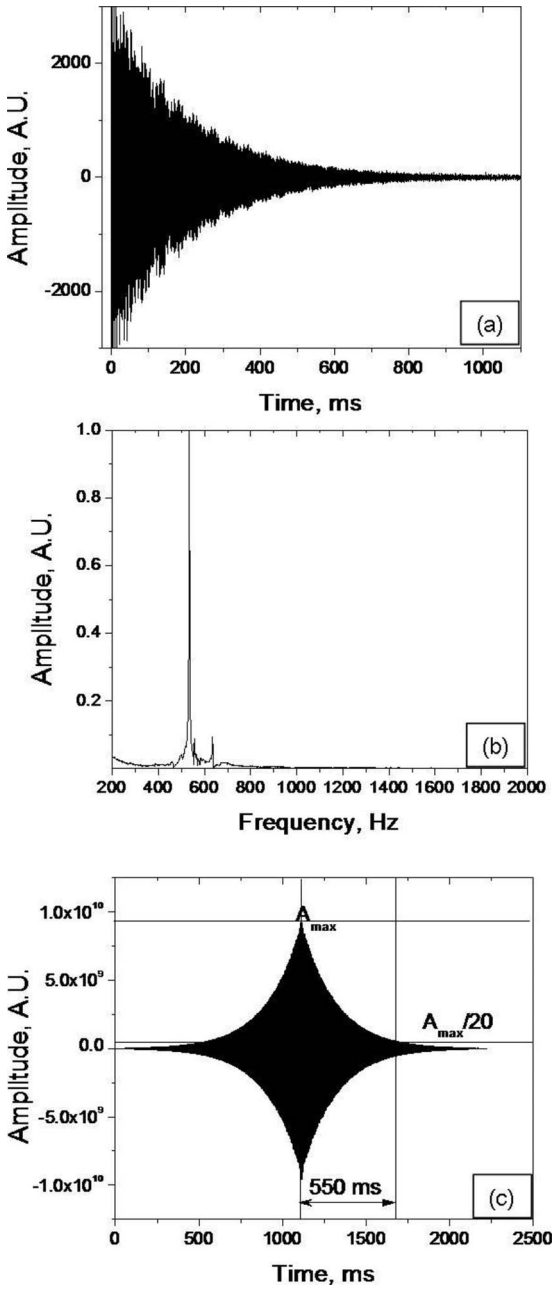


FIG. 4. (a) The time domain signal for the sound produced by tapping on one of the musical columns in pillar 16, which is popularly known to produce the sound of a bell. The amplitude spectrum and autocorrelation function of (a) are shown in (b) and (c), respectively. The time for decay of amplitude to 1/20 of the peak in autocorrelation function is also shown in (c).

B. *In situ* metallography

Figure 5 shows typical microstructures observed on the broken columns in pillars 2 and 19, respectively. The microstructures at both places are found to be similar to that of a typical granite stone.

C. Dimensional measurements

Table I also provides the range of length and the average diameter of each musical column of the 11 pillars. The columns were found to have larger dimension at the bottommost regions and it decreased almost continuously to the mini-

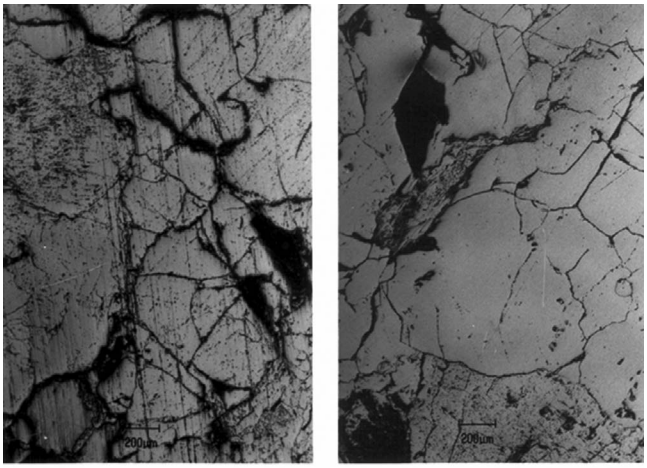


FIG. 5. Typical as polished microstructure in cut surface of the musical columns: (a) Pillar 2 and (b) pillar 19.

mum at the uppermost regions in most of the columns. The maximum variation in the diameter of the columns was found to be about 10% across about 1 m length of the column from the bottom to the top portion. In view of this, the columns are considered as cylindrical and the average diameter is reported and used for all the calculations.

D. Low frequency ultrasonics and impact echo testing

The low frequency ultrasonics and impact echo testing indicated that the musical columns are solid shafts in nature. Ultrasonic velocity in the musical columns is found to be in the range of 4800–5500 m/s in almost all the pillars, as can be seen in Table I. However, ultrasonic velocity in columns of a few of the pillars (pillars 15, 15A, 17) that exhibited damage (presumably by fire) is found to be much lower, i.e., in the range of ~3800 to 4200 m/s.

E. Analysis of the sound produced from the musical columns

The sound pattern for one of the columns of pillar 11 (popularly known to produce the sound of a bell), the corresponding frequency spectrum, and its autocorrelation function are shown in Fig. 4. A detailed analysis of the frequency content of the sound generated in different columns of various pillars has been carried out. The frequency range of the sound generated in different pillars is also given in Table I. In a few of the pillars, frequency corresponding to higher overtone has also been observed, as given in Table I. The average time for decay of amplitude to 1/20 of the peak in autocorrelation function (ms) corresponding to various columns in the pillars is also given in Table I. Proper coupling between the musical columns led to a large duration (~500 ms) of sound waves produced by the columns. The duration of the sound waves corresponding to a damaged pillar, which did not produce any musical sound, is also given in Table I for ready comparison. It can be seen very clearly that the duration of the sound produced is much higher for the pillars producing musical sounds as compared to the damaged one. Further, the duration of the sound is also found to be different in different musical pillars. Pillar 11, which is known to produce the sounds of bells, exhibited the highest duration of sound waves (>500 ms). The sound produced from the columns in this pillar exhibited almost exponential decay of the sound amplitude [Fig. 4(a)], inharmonic component [Fig. 4(b)], and nearby peaks [Fig. 4(b)] in the amplitude spectrum. These are the characteristic features of the bell-like sound.[4] Further, different musical columns in the same pillar produced sounds of different frequencies, usually in an increasing order in a sequence. Hence, when these columns are struck in a sequence, they produce the effect of playing of a musical instrument in increasing tone. A few of the columns, which have rectangular cross section, produce different sounds when struck on the two perpendicular sides. By considering the two moments of inertia in the two perpendicular directions for the rectangular columns, the ratio of the flexural resonance frequencies would be equal to the ratio of the thickness and the breadth (area moment of intertia, $I \propto bt^3$; and $f^2 \propto I$).[5] The frequency of the sound produced by striking the musical columns in the pillars is presumed to be tailored by adjusting the height and diameter of the columns. The correlation among the dimensions, ultrasonic velocity, and frequency of the sound waves generated in different columns is discussed in the following.

F. Correlation among dimensions of columns, velocity, and peak frequencies

For the sake of simplicity in the correlation, the conical (less than 10% variation in diameter over about 1 m length) columns are considered as cylinders. The diameter is taken as the average of the minimum and maximum diameters at top and bottom portions of the columns, respectively. For homogeneous columns with uniform cross section, bending and torsional vibrations can be described by a fourth- and a second-order partial differential equation, respectively. The equation of motion for the bending modes is given by[5,6]

$$EI\frac{\partial^4 y}{\partial x^4} + \rho A\frac{\partial^2 y}{\partial t^2} = 0, \tag{2}$$

where E is the Young's modulus, I $(=\pi D^4/64)$ is the area moment of inertia, ρ is the mass density, A $(=\pi D^2/4)$ is the cross-sectional area and D is the diameter. Here, x is the coordinate in the longitudinal direction of the column and $y(x)$ is the excursion from the rest position of the length element at x. A general solution for Eq. (2) can be written as

$$y(x,t) = (a_1 e^{kx} + a_2 e^{-kx} + a_3 e^{ikx} + a_4 e^{-ikx}) e^{i\omega t}. \tag{3}$$

Here, $\omega = 2\pi f$ is the angular frequency and $k = 2\pi/\lambda$ is the flexural wave number. Inserting Eq. (3) in Eq. (2) yields the following generalized dispersion relation:

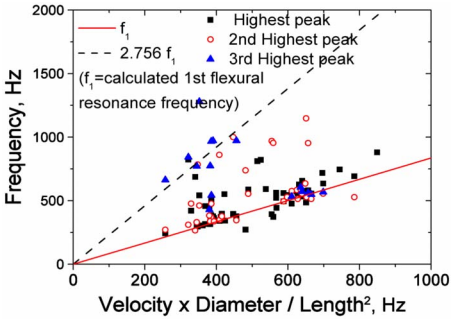


FIG. 6. (Color online) Variations in the peak frequencies with (velocity $\times$ diameter/length^2) for 87 columns in the 11 pillars. Solid and dashed lines show the variations in the first (f_1) and second (f_2) flexural resonance frequencies, respectively, with (velocity $\times$ diameter/length^2) as per Eqs. (8) and (9).

$$EIk^4 - \rho A\omega^2 = 0 \quad \text{or} \quad \omega = k^2\sqrt{\frac{EI}{\rho A}} \quad \text{or} \quad f = \frac{k^2}{2\pi}\sqrt{\frac{EI}{\rho A}} \tag{4}$$

$$f = \frac{k^2}{2\pi}\sqrt{\frac{EI}{\rho A}} \quad \text{or} \quad f = \frac{(kl)^2}{2\pi l^2}\sqrt{\frac{EI}{\rho A}},$$

where, l is the length of the column. As the columns analyzed in the present study are clamped at both the ends, the boundary conditions are

$$y = 0, \quad \frac{\partial y}{\partial x} = 0 \quad \text{at} \quad x = 0, \quad x = L. \tag{5}$$

These boundary conditions lead to the following frequency equation:

$$(\cos kl)(\cosh kl) = 1. \tag{6}$$

For Eq. (6), the first (f_1) and the second (f_2) flexural resonance frequencies can be derived as

$$f_1 = 0.89\frac{D}{L^2}\sqrt{\frac{E}{\rho}}, \tag{7}$$

$$f_2 = 2.7564 f_1. \tag{8}$$

Equation (7) can be reduced to the experimentally measured quantities, i.e., frequency, d, L, and V_L by substituting

$$\sqrt{\frac{E}{\rho}} = V_L\sqrt{\frac{(1+\nu)(1-2\nu)}{(1-\nu)}},$$

where ν is the Poisson's ratio. The values of ν for granites are reported to be in a range of 0.1–0.33.[7] As the ν is not measured experimentally for the columns, ν=0.215 (mean value for granites) is considered for the correlation. Substituting these values in Eq. (7) leads to

$$f_1 = 0.84 DV_L/L^2, \tag{9}$$

where D and L are in meters and V_L is in meters per second.

Based on Eq. (9), the variation in the peak frequency with DV_L/L^2 is plotted in Fig. 6 for 87 columns in 11 pillars. In a few of the pillars, more than one peak (up to three peaks) was observed in the frequency spectra. The frequencies corresponding to the highest amplitude, the second highest amplitude, and the third-highest amplitude are shown as square, circular, and triangular data points in Fig. 6. The average diameter of the top and bottom regions of the column is used in the correlation. For columns with square cross section (about 10 out of 87 columns analyzed), average thickness is used in place of the diameter. For the same range of DV_L/L^2, the calculated first and second flexural resonance frequencies are also plotted in Fig. 6 for ready comparison. The calculated resonance frequencies are found to be in good agreement with those observed experimentally. This indicates that the sound produced in the columns is essentially arising due to the flexural mode of vibrations. The deviations from the calculated frequencies for a few of the columns could be attributed to the structural factors, such as localized damage to the columns, varied shape or diameter over its length, interference from the nearby columns, or varied integrated geometries of the columns with the base and top structures. Further, as the columns can be excited to produce audible sound by striking with the thumb, it indicates that the driving point impedance of the columns is extremely low. This aspect is being investigated separately.

IV. CONCLUSION

This study reports the first scientific investigation on the acoustic properties of the musical columns in the pillars of the mahamandapam of the Vitthala Temple at Hampi, a world heritage site in India. The study was concentrated on the 11 most popular pillars, which are widely known to produce sounds of different Indian musical instruments. Various nondestructive techniques, such as low frequency ultrasonic testing, impact echo testing, and *in situ* metallography, were employed on the musical columns of the pillars. The *in situ* metallography revealed the microstructure of the pillars to be similar to a typical granite microstructure. The low frequency ultrasonic and impact echo testings revealed that all the musical columns are solid shafts. Further, ultrasonic velocity was found to be almost uniform in all the good pillars. The velocity in the damaged pillars is found to be considerably lower as compared to that in the good pillars. The frequency of the sound produced from the musical columns could be correlated well with the calculated flexural resonance frequencies based on the dimensions (height and diameter) and the velocity of the sound waves in the columns.

ACKNOWLEDGMENTS

The author acknowledge P. Sukumar of the Indira Gandhi Centre for Atomic Research for his assistance with *in situ* metallography, dimensional measurements, and photography. The authors are also thankful to the Indian National Science Academy (INSA), New Delhi, and Archeological Survey of India (ASI), Hampi, for their support in carrying out this study. P. A. greatefully acknowledges the financial support received from INSA.

[1]D. Devakunjari, *Hampi*, 4th ed. (The Director General, Archeological Survey of India, New Delhi, 1998), pp. 63–66.
[2]A. Kumar, B. Raj, P. Kalyanasundaram, T. Jayakumar, and M. Thavasimuthu, "Structural integrity assessment of the containment structure of a pressurized heavy water nuclear reactor using impact-echo technique," NDT & E Int. **35**, 213–220 (2002).

[3]Y. Lin and M. Sansalone, "Transient response of thick circular and square bars subjected to transverse elastic impact," J. Acoust. Soc. Am. **91**, 886–893 (1992).
[4]M. Karjalainen, V. Valimaki, and P. A. A. Esquef, "Efficient modeling and synthesis of bell-like sounds," in *Proceedings of the Fifth International Conference on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, 26–28, September 2002.
[5]W. F. Stokey, "Vibration of systems having distributed mass and elasticity," in *Shock and Vibration Handbook*, 2nd ed., edited by C. M. Harris and C. E. Crede (McGraw-Hill, New York, 1976), pp. 7–14.
[6]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, Berlin, 1991), 56 pp.
[7]H. Gercek, "Poisson's ratio values for rocks," Int. J. Rock Mech. Min. Sci. **44**, 1–13 (2007).

Defect detection and localization in orthotropic wood slabs by inversion of dynamic surface displacements

Anthony J. Romano
Naval Research Laboratory, Physical Acoustics Branch, Code 7130, 4555 Overlook Avenue, S.W., Washington, DC 20375

Joseph A. Bucaro[a),b)]
SET, Inc., Springfield, Virginia 22150

Saikat Dey[b)]
SFA, Inc., Crofton, Maryland 21114

(Received 14 February 2008; revised 19 May 2008; accepted 24 May 2008)

The nondestructive evaluation inversion and generalized force-mapping techniques developed and demonstrated for isotropic thin plates by Bucaro *et al.* [(2004). "Detection and localization of inclusions in plates using inversion of point actuated surface displacements," J. Acoust. Soc. Am. **115**, 201–206] are extended to the case of orthotropic plates. The extended techniques are applied to a finite-element generated numerical database for point excited wooden slabs with and without an internal defect at 5 and 10 kHz. Operation of the original isotropic algorithms on the wood surface displacements is shown to fail in recovering the uniform elastic parameters or in detecting and locating the defect. The new algorithms based on the wave equation for a thin, orthotropic plate successfully convert the surface displacements on the uniform wooden slab to elastic parameter maps which serve to detect and localize the defect in the flawed plate. The results, particularly at the higher frequency, indicate that the onset of failure in the thin plate approximation impacts both the inversion and the generalized force-mapping accuracy. However, in this case use of the inversion algorithm to obtain modified wave equation coefficients followed by operation of the force-mapping algorithm with these new parameters inserted is shown to successfully mitigate this effect.
[DOI: 10.1121/1.2945706]

PACS number(s): 43.40.Le, 43.40.Dx, 43.60.Pt [YHB] Pages: 918–925

I. INTRODUCTION

Nondestructive evaluation (NDE) in wood and wooden structures using the propagation of acoustic waves and the related vibration is of interest in a number of areas (Bucur, 2003). For example, acoustic-based NDE techniques that can determine wood material properties in a cost effective manner are of considerable importance to the timber industry for both health monitoring of living trees and defect detection in quality sorting of structural products. Many historic buildings have a significant amount of wooden members, and preservationists require such NDE techniques to determine the extent of deterioration so that degraded members may be replaced or repaired to avoid structural failure. Furthermore acoustic evaluation techniques also play an important role in wood selection for the making of quality musical instruments.

Compared to their use in nonwooden members or structures—e.g., those made of steel, plaster, concrete, etc.—the application of acoustic-based NDE techniques to wooden materials is somewhat more complex owing to the orthotropic nature of wood. Notwithstanding the complications arising from a wooden article's unique and independent mechanical properties along three mutually perpendicular axes, considerable progress has been made in the development of acoustic measurement and related NDE techniques for wood. For example, motivated by such applications Okyere and Cousin (1980) and Bucur and Feeney (1992) reported successful ultrasonic measurements of wood elastic wave attenuation, and Bucur and Archer (1984) wood elastic constants. Berndt *et al.* (1999) implemented high resolution imaging of transmitted and reflected ultrasonic energy in wood slabs and were able to find signal components related to the anatomical features of the wood. Tomikawa *et al.* (1986) developed and demonstrated a tomographic technique for the inspection of wooden poles.

Here we explore the extension of our previously reported NDE techniques (Bucaro *et al.*, 2004 and Romano *et al.*, 2007) based on inversion of two dimensional (2D) surface vibration maps measured on isotropic materials to those obtained on orthotropic wood structures. In these techniques, the structure is weakly excited into vibration by a broadband, localized shaker or by a broadband incident sound field. The resulting surface dynamic velocity, which is mapped spatially with a laser Doppler scanning vibrometer, is inverted *locally* using specially designed algorithms based on the flexural wave equation.

These techniques were first demonstrated (Bucaro *et al.*, 2004) using numerical surface displacement data generated with a finite-element code for solid, isotropic plates of both

[a)]Author to whom correspondence should be addressed. Electronic mail: bucaro@pa.nrl.navy.mil
[b)]On-site at The Naval Research Laboratory, Washington, D.C.

steel and mortar which were homogeneous except for a small, softer inclusion. The inversion algorithms successfully detected and localized the defect. Furthermore, one of the inversion techniques was shown to be capable of quantitatively recovering the local bending stiffness of the plate. More recently, these inversion techniques were applied to an experimental study on thin metal and composite ribbed plate structures (Romano *et al.*, 2007). Here, the inverted displacement (velocity) maps successfully detected and located rib detachment purposely created along a short segment of one of the ribs.

In this paper, we describe progress we have made in extending and applying these techniques to wooden, orthotropic slabs. In Sec. II, we present a short mathematical review of the flexural wave inversion algorithms for isotropic materials and then describe our extension of these algorithms to the orthotropic thin plate case. In Sec. III, we present application of the new algorithms to a numerical database which we have generated for a locally excited wooden slab structure with and without an internal defect.

II. INVERSION ALGORITHMS

A. Isotropic plates

As summarized below, the case for flexural wave inversion of surface displacements on isotropic thin plates was first treated by Bucaro *et al.* (2004) who developed inversion algorithms exploiting virtual functions to weaken the deleterious effect of spatially dependent noise on the resulting inversion. In the first of these techniques named "weak flexural wave inversion (WFWI)," with x,y coordinates in the plane of the plate and z in the thickness direction, the equation of motion appropriate to a thin, isotropic plate

$$D\left(\frac{\partial^4 u_z}{\partial x^4}+2\frac{\partial^4 u_z}{\partial x^2\partial y^2}+\frac{\partial^4 u_z}{\partial y^4}\right)-\rho\omega^2 h u_z=f_z(x,y,\omega) \tag{1}$$

is expressed in variational form as

$$\int_x\int_y\left(v_z D\left(\frac{\partial^4 u_z}{\partial x^4}+2\frac{\partial^4 u_z}{\partial x^2\partial y^2}+\frac{\partial^4 u_z}{\partial y^4}\right)-v_z\rho\omega^2 h u_z\right)dxdy = \int_x\int_y v_z f_z(x,y,\omega)dxdy, \tag{2}$$

where h is the plate thickness, ρ is the plate density, D is the flexural rigidity defined as $Eh^3/12(1-\nu^2)$ (where E is Young's modulus and ν is Poisson's ratio), f_z is an applied normal surface force, ω is the angular frequency, and ν_z is our 2D virtual function defined as $\nu_z=(1-\bar{x}^2)^2(1-\bar{y}^2)^2$. For a surface area element of sides L_m^x, L_m^y, centered on the global coordinates (x_m,y_m), the local coordinates $(\bar{x},\bar{y})$ are defined as $\bar{x}=2(x-x_m)/L_m^x$, and $\bar{y}=2(y-y_m)/L_m^y$. Integration of Eq. (2) by parts four times yields

$$D\left[\int_y\left[\left.\frac{\partial^2 v_z}{\partial x^2}\frac{\partial u_z}{\partial x}\right|_x-\left.\frac{\partial^3 v_z}{\partial x^3}u_z\right|_x+\int_x u_z\frac{\partial^4 v_z}{\partial x^4}dx\right]dy + 2\int_x\int_y u_z\frac{\partial^4 v_z}{\partial x^2\partial y^2}dxdy+\int_x\left[\left.\frac{\partial^2 v_z}{\partial y^2}\frac{\partial u_z}{\partial y}\right|_y-\left.\frac{\partial^3 v_z}{\partial y^3}u_z\right|_y + \int_y u_z\frac{\partial^4 v_z}{\partial y^4}dy\right]dx\right]-\int_x\int_y\rho\omega^2 h v_z u_z dxdy = \int_x\int_y v_z f_z dxdy. \tag{3}$$

The important result here is that the variational form has forced all derivatives higher than first order onto the virtual functions, thereby reducing the effects of spatially varying noise in the process of calculating higher order derivatives.

For simplicity, the integrands multiplying the flexural rigidity, D, are labeled as I and Eq. (3) is rewritten as

$$DI-\int_x\int_y\rho\omega^2 h v_z u_z dxdy=\int_x\int_y v_z f_z dxdy. \tag{4}$$

Avoiding the location of any applied force (so that $f_z=0$) and dividing by the quantity ρh, Eq. (4) can be rearranged to yield

$$\left(\frac{D}{\rho h}\right)_m=\frac{\int_x\int_y\omega^2 v_z u_z dxdy}{I}. \tag{5}$$

As can be seen, Eq. (5) allows one to obtain the ratio of the elastic parameter $D/(\rho h)$ by use of a fairly simple operator acting on the measured displacement field u_Z. Accordingly, if the defect has an appreciable impact on the *local* bending stiffness, processing the measured surface displacements according to Eq. (5) can serve to detect and localize the defect.

The second technique called "generalized force-mapping (GFM)" results from a straightforward application of the variational form of the inhomogeneous equation of motion as portrayed in Eq. (4). The approach, which strictly speaking is not an "inversion" algorithm *per se*, provides a mapping of forces through a forward calculation of Eq. (4). In this method, one assumes knowledge not only of the measured displacement field, $u_z(x,y,\omega)$, but also of the material parameters and plate thickness, all of which are assumed to be constant with respect to space within each local voxel of interrogation. Dividing both sides of Eq. (4) by the quantity ρh, one can see that for a homogeneous plate away from the applied force, the two terms on the left-hand side sum identically to zero. For a plate with an inclusion (or applied surface traction), a nonzero result for the left-hand side (away from the known position of the applied force) is interpreted as a generalized force G_m defined as $G_m=\int_x\int_y(\nu_z f_z(x,y,\omega)/\rho h)dx$ dy which is present in the affected region as a consequence of the defect. These cases may be expressed as

$$\left(\frac{DI}{\rho h}\right)_m - \int_x \int_y \omega^2 v_z u_z dxdy = \begin{cases} 0 & \text{if the plate is homogeneous} \\ G_m & \text{if a force or defect is present.} \end{cases} \quad (6)$$

Therefore, as the left-hand side of Eq. (6) is calculated at each 2D voxel centered at the location x_m, y_m over the surface of the plate, nonzero values illuminate and map the spatial distributions of any defects (or applied forces). We note that being based on the variational form of the equation of motion, this algorithm like WFWI is also insensitive to spatially dependent noise.

In isotropic thin plate structures, both WFWI and GFM have been shown to be effective in detecting and localizing defects, and WFWI in mapping spatially dependent elastic parameters as well (Bucaro *et al.*, 2004; Romano *et al.*, 2007). In the next section, we apply the WFWI inversion algorithm to a database generated numerically on wooden slabs. As will be seen, applying the WFWI algorithm derived for the isotropic case to the displacements found on orthotropic slabs fails to detect the defect or to determine the elastic stiffness correctly. Recognizing that this failure is a consequence of the anisotropic nature of wood, we now reformulate the inversion and generalized force machinery based on the theory of wave propagation in *orthotropic* materials.

B. Orthotropic plates

In the case of an orthotropic thin plate, the differential equation for the transverse bending is given by Leissa (1993)

$$D_x \frac{\partial^4 u_z}{\partial x^4} + 2D_{xy} \frac{\partial^4 u_z}{\partial x^2 \partial y^2} + D_y \frac{\partial^4 u_z}{\partial y^4} - \rho \omega^2 h u_z = f_z(x,y,\omega), \quad (7)$$

where

$$D_x = \frac{E_x h^3}{12(1-\nu_{xy}\nu_{yx})},$$

$$D_y = \frac{E_y h^3}{12(1-\nu_{xy}\nu_{yx})},$$

$$D_{xy} = D_x \nu_{yx} + 2D_k,$$

$$D_k = \frac{G_{yx} h^3}{12}. \quad (8)$$

In these equations D_x, D_y, and D_{xy} are the flexural rigidities, E_x, E_y, and G_{yx} are the orthotropic elastic moduli, and ν_{xy}, ν_{yx} the relevant Poisson's ratios. The equivalent variational form can be expressed as

$$\int_x \int_y \left(v_z \left(D_x \frac{\partial^4 u_z}{\partial x^4} + 2D_{xy} \frac{\partial^4 u_z}{\partial x^2 \partial y^2} + D_y \frac{\partial^4 u_z}{\partial y^4} \right) - v_z \rho \omega^2 h u_z \right) dxdy = \int \int v_z f_z(x,y,\omega) dxdy, \quad (9)$$

where the virtual functions, v_z, are defined above. Integration of Eq. (9) by parts four times yields

$$D_x \int_y \left[\frac{\partial^2 v_z}{\partial x^2} \frac{\partial u_z}{\partial x}\bigg|_x - \frac{\partial^3 v_z}{\partial x^3} u_z \bigg|_x + \int_x u_z \frac{\partial^4 v_z}{\partial x^4} dx \right] dy + 2D_{xy} \left[\int_x \int_y u_z \frac{\partial^4 v_z}{\partial x^2 \partial y^2} dxdy \right] + D_y \int_x \left[\frac{\partial^2 v_z}{\partial y^2} \frac{\partial u_z}{\partial y}\bigg|_y - \frac{\partial^3 v_z}{\partial y^3} u_z \bigg|_y + \int_y u_z \frac{\partial^4 v_z}{\partial y^4} dy \right] dx - \int_x \int_y \rho \omega^2 h v_z u_z dxdy = \int_x \int_y v_z f_z(x,y,\omega) dxdy. \quad (10)$$

As before, one can see that the variational form has forced all but first order derivatives onto the virtual functions, thereby reducing the effects of spatially varying noise in the process of calculating higher order derivatives.

1. Orthotropic inversion

As in the isotropic case, Eq. (10) can be utilized to solve for the orthotropic flexural rigidities away from any applied force [where $f_z(x,y,\omega)=0$]. Since there are now three unknowns, a minimum of three excitation locations is required to provide a linearly independent set of equations. Labeling the integrals in Eq. (10) as A_{ij}, $i,j=1,2,3$, and the displacements due to each excitation as u_z^i, the set of equations may be expressed in matrix form as

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} \frac{D_x}{\rho h} \\ \frac{D_{xy}}{\rho h} \\ \frac{D_y}{\rho h} \end{bmatrix}_m = \begin{bmatrix} \iint_{xy} \omega^2 v_z u_z^1 dxdy \\ \iint_{xy} \omega^2 v_z u_z^2 dxdy \\ \iint_{xy} \omega^2 v_z u_z^3 dxdy \end{bmatrix}. \quad (11)$$

Letting the 3×3 matrix in Eq. (11) be represented as A, the unknown flexural rigidity column matrix as D, and the right-hand side as R, then the solution can be expressed as $D = A^{-1}R$, $(\text{Det}\, A \neq 0)$. This set of equations may additionally be over specified and solved using algorithms such as the conjugate gradient least squares (Hansen, 1998) method to site one example. For orthotropic plates, Eq. (11) replaces Eq. (5), its counterpart for the isotropic case.

2. Orthotropic generalized force mapping

In the case of orthotropic plates, the GFM operator assumes the following form:

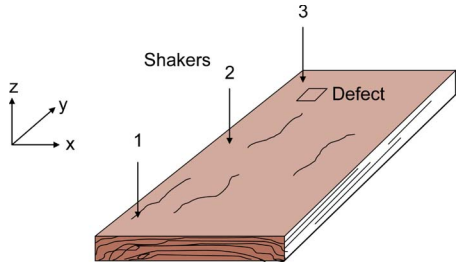


FIG. 1. (Color online) Wooden plate and Cartesian axis system. The x and z axes are along the wood tangential and radial growth ring directions, and the z axis along the longitudinal or fiber direction. The defect is in the form of a rectangular inclusion centered midway through the slab depth. Also shown are three independent shakers used to excite the slab.

$$\int_x \int_y \left(v_z \left(\frac{D_x}{\rho h} \frac{\partial^4 u_z}{\partial x^4} + 2 \frac{D_{xy}}{\rho h} \frac{\partial^4 u_z}{\partial x^2 \partial y^2} + \frac{D_y}{\rho h} \frac{\partial^4 u_z}{\partial y^4} \right) - v_z \omega^2 u_z \right) dxdy = \begin{cases} 0 & \text{if the plate is homogeneous} \\ G_m & \text{if a force or defect is present,} \end{cases} \tag{12}$$

where, as discussed above, D_x, D_{xy}, and D_y are the orthotropic flexural rigidities, ρ and h are the density and plate thickness, respectively, and $G_m = \int_x \int_y (v_z f_z(x,y,\omega)/\rho h) dxdy$. In comparison to Eq. (6) for the isotropic plate, as the left-hand side of Eq. (12) is calculated at each 2D voxel centered at the location x_m, y_m over the surface of the plate, nonzero values illuminate and map the spatial distributions of any defects (or applied forces).

This particular representation will be applied in two different ways. The first is one in which we utilize the known values for the flexural rigidities which are based on the elastic moduli used in the forward finite-element calculations. The second is one in which we utilize the values for the flexural rigidities obtained from the orthotropic inversions. In this latter method, the values in Eq. (12) are therefore "calibrated" to the solution as provided by the application of the inversions. As will be seen, such an approach can have a dramatic effect in the performance of the GFM method.

III. INVERSION AND FORCE MAPPING RESULTS

We focus our study and demonstration of these techniques on a solid, homogeneous, orthotropic wooden slab of dimensions length=60 cm, width=30 cm, and thickness =2.54 cm with and without an internal defect. As depicted in Fig. 1, we take the x, y, and z axes to be aligned with tangential, longitudinal (fiber), and radial wood grain and growth ring directions, respectively. The nine independent stiffness constants and density (450 kg/m^3) are taken to be that of Douglas coastal fir (see Table I) as computed (see Appendix) from the elastic and Poisson's ratios reported by Green *et al.*(1999). The defect is in the form of a rectangular inclusion (see Fig. 1) of length=2 cm, width=1.5 cm, and thickness=0.5 cm (centered at $x=-8.7$ cm, $y=19$ cm, and $z=1.27$ cm) with the density unchanged but with each of the nine stiffness constants reduced to 0.05 its value in the normal wood.

TABLE I. Nine independent stiffness matrix elements for Douglas fir (10^8 Pa).

C_{11}	8.450
C_{22}	150.6
C_{33}	11.44
C_{44}	11.50
C_{55}	1.030
C_{66}	9.430
C_{12}	4.784
C_{13}	3.390
C_{23}	4.851

The damping factor is taken to be 5% and is accounted for by adding the appropriate imaginary component to each elastic moduli (Hosten, 1991). Some support we can offer for choosing 5% is that the 2 and 15 dB/cm longitudinal and shear attenuations in Douglas fir at 1 MHz reported by Bucur (1999) correspond to damping factors of 4.4% and 2.2%, respectively. Of course, our results are obtained at much lower frequencies (5–10 kHz) and for flexural waves. Also, damping factors for various wave types in white oak reported by Kerlin (1966) fall anywhere between 1% and 3% in our frequency band. We favored erring on too high rather than too low a damping factor recognizing that the performance of our inversions is expected to decline with larger attenuation. In any case, the results we will present should not be too dependent on the actual damping number used in the range given above.

A. Finite-element database

The displacement response of the orthotropic wood slab was obtained using a parallel hp-version finite-element technique (Dey and Datta, 2006) with a mesh consisting of 9000 volume elements. After a p-convergence check, a cubic ($p=3$) discretization was used consisting of 208,448 complex-valued displacement degrees of freedom. The surface displacements were obtained resulting from a normal point force applied one at a time to three different positions on the slab, one near the lower left corner, one directly above it halfway up the slab, and the third near the upper left corner (see Fig. 1). In the finite-element model, the boundary conditions were taken to be fixed, i.e., all three displacement components at the edges are zero. Particular boundary conditions should have little effect on our inversions since they are based on infinite plate, free-wave propagation. Furthermore except for "edge distortion" at positions very near the boundaries of the slab (i.e., within distances $\ll$ structural wavelength), the displacement responses are well represented by linear combinations of the infinite plate, free-wave solutions (Skudrzyk, 1968).

The responses were computed for two frequencies (5 and 10 kHz) defining a band which we believe would be practical from both a force application and a scanned surface displacement measurement point of view. For each case we compute the normal surface displacement $u_z(x,y,\omega)$ on a rectangular grid with a spacing of 0.25 cm. The displacements at each frequency are shown in Fig. 2 for both the

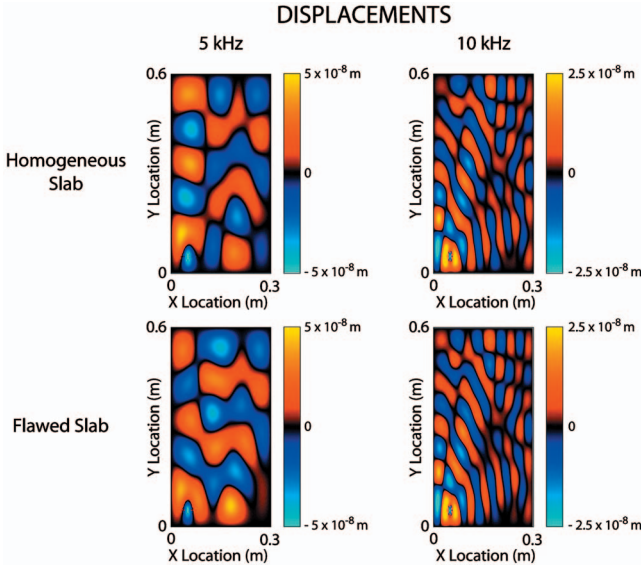


FIG. 2. Calculated color-coded surface displacement at 5 and 10 kHz when the shaker in the lower left position is excited at each frequency. Positive and negative relative phases are indicated by red to yellow and dark to light blue, respectively.

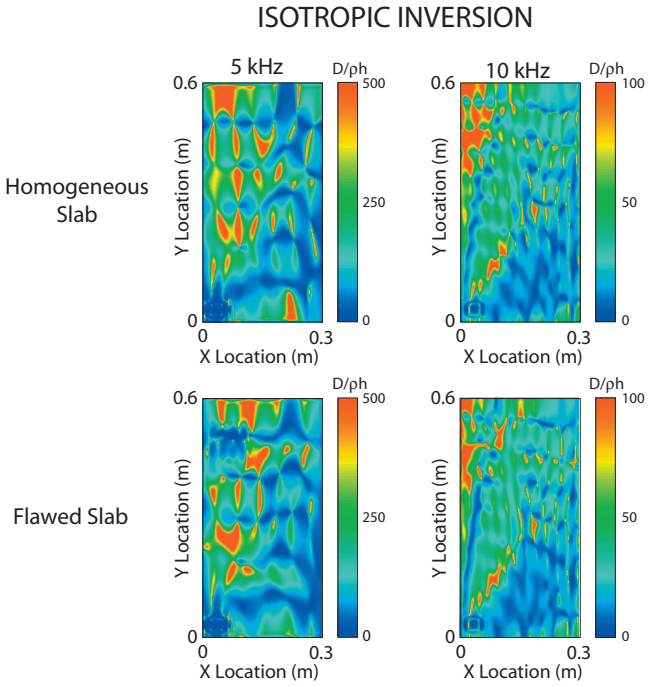


FIG. 3. Result of applying weak flexural wave inversion operator (WFWI) developed for the isotropic case, Eq. (5), to the displacement data (lower left driver excited) at each frequency for both the homogeneous wooden slab and the flawed wooden slab. The magnitude of the inversion in units of $D/\rho h$ is color coded from blue to red.

homogeneous slab and the flawed slab. It should be pointed out that the displacement maps themselves show little indication of the presence of the internal defect.

B. Results

To provide the inversion results, the integrations in Eqs. (5) and (11) were calculated over 16×16 data points, centered at each of 106×226 pixels (x_m,y_m), i.e., $x_m\pm8\Delta x$, $y_m\pm8\Delta y$, yielding surface elements with sides $L_m^x=L_m^y=4$ cm. First, we can apply the inversion operator, Eq. (5), developed from the isotropic wave equation to the numerical database of displacement on the wooden slabs. In Fig. 3 we show the result of this operation on both the homogeneous slab and on the flawed slab for one of the driver positions (lower left). As can be seen, the inversion algorithm fails in two respects: (1) it does not recover the known spatial uniformity of the effective stiffness for either slab (note the presence of large stiffness parameter values at a number of locations), and (2) it does not do well in detecting and localizing the internal defect. These failures are not surprising and are a consequence of applying an inversion operator built on the isotropic wave equation to an orthotropic structure. The same problems persisted when inverting the data associated with the other two shaker locations as well, although the results are not shown here.

Next, we apply the orthotropic inversion algorithm as described in Eq. (11). Figure 4 shows the inversions yielding the three stiffness parameters $D_x/\rho h$, $D_{xy}/\rho h$, and $D_y/\rho h$, respectively, mapped spatially over the homogeneous wooden slab. As can be seen, away from the three shaker positions, each inversion correctly produces a spatially uniform stiffness parameter. In Fig. 5 we show the inversion results for the flawed slab. As can be seen, all three inversions successfully detect and localize the rectangular internal inclusion through the three D coefficients. From the perspective of contrast, the inversion yielding $D_y/\rho h$ seems somewhat superior in performance to that for $D_x/\rho h$ or $D_{xy}/\rho h$. This is perhaps related to the fact that the displacement response may be determined predominantly through D_y as appears to be the case based on the modal spatial patterns seen in Fig 2. For example, from these displays we find an estimate of the dominant structural wave number, k, to be consistent with $\omega/k=\omega^{1/2}(D_y/\rho h)^{1/4}$, the latter being the phase speed of a flexural wave in a plate with stiffness given by D_y. In any case, these results confirm the efficacy and correctness of our extension of the flexural wave inversion technique to the orthotropic case.

Not unexpectedly, the actual numerical values for the D's obtained through the inversions are different from the values used in the forward finite-element calculations. At 5 kHz the ratios of the values obtained through inversion to the known values are 0.4, 0.5, and 0.6 for D_x, D_{xy}, and D_y, respectively, and at 10 kHz 0.1, 0.2, and 0.3. We surmise that these inconsistencies are related to the use of the thin plate approximation inherent in the free-space wave equation, Eq. (7), which is increasingly in error as frequency increases. For example, Cremer and Heckl (1988) argue that failure in this approximation results in less than a 10% difference in the computed flexural wave speed when the flexural wavelength, $\lambda_f>6h$. At 5 kHz, for our slab thickness we have $\lambda_f/h\sim8$ and 3.7 in the y and x directions, respectively, and the thin plate approximation begins to breakdown, albeit weakly. However, at 10 kHz we have $\lambda_f/h\sim5.9$ and 2.7 for y and x directions. The computed flexural wave speeds are now off by 10% or more, and this should begin to have a significant impact on the inversions. A rough argument as to the magnitude of this error can be made by thinking of the inversion

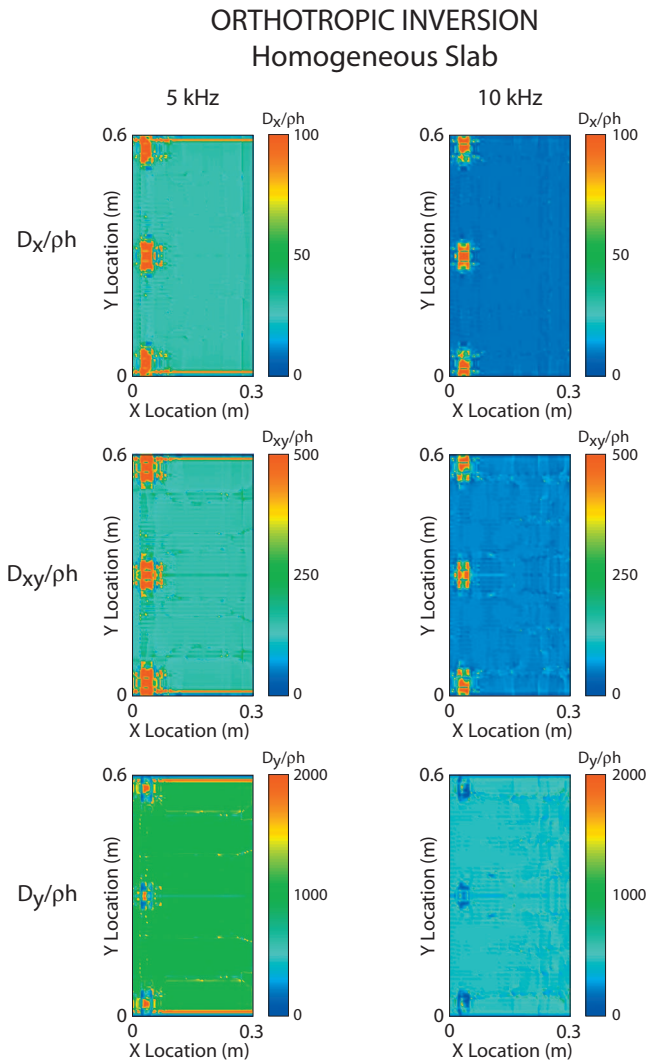


FIG. 4. Result of applying the weak flexural wave inversion operator (WFWI) developed for the orthotropic case, Eq. (11), to the displacement data at each frequency for the homogeneous wooden slab to obtain the three stiffness parameters $D_x/\rho h$, $D_{xy}/\rho h$, and $D_y/\rho h$. The magnitude of the inversion in units of the respective $D/\rho h$ values is color coded from blue to red. The three artifacts seen in each figure are due to the presence of the shakers.

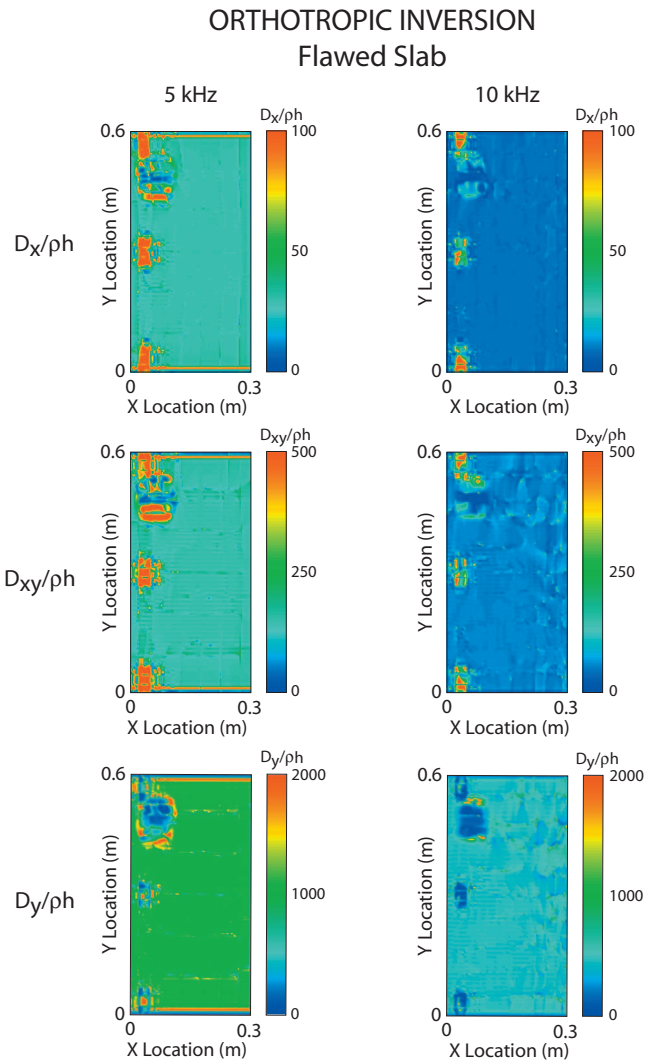


FIG. 5. Result of applying the weak flexural wave inversion operator (WFWI) developed for the orthotropic case, Eq. (11), to the displacement data at each frequency for the flawed wooden slab to obtain the three stiffness parameters $D_x/\rho h$, $D_{xy}/\rho h$, and $D_y/\rho h$. The magnitude of the inversion in units of the respective $D/\rho h$ values is color coded from blue to red. The three artifacts again seen in each figure are due to the presence of the shakers. The defect which is located in the upper left of the wooden slab can be seen in each display.

operator as a λ estimator and recognizing that a generalized modulus, M, would be given by $M=\rho C^2$ with $C=\lambda\omega/2\pi$. Accordingly, a 20% error in λ (or C) would give about a 40% error in the stiffness. We believe that this is probably responsible for the large error in the D coefficient values at 10 kHz. In any case, our goal in the inversions is to detect spatial anomalies in the local stiffness and thus to detect the presence of a defect causing this variation. In this respect, errors in the magnitude of the stiffness obtained as a by-product of the inversions are not of any serious consequence.

Consider next the generalized force method. Inserting the known (i.e., those used in the numerical calculations) values for $D_x/\rho h$, $D_{xy}/\rho h$, and $D_y/\rho h$ into Eq. (12), we now apply the GFM algorithm to the numerical displacement data. As can be seen in Fig. 6, the algorithm is very effective at detecting and locating the internal defect at 5 kHz. Once again the results at the higher frequency are worse, and this time they almost completely fail. Earlier we saw that the inversions produced values for the D coefficients which were different from those used in the calculations, presumably from the effect of the thin plate approximation, and that these differences grow with increased frequency. We expect that this effect is responsible for the very poor result observed in the GFM processing at 10 kHz. In place of inputting the known values for the D's, we subsequently used the values produced by the earlier inversions and then reapplied the GFM algorithm; this result is shown in Fig. 7. Notice that the results at 10 kHz are no longer degraded. In fact, even the already good results at 5 kHz are now improved. Apparently, differences associated with the onset of failure of the thin plate approximation have been in a sense "corrected" in an *adaptive* sense by the inversion choosing coefficients in the wave equation [Eq. (7)] which more correctly describe the dynamics at that particular frequency. In fact, we have been developing a more generalized adaptive inversion algorithm

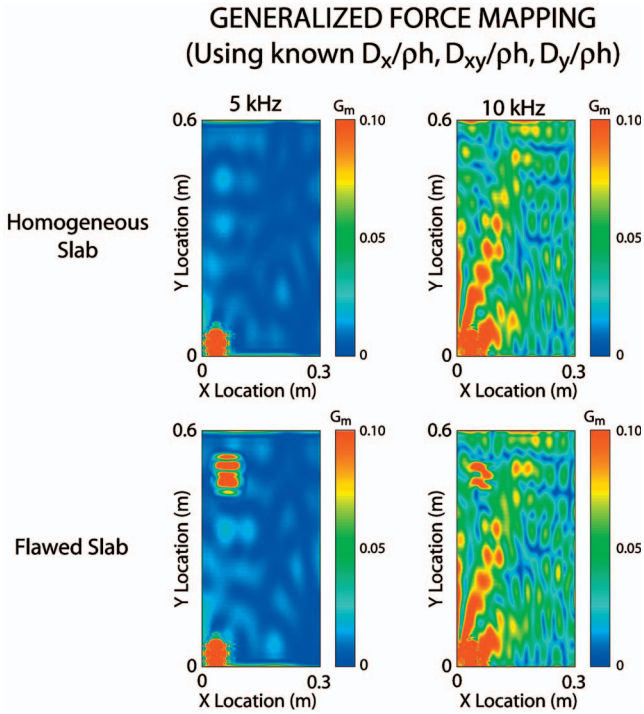


FIG. 6. Result of applying the GFM operator developed for the orthotropic case, Eq. (12), to the displacement data at each frequency for the homogeneous and flawed wooden slab when the driver in the lower left is excited. The resulting generalized force values are shown color coded from blue to red. The operator defined in Eq. (12) has been determined inserting the values for $D_x/\rho h$, $D_{xy}/\rho h$, and $D_y/\rho h$ that were used in the forward finite-element calculation of displacements. The highlight seen in each figure in the lower left is due to the presence of the shaker. The inclusion located in the upper left of the flawed wooden slab can be seen best in the 5 kHz result.

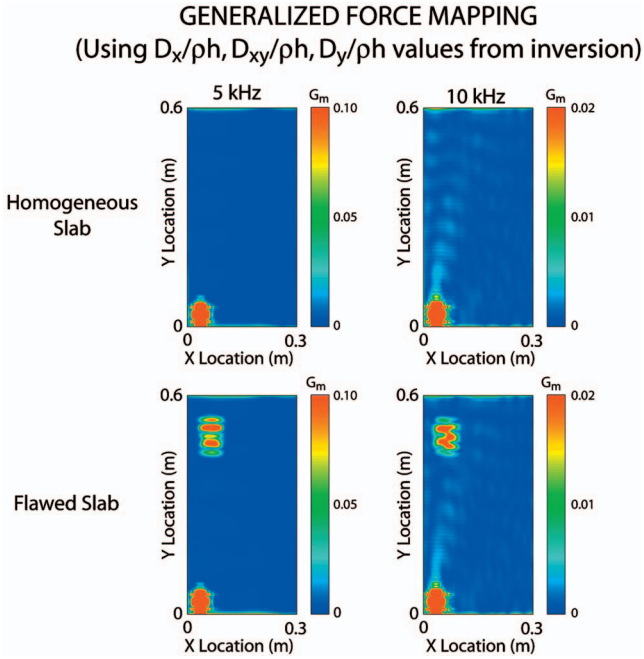


FIG. 7. Result of applying the GFM operator developed for the orthotropic case, Eq. (12), to the displacement data at each frequency for the homogeneous and flawed wooden slab when the driver in the lower left is excited. The resulting generalized force values are shown color coded from blue to red. The operator defined in Eq. (12) has been determined inserting the values for $D_x/\rho h$, $D_{xy}/\rho h$, and $D_y/\rho h$ that were obtained by applying the inversion operator to the displacement data. The highlight seen in each figure in the lower left is due to the presence of the shaker. The inclusion located in the upper left of the flawed wooden slab can be clearly seen at both frequencies.

which works in this way, and this will be the subject of another, later publication.

IV. CONCLUDING REMARKS

The orthotropic inversion and force mapping algorithms described and demonstrated here, together with their isotropic counterparts previously reported, provide a powerful set of algorithms which can convert surface vibration maps on thin plates and platelike structures to information about the local elastic properties and the presence of internal defects. In the work presented here, we have chosen excitation frequencies (5 and 10 kHz) which we believe to be experimentally practical; but we have done so with some knowledge that the corresponding structural wavelength size would be in a satisfactory range. What is generally required is that the platelike structure be mechanically excited at frequencies where structural wavelengths are short enough compared to the defect size, viz., no more than an order of magnitude larger, and that the dynamic surface displacements be mapped with a spatial resolution an order of magnitude smaller than the defect size. These two criteria loosely bound the required excitation frequency, the former criteria ensuring that the defect has a measurable effect on the surface vibration while the latter that the defect can be spatially localized.

These inversion and force-mapping algorithms can provide quantitative elastic parameter and internal defect information on structures which behave as thin solid plates. We are currently extending approaches like these to more general structures by developing and exploiting an adaptive approach to the equations of motion.

ACKNOWLEDGMENTS

This work was supported by the SERDP Program and the Office of Naval Research. One of the author (J.A.B.) would like to acknowledge the insights he gained into plate dynamics in discussions with Earl G. Williams.

APPENDIX: ORTHOTROPIC PARAMETERS

In orthotropic materials, there are nine independent moduli comprising the elastic Hooke's tensor. These materials have three orthogonal planes of elastic symmetry and are often defined in terms of the so-called "engineering constants" E_{ii}, G_{iJ}, and $\nu_{ij}^{(1)}$. In the case of Douglas coastal fir (Green, *et al.*, 1999), we have the following values for these constants:

$$E_x = 0.737 \times 10^9 \text{ Pa}, \quad G_{yx} = 1.150 \times 10^9 \text{ Pa}, \quad \nu_{xz} = 0.2868,$$

$$E_y = 14.7 \times 10^9 \text{ Pa}, \quad G_{zx} = 0.103 \times 10^9 \text{ Pa}, \quad \nu_{zy} = 0.01986,$$

$$E_x = 0.737 \times 10^9 \text{ Pa}, \quad \nu_{yz} = 0.292, \quad \nu_{xy} = 0.02245.$$

$$E_z = 1.002 \times 10^9 \text{ Pa}, \quad \nu_{yx} = 0.449,$$

$$G_{yz} = 0.943 \times 10^9 \text{ Pa}, \quad \nu_{zx} = 0.390,$$

It can be verified by inspection that these quantities satisfy the necessary symmetry relations defined as follows:

$$\frac{\nu_{ij}}{E_i} = \frac{\nu_{ji}}{E_j}, \quad i \neq j, \quad i,j = x,y,z. \tag{A1}$$

The corresponding representation for the stiffness components of Hooke's tensor (C_{ij}) can be obtained from these engineering constants using the following relationships (Rand and Roveski, 2005):

$$D0 = 1.0 - \nu_{xy}\nu_{yx} - \nu_{xz}\nu_{zx} - \nu_{yz}\nu_{zy} - \nu_{xy}\nu_{yz}\nu_{zx} - \nu_{xz}\nu_{yx}\nu_{zy},$$

$$C_{11} = E_x(1.0 - \nu_{yz}\nu_{zy})/D0, \quad C_{66} = G_{yz},$$

$$C_{22} = E_x(1.0 - \nu_{xz}\nu_{zx})/D0, \quad C_{12} = E_x(\nu_{yx} + \nu_{yz}\nu_{zx})/D0,$$

$$C_{33} = E_x(1.0 - \nu_{xy}\nu_{yx})/D0, \quad C_{13} = E_x(\nu_{yx}\nu_{zy} + \nu_{zx})/D0,$$

$$C_{44} = G_{yx}, \quad C_{23} = E_x(\nu_{zy} + \nu_{xy}\nu_{zx})/D0,$$

$$C_{55} = G_{zx}.$$

As a result of these definitions, the components of the unrotated Hooke's tensor have the values listed in Table I.

Berndt, H., Schniewind, A. P., and Johnson, G. C. (**1999**). "High-resolution ultrasonic imaging of wood," Wood Sci. Technol. **33**, 185–198.

Bucaro, J. A., Romano, A. J., Abraham, P. B., and Dey, S. (**2004**). "Detection and localization of inclusions in plates using inversion of point actuated surface displacements," J. Acoust. Soc. Am. **115**, 201–206.

Bucur, V. (**1999**). "Acoustics as a tool for the nondestructive testing of wood," NDTISS'99 International Symposium on NDT Contribution to the Infrastructure Safety Systems, Nov. 22–26, Torres, VFSM, Santa Maria, R.S. Brazil.

Bucur, V. (**2003**). *Nondestructive Characterization and Imaging of Wood*, Springer Series in Wood Science (Springer-Verlag, Berlin).

Bucur, V., and Archer, R. R. (**1984**). "Elastic constants for wood by an ultrasonic method," Wood Sci. Technol. **18**, 255–265.

Bucur, V., and Feeney, F. (**1992**). "Attenuation of ultrasound in solid wood," Ultrasonics **30**, 76–81.

Cremer, L., and Heckl, M. (**1988**). *Structure-Borne Sound* (Springer-Verlag, New York), pp. 109–115.

Dey, S., and Datta, D. K. (**2006**). "A parallel hp-FEM intrastructure for three-dimensional structural acoustics," Int. J. Numer. Methods Eng. **68**, 583–603.

Green, D. W., Winandy, J. E., and Kretschmann, D. E. (**1999**). "Mechanical properties of wood," *Wood Handbook:Wood as an Engineering Material* (USDA Forest Service, Forest Products Laboratory, Washington, DC), GTR-113, Chap. 4, pp. 1–45.

Hansen, P. C. (**1998**). *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA).

Hosten, B. (**1991**). "Reflection and transmission of acoustic plane waves on an immersed orthotropic and viscoelastic solid layer," J. Acoust. Soc. Am. **89**, 2745–2752.

Kerlin, R. (**1966**). "Internal friction studies on bell metal and wood," Technical Memorandum No. TM603.2811-05, Pennsylvania State University, University Park, PA. Technical Information Center, Accession #AD0630798.

Leissa, A. (**1993**). *Vibration of Plates* (Acoustical Society of America, Melville, NY).

Okyere, J. G., and Cousins, A. J. (**1980**). "On flaw detection in live wood," Mater. Eval. , 43–47.

Rand, O., and Rovenski, F. (**2005**). *Analytical Methods in Anisotropic Elasticity* (Birkhauser, Boston), pp. 56–58.

Romano, A. J., Bucaro, J. A., Vignola, J. F., and Abraham, P. B. (**2007**). "Detection and localization of rib detachment in thin metal and composite plates by inversion of laser Doppler vibrometry scans," J. Acoust. Soc. Am. **121**, 2667–2672.

Tomikawa, Y., Iwase, Y., Arita, K., and Yamada, H. (**1986**). "Nondestructive inspection of a wooden pole using ultrasonic computed tomography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-33**, 354–358.

Skudrzyk, E. (**1968**). *Simple and Complex Vibratory Systems* (The Pennsylvania State University Press, University Park, PA), pp. 200–214.

Nature of orchestral noise

Ian O'Brien[a)]
The Queensland Orchestra; School of Music, The University of Queensland, St Lucia Brisbane QLD 4072, Australia

Wayne Wilson
School of Health and Rehabilitation Sciences, The University of Queensland, St Lucia Brisbane QLD 4072, Australia

Andrew Bradley
School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia Brisbane QLD 4072, Australia

(Received 10 January 2008; revised 12 May 2008; accepted 14 May 2008)

Professional orchestral musicians are at risk of exposure to excessive noise when at work. This is an industry-wide problem that threatens not only the hearing of orchestral musicians but also the way orchestras operate. The research described in this paper recorded noise levels within a professional orchestra over three years in order to provide greater insight to the orchestral noise environment; to guide future research into orchestral noise management and hearing conservation strategies; and to provide a basis for the future education of musicians and their managers. Every rehearsal, performance, and recording from May 2004 to May 2007 was monitored, with the woodwind, brass, and percussion sections monitored in greatest detail. The study recorded dBALEQ and dBC peak data, which are presented in graphical form with accompanying summarized data tables. The findings indicate that the principal trumpet, first and third horns, and principal trombone are at greatest risk of exposure to excessive sustained noise levels and that the percussion and timpani are at greatest risk of exposure to excessive peak noise levels. However, the findings also strongly support the notion that the true nature of orchestral noise is a great deal more complex than this simple statement would imply.
[DOI: 10.1121/1.2940589]

PACS number(s): 43.50.Rq, 43.50.Yw, 43.75.Cd [KA] Pages: 926–939

I. INTRODUCTION

The livelihood of orchestral musicians is often dependent upon creating a great deal of sound. By repeated exposure to this sound, orchestral musicians risk suffering noise induced hearing loss (NIHL) that not only threatens their livelihood, but also their quality of life (Sataloff and Sataloff, 2006). Adding to these concerns, large orchestras produce noise levels that breach workplace health and safety laws, bringing into question the viability of the industry itself.

Noise is generally defined as unwanted or unpleasant sound. Clearly, the sound levels that orchestras create through musical performance are not unwanted; in fact they are necessary in order to continue the art form. The problem arises when the otological health of the musicians is jeopardized when performing this music. As the word 'noise' alone does not adequately describe the phenomenon, this paper will refer to high-level orchestral sound as *orchestral noise*. The report defines the nature of orchestral noise in detail.

II. RELEVANT LITERATURE

A. Review of past studies

There have been many studies into orchestral noise in professional orchestras. These can be broken down into three basic categories: level studies, loss studies, and linking studies.

"Level" studies have sought to describe the nature of orchestral noise, usually measuring noise levels in one or more orchestras and allocating potential risk to musicians based on how these levels relate to existing data regarding NIHL. "Loss" studies have sought to draw conclusions about the risks of NIHL to professional musicians by studying the otological health of the musicians themselves. "Linking" studies have sought to demonstrate a link between orchestral noise and NIHL in orchestral musicians by measuring sound levels in one (or a number) of specific orchestras and then assessing the otological health of the musicians in these orchestras.

As the current study is more concerned with the nature of orchestral noise than with attempting to establish a correlation between exposure to music and NIHL, the following literature review will focus upon previous level studies and descriptions of the orchestral noise environment presented in previous linking studies. Table I summarizes these studies and their findings. It must be noted that comparing previously obtained orchestral noise data is difficult, as these data have often been recorded from only a small sample of any given orchestra's activities. Previous studies have also used a range of sound level measures, the most common being

[a)]Electronic mail: enobrian@optusnet.com.au

TABLE I. Scope of studies to date.

Study	Venue type	Repertoire type	Number or duration of measurements	Number of data points	Microphone positions specified?	Orchestra set-up specified?	Findings
Axelsson and Lindgren 1981	Orchestra pit and concert hall	Symphonic (small to large), opera and ballet	89	14	No	No	Highest levels in brass and woodwind. Differences between positions were between 4 and 6 dB. Wide variation level over very short periods. Difference in dynamic range between venues: 85.5–92.5 dBALEQ for the pit and 77.9–94.4 dBALEQ in the concert hall.
Westmore and Eversden 1981	Concert hall, rehearsal hall and orchestra pit	Symphonic (small to large), opera and ballet	14.4 h	3	No	No	Description of microphone position is vague. Because of this the survey yields little useable information on the nature of orchestral noise apart from broad statements about the level of various pieces and the level of transient peaks (in excess of 120 dBA).
Jansson and Karlsson 1983	Concert hall, rehearsal hall and orchestra pit	Symphonic (small to large), opera and ballet	42	11?	Yes	No	Repertoire separated into heavy, medium, and light and positions into exposed and normal. Found noise levels exceeded 85 dBALEQ8 h in exposed positions after 10 h of orchestral work and after 25 h in normal positions during heavy repertoire.
Woolford 1984	Concert Hall, rehearsal hall, orchestra pit	Symphonic (small to large), opera and ballet	41	8	Yes	Yes	Exposure dependent upon repertoire and venue. Highest peaks in percussion and those in close proximity. Highest average levels amongst the brass (up to 93.4 dBALEQ) and horns (to 91.5 dBALEQ). Levels in the orchestra pit during the Tchaikovsky ballet *Swan Lake* found to be 95.9 dBALEQ in front of trombones.
Schacke 1987 (reviewed in Woolford *et al.*, 1988)	Orchestra pit	Opera	30	30	No	No	Found exposure dependent upon position, instrument played, and, less importantly, composer. 87–96 dBALEQ and peaks up to 122 dBA in the brass; 88–97 dBALEQ and peaks up to 117 dBA in the woodwind; 86–93 dBALEQ and peaks up to 110 dBA in the violins and violas; and 81–87 dBA in the cellos and basses
van Hees 1991	Concert hall, rehearsal hall	Symphonic	38	19	Yes	Yes	Found sound exposure to be dependent upon repertoire (separating repertoire into traditional and modern), venue and orchestral zone, concluded that instrument type was less significant in determining exposure levels. Extrapolated annual exposure from data.
Camp and Horstman 1992	Orchestra pit	Opera	41	17	Yes	No	Found the horn, trombone, tuba, and trumpet sections to be consistently exposed to levels above the other musicians. Levels reported in noise dose (according to 90 dBALEQ8 h) with varying exposure times. Peaks in trumpets up to 111 dBA, percussion not reported.
Sabesky and Korczynski 1995	Concert hall, rehearsal hall, orchestra pit	Symphonic (small to large), ballet	50	10?	Yes	Yes	Found highest exposure in the brass and woodwind and provide maps of the three different orchestral setups. Found mean readings of 87–92 dBALEQ. Exact positions of microphones not given. Concluded venue and repertoire account for only slight variations in exposure.
Mikl 1995	Orchestra pit, concert hall, rehearsal hall	Opera and ballet	84	6	Yes	Yes	Concluded noise exposure to musicians was dependent upon position (higher towards the rear of the orchestra); venue (worse in the orchestra pit); and repertoire. No specific data were given. Recordings limited to a representative sample from each performance.
Laitinen et al. 2003	Orchestra pit and individual rehearsals	Opera and ballet	87	10	Yes	Yes	Recorded highest noise exposure in percussion (95 dBA), brass (92–94 dBA) and flute/piccolo (95 dBA) players. Found exposure was repertoire dependent.

TABLE I. (*Continued.*)

Study	Venue type	Repertoire type	Number or duration of measurements	Number of data points	Microphone positions specified?	Orchestra set-up specified?	Findings
Lee *et al.* 2005	Orchestra pit	Opera	73	13	Yes	Yes	Found exposure repertoire dependent. Mean dBALEQ exposure for a moderately loud opera was 93.7 amongst the trumpets, 91.7 for the horns, 90.3 in the trombones, and 90.3 and 91.7 in the piccolo/flutes. Each finding was less for smaller repertoire.
Current study	Orchestra pit, reheasal hall, concert hall	Symphonic (small to large), opera and ballet	1608	27	Yes	Yes	See Table II.

LEQ, or the equivalent steady sound level that would produce the same energy as the variable sound level over the measurement period.

Overall, the studies reviewed show that orchestral noise is highly variable and that generally the brass, percussion, and some woodwind are at greatest risk of exposure to potentially damaging noise levels. However, these studies are limited by a number of factors.

The studies by Westmore and Eversden (1981), Jansson and Karlsson (1983), Woolford (1984), Royster *et al.* (1991), Mikl (1995), and Laitinen *et al.* (2003) suffered from a lack of data points (see column 5 in Table I). Attempting to accurately extrapolate noise exposure for individual positions using limited sample points is likely to provide misleading results, particularly in an orchestral environment where some of the highest noise levels are produced by instruments, such as trumpets and trombones, that have extremely directional characteristics (Chasin, 1996).

The studies by Axelsson and Lindgren (1981), Westmore and Eversdon (1981), Jansson and Karlsson (1983), Schacke (1987), Woolford *et al.*, 1988, Royster *et al.* (1991), and Camp and Horstman (1992) suffered from vague or inadequate reportage of either microphone position, orchestral setup, or both. This insufficiently describes the nature of the varying noise levels across the orchestra and prevents the results from being generalized to other orchestras.

All the studies reported previously also failed to provide an adequate number of samples to clearly represent the highly variable and complex noise environment of an orchestra.

B. Complexity of orchestral noise

The previous literature on orchestral noise cannot be generalized to the wider orchestral community as it does not adequately address the complexity of orchestras and orchestral noise. This complexity would appear to be due to two key factors: position and variability.

1. Position

"Position" refers to the physical position occupied by a musician and their instrument in the orchestra. Orchestral musicians have very specific jobs, such as second clarinet, third trumpet, or first horn. Regardless of the wider orchestral setup, where an individual musician sits relative to their colleagues is largely dictated by the job they hold and is relatively constant through a musician's career. This has a significant impact on the nature of the sound to which they are exposed. It is partly for this reason that generalizing about noise exposure by instrument family (such as woodwinds) or even by instrument section (such as clarinets) oversimplifies the nature of the noise exposure experienced by individual musicians.

Further, some positions are called on to play louder and in a higher register than most others more often. When the trumpet section plays, for example, the dominant line is usually played by the first trumpet, with the second and/or third trumpets playing more of a supporting role. In terms of pitch—in the vast majority of cases—the first trumpet will play at least a tone and a half above the second trumpet. This is similar amongst all of the instrumental sections in an orchestra. If it is accepted that the sound of an individual's own instrument significantly contributes to their sound exposure levels (a musician must hear their own instrument above all others in order to play, after all) then the implications of this arrangement are quite far reaching.

If it can be shown that musicians in some positions in the orchestra are consistently exposed to higher orchestral noise levels than their colleagues, then these data could be used to inform further research into solutions and potential management strategies to deal with this issue. Detailed maps of noise levels within a symphony orchestra may also provide greater insight into both the nature and source of excessive noise levels and allow for a thorough risk assessment for individual musicians by position.

In order to assess the positions within the orchestra that are consistently exposed to greater noise levels, each position in the orchestra must be studied separately—over a long period of time and in a variety of venues—whilst playing a variety of repertoire.

2. Variability

Variability refers to the range in type and level of music played by an orchestral musician over the course of their day-to-day work. This variability can be enormous due to changes in repertoire, venue, rehearsal format, orchestral setup, individual variations, and personnel.

a. Repertoire Orchestral noise levels are clearly dependent upon the repertoire being played. For example, Richard Strauss wrote symphonies for large orchestras of over 100 instruments often playing *fortissimo* (at full volume), whereas Josef Haydn wrote symphonies for smaller orchestras of less than 30 instruments that rarely play beyond *forte* (loud). An orchestra of 100 musicians playing at full volume is clearly going to generate substantially more orchestral noise than an orchestra of less than 30 musicians playing at a moderate volume. Aside from these extremes, the variability in noise levels from piece to piece is usually quite pronounced.

b. Venue Orchestras generally work in a variety of venues according to the type of program they are undertaking. A standard orchestra will usually have a rehearsal studio (often a large, acoustically treated hall), where the majority of rehearsals take place and occasional concerts or recordings occur. Standard performance venues are usually concert halls, orchestra pits when accompanying opera and ballet, town halls, occasionally stadiums, and a range of other minor venues.

Each venue differs acoustically and musicians will adopt different playing styles in order to project a properly balanced orchestral sound into the space in question, both at their own initiative and at the conductor's request. With the exception of the orchestra pit, stages are generally open and reflective surfaces (apart from the floor) and are generally well clear of the musicians. In orchestra pits, the ensemble is essentially enclosed at the sides, front, rear, and sometimes partially from above when the stage overhangs the pit.

Aside from the acoustic differences, different venues are used for specific activities. The rehearsal studio is used mostly for stop/start rehearsals, whereas at performance venues orchestras tend to play straight through pieces without stopping. Further, orchestras perform specific repertoire at the different venues—for instance, symphonies are generally played at concert halls but are rarely performed in orchestra pits.

c. Rehearsal format. When in rehearsal, orchestras will play en masse a great deal of the time, but will also frequently stop and work through parts of a piece with only a few musicians playing. Orchestras may also spend a great deal of time on one particular passage and play through other, simpler passages only once or twice. Depending on the length of the piece, it may not be played in its entirety until the dress rehearsal at the performance venue where adjustments will be made by the musicians under the guidance of the conductor to account for changes in acoustics.

d. Orchestral setup. An orchestra will set up in particular ways depending on repertoire, venue, orchestra size, the conductor's request, and anticipated noise exposure. An orchestra will also have several standard venue-dependent setups, the most markedly different being in an orchestral pit.

e. Individual variations and personnel. Players will play the same material at slightly differing volumes depending on many factors. Often fatigue will play a role, as will the volume of those around the individual musician. A conductor may also request greater or lesser volume as they shape the musical work. In addition, some players will play the same part significantly louder than their colleagues. Often this is merely due to stylistic interpretations on the part of the musicians; sometimes it is due to differences in the instrument and sometimes due to techniques employed in playing the instrument.

The limited nature of previous studies has meant that these variables are only partly incorporated in the presented results. In order for a study to have any statistical reliability in the face of these variables it is necessary to take samples over a very long period and to at least group the samples according to exact orchestral position and venue. Information about the nature of orchestral noise for individual positions may then be extracted taking into account orchestral activities over the sample period, and including variations in repertoire, orchestral setup, and rehearsal format.

III. PROBLEM STATEMENT

We believe the nature of orchestral noise is more complex than has previously been reported. Previous investigations have suffered either from a lack of data points, inadequate reportage of microphone positions/orchestral setup, or insufficient allowances for the variability in the orchestral playing environment. Due to large shifts in orchestral activity in a standard orchestra on a day-to-day basis, attempting to accurately extrapolate noise exposure for individual positions using limited sample points is likely to provide misleading results. In addition there is as yet no study that has comprehensively mapped the variations in noise level from musician to musician across the orchestra, particularly in the acknowledged higher-risk areas of the brass and woodwind. The lack of insight into the true nature of orchestral noise is an impediment to devising appropriate solutions to this wide-ranging problem.

IV. AIM

This study has three aims:

(1) To investigate the nature of orchestral noise in greater detail than has previously been accomplished.
(2) To act as a reference guide to future research into solutions and management strategies for the complex problem of excessive noise exposure in symphony and pit orchestras.
(3) To provide an easily understandable basis for both planning and education of musicians and their managers into the nature of the noise in their work environment and the risks (both legal and physiological) they may face.

This project is not intended to demonstrate a link between orchestral noise and NIHL. It takes as its basis for risk recommendations for safe noise exposure as dictated by current Australian occupational health and safety legislation which states an accumulated daily noise exposure in excess of 85 dBALEQ 8 hr and transient peak levels exceeding 140 dBC are likely to be hazardous. Under this legislation an employer is bound to take action if noise levels go beyond these benchmarks (AS/NZS, 2005a).

V. METHOD

Although week-to-week there is limited repeatability in an orchestral musician's activities, year-to-year these activities are quite similar. This is because orchestras tend to plan their activities annually with the aim of maintaining various

playing series and any pit services. In order to represent this accurately the current study has taken a long-term approach. The data collection described covers the period from May 2004 to May 2007 at The Queensland Orchestra.

A. Orchestra

This study obtained all of its recordings from The Queensland Orchestra (TQO). TQO is the only full-time professional orchestra in Brisbane, Australia and, as such, it has a varied role. It performs, on average, 95 times per year, which includes the orchestra's subscription series, pops concerts, small ensemble concerts, pit services to the local and national ballet and opera companies, collaborations with touring artists, and festival events. The orchestra employs 86 musicians and is Australia's third largest orchestra. It has a national and international reputation as a high quality ensemble.

B. Venues

When rehearsing, TQO is usually based at its rehearsal studios (Studio 420) where it also occasionally performs a concert, recording, or radio broadcast. When performing, TQO spends roughly one-third of each year working in the orchestra pit of the Lyric Theatre for resident and touring opera and ballet companies, roughly one-third of a year in Queensland Performing Arts Centre's (QPAC's) concert hall and the remaining one-third of each year is spent in assorted other venues such as Brisbane's City Hall, QPAC's Playhouse orchestra pit and the Queensland Conservatorium Concert Hall. Throughout the year each program is rehearsed in the orchestra's rehearsal studio before taking it to the relevant venue.

1. Studio 420

This rehearsal venue is a large, acoustically treated, purpose-built orchestral rehearsal hall with a floor space of 28 m by 20 m and a height of 14 m. The floor is wooden parquetry and the walls and ceiling are constructed with large preformed diffusive plaster blocks. The orchestra can set up in Studio 420 in various ways, but most often the configuration is identical to the set up it uses in QPAC's Concert Hall, including riser heights, screens, and distances between instrument sections. When rehearsing opera or ballet, however, the orchestra will assume the configuration it uses in the orchestra pit. In addition to rehearsals, the orchestra uses the studio as a venue for formal concerts, children's concerts, and recordings. In terms of repertoire, the studio sees the greatest variability of all the venues—opera, ballet, symphonies, pops programs, small ensemble work, and so on.

2. QPAC Concert Hall

The stage of QPAC's Concert Hall measures 15 m deep and 18 m wide. The floor is made from polished wooden floorboards and the surrounding walls are a mix of concrete and wood. The stage is open to the 1800 seat auditorium to the front and has tiered choir stalls to the rear. The Concert Hall is where the orchestra performs the vast majority of its main stage concerts. These can be a standard symphony orchestra program, light classical concerts, educational programs, festivals, special events, pops, and Christmas programs. The orchestra uses the venue for performance and for rehearsal.

3. QPAC Lyric Theatre orchestra pit

The orchestra pit at QPAC's Lyric Theatre is 18 m at its widest point, 5 m from front to back and 2 m below the level of the stage, with the stage overhanging the pit by 2 m. The rear of the pit is treated with diffusive panels and the floor and surrounding walls are made from painted plywood. It was built, along with the rest of the venue, in the early 1980s when noise concerns were only just beginning to be studied in any depth. There is, however, enough room for some engineered controls, such as screens and a small riser (a raised platform upon which the musicians sit) for the back row of woodwinds.

4. Other venues

The City Hall (Brisbane) is an open wooden floored stage in a large auditorium and the Queensland Conservatorium Concert Hall is similar. Both have very open stages similar to the concert hall. QPAC's Playhouse pit is similar to the Lyric Theatre pit, but smaller. Samples from these venues are included in the overall database but were not extracted specifically due to their less frequent use by the orchestra.

C. Recording sessions

Noise monitoring was planned on a project-by-project basis between May 2004 and May 2007. During this period, many of TQO's orchestral projects consisted of four rehearsals, a dress rehearsal, and a performance. In the case of an opera or ballet there were slightly more rehearsals and as many as 10 or 15 performances. As the orchestra spent a great deal more of its time in rehearsal than performance, it was decided at the outset that both performances and rehearsals should contribute to the data. Occasionally the orchestra split to play more than one project at one time. In cases like these, preference was given to the project judged to be more likely to exceed legally allowable noise exposure.

D. Equipment

Three Cassella USA CEL-460 data-logging dosimeters were used for this project. Each unit was programed to record noise dose, dBALEQ, and dBC peak values. Noise dose was calculated using an exchange rate of 3 dB and a criterion of 85 dBALEQ 8 h according to the Australian/New Zealand Standard (AS/NZS, 2005b).

For all dosimetry runs the units were left on for the duration of the orchestral call—including breaks and intervals—in order to give a more accurate indication of the noise exposure of the musicians. The units were programed to automatically begin recording at the start of the rehearsal/performance (call) and to switch off at the end of the call to reduce possible artifact by handling noise. If the call finished early, the units were manually switched off *in situ*. Immediately prior to and at the conclusion of each and every run the

units were individually calibrated using a matching CEL-282 acoustic coupler to account for temperature and humidity fluctuations. The units were professionally recalibrated, serviced, and checked for faults annually according to the manufacturer's specifications.

E. Positioning of dosimeters

The three dosimeters were placed in different positions during different recording sessions until sufficiently detailed recordings were obtained for all chosen positions in all venues. At the start of each orchestral project a series of dosimeter positions was devised, taking into account any possible variations, such as orchestral setup and repertoire to be rehearsed or performed. Instrument groups that indicated consistently high levels of exposure were monitored in more detail than those with consistently low levels, but a broad cross section of recording positions was the aim for each orchestral project.

Each dosimeter was positioned on a microphone stand for the duration of the call. The microphone of each dosimeter was placed on a boom at the ear level of the musician being monitored, with the microphone itself positioned greater than 20 and less than 60 cm from his or her most exposed ear. If a wraparound head screen (made from absorbent high density foam) was in place, the microphone was positioned within the screen. In the case of percussionists, a stand-mounted unit was not practical as the musicians constantly moved from instrument to instrument.

Due to this, percussionists wore the units on their belt with the microphone mounted high on the shoulder of what the musician judged to be his or her most exposed ear. When surveying these data, it is important to remember that this may have caused levels amongst the percussion to appear slightly elevated due to the reflection of sound back from the head and neck. It may also have increased the likelihood of artifacts and for this reason percussion readings were error checked in greater depth than other stand-mounted readings.

F. Data gathering/record keeping

The following data were gathered from each recording session: venue, orchestral project, repertoire, precise position of dosimeter (including left or right ear), duration, the equivalent steady sound level present for the duration of the call (expressed as dBALEQ), the C-weighted peak, and the noise dose expressed as a percentage.

The dosimeters were programed to plot a new data point on a profile graph sampling every 30 s. This plot was used in analysis to eliminate artifacts, such as accidental bumping of the microphone by a musician or operations staff. Artifacts were easily identified by comparing peak level graphs from the three units to determine whether a "spike" could be seen on all units at the same point in time. If not, the next highest peak was recorded.

At the end of each dosimetry run, data from the units were entered onto a paper record and information from the units was downloaded to computer every three or four days. The data were also entered into a database for later analysis and to check for anomalies. Orchestral layout maps were

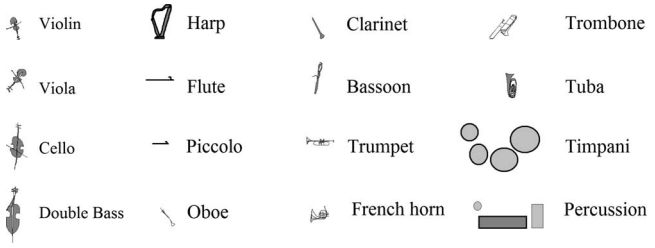
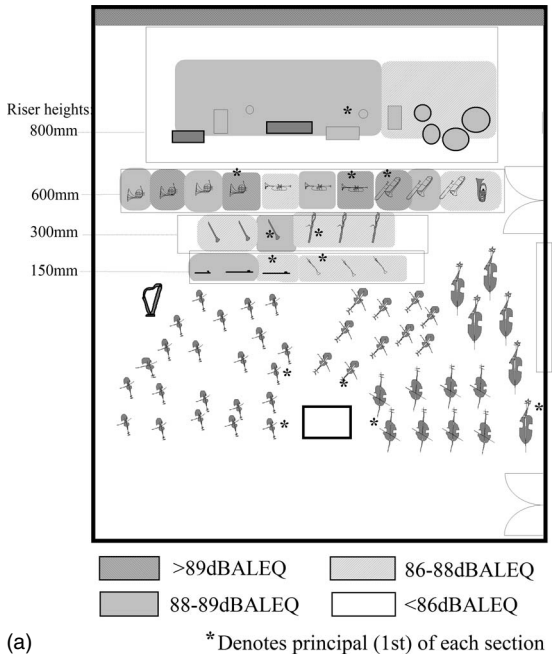


FIG. 1. Key to instruments in sound maps.

stored with the data to indicate the position of the readings relative to the rest of the orchestra. At the time of the final analysis, the database held in excess of 1600 separate noise-monitoring sessions.



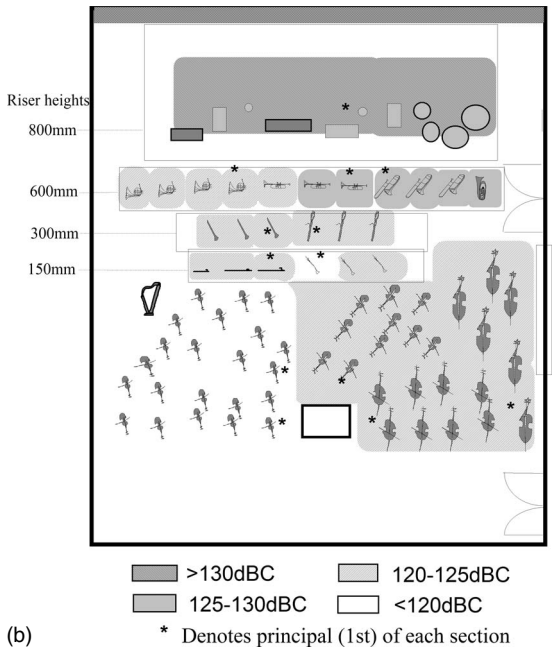


FIG. 2. (a) Mean LEQ—all venues, 2004–2007 and (b) median dBC peak—all venues, 2004–2007.

TABLE II. Summarized data for all readings, 2004–2007.

Position	dBALEQ mean	dBA range	dBC peak median	dBC peak range	Number of samples
Violin 1	84.4	77.4–90.6	119.6	107.1–131.2	24
Violin 2	84.7	78.7–90.7	119	111.1–134.4	62
Viola	85.3	76.1–94.9	121.8	103.7–131.6	78
Cello	84.5	76.2–89.3	121.1	108.9–135.9	49
Bass	84.3	78.1–91.7	122.7	114.4–135	75
Harp	85.2	77.3–90.5	118.6	113.4–127.4	24
Flute 1	87.8	78.1–93.1	121.2	106.5–132.8	84
Flute 2/Piccolo	88.2	80.2–94.4	121	116.7–131.3	63
Oboe 1	87.1	80.5–91.5	119.8	110.8–130.4	48
Oboe 2/Cor Anglais	87	81.3–92.3	120.7	112.3–127.6	42
Clarinet 1	88.5	80.4–93.9	122.8	112.1–135.5	111
Clarinet 2/bass clarinet	86.6	79.8–91.7	120.7	113.6–129.8	42
Bassoon 1	87.7	79.9–93.2	122.9	98.7–138.2	77
Bassoon 2/contra bassoon	87.5	79.7–94.4	124.2	116.1–139.2	57
Trumpet 1	89.8	82.1–95.1	125.3	108.5–137.7	128
Trumpet 2	88.9	80.2–95.3	126.1	115.3–138.3	60
Trumpet 3	87.9	79.7–93.6	124.1	118–131.3	16
Horn 1	89.5	81.4–95.2	122.8	107.3–133.1	119
Horn 2	88.8	81.6–95	122.1	116.5–131.1	46
Horn 3	89.3	82.7–95.9	122.8	117.4–133	71
Horn 4	88.7	82.9–93.5	122.5	118.8–133.8	26
Trombone 1	89.1	82.4–94.1	127.7	117.1–136.1	89
Trombone 2	88.8	78.3–95.4	129.5	118.9–137.1	31
Bass trombone	87	80.9–93.6	126.9	113.4–135.7	34
Tuba	86.9	78.5–92.1	127.9	113.6–137.4	23
Percussion	88.8	81.5–96.1	135.5	120–146.9	66
Timpani	87.7	80.5–96.3	132.9	122.5–144	63
				Total samples:	**1608**

For the purposes of this study a mean average of the recorded A-weighted equivalent sound levels (LEQ) and a median of the C-weighted peaks were used in describing the orchestral environment. The use of the A-weighting to describe the equivalent sound level was chosen as workplace health and safety legislation refers to allowable levels in A-weighted decibels (AS/NZS, 2005a) and the majority of recent studies into orchestral noise use the "A" weighting, making comparisons more meaningful. The use of the "C" weighting in reporting the peak level was also due to current Australian reporting requirements (AS/NZS, 2005a).

LEQ or level equivalent is an expression of the steady sound level that would produce the same energy as the fluctuating level actually occurring (AS/NZS, 2005a). As orchestral musicians work irregular hours (call times are often shortened and actual hours worked over the course of a week vary markedly from player to player) using time-dependent expressions, such as noise dose or dBALEQ8 hr, is less useful than providing a simple dBALEQ, which is more indicative of actual levels experienced. Actual exposure and projected exposure can easily be extrapolated for specific positions using the base dBALEQ figure and adjusting for the amount of hours an individual musician has worked or will be working over any given period.

G. Sound maps

Sound maps were generated upon completion of the data gathering. The dBALEQ and dBC peak levels for each instrumental position in all venues and in each of the three key venues was extracted from the database, averaged (mean average dBALEQ and median average dBC peak) and ranked. Each instrument was individually rendered and the most common orchestral setup for each venue was graphically represented. Instrument positions were then shaded according to the range of exposure indicated by the averaged results.

H. Ethics

Unconditional ethical clearance to conduct this study was granted by the Director of Human Resources at The Queensland Orchestra and the University of Queensland's Behavioural and Social Sciences Ethical Review Committee (reference number 2008000031).

VI. RESULTS

Data are presented both overall and by key venues, with details of venue, setup, repertoire common to venue (including style of rehearsal/performance) and acoustics. Included for the sake of illustrating the impact of repertoire upon noise exposure is a graph of a single performance at QPAC's concert hall.

The sound maps are based upon data summarized from the complete database. The shading of the maps indicates increasing levels of exposure according to the attached legend. When viewing these maps it must be remembered that this is an averaged level, some calls may have exceeded this

and some may have been much less, according to the key variables of repertoire and venue. Range information is given in the relevant tables.

There are only a limited number of occasions where noise peaks were recorded above the legislated allowable level of 140 dBC (all in the percussion and timpani), but an illustration of peak levels is essential in determining the nature of orchestral noise in a particular section of the orchestra. For instance, a low dBALEQ combined with a high dBC peak would indicate the presence of a quieter instrument in close proximity to an instrument with high transient peaks such as a double bass situated in front of the percussion. This can critically affect a musician's comfort levels and their attitudes to the sounds they are hearing. Due to existing evidence of the subjective nature of hearing loss, particularly in relation to music (Chasin, 1996) the comfort of musicians is of great importance in fully assessing the potential impact of noise exposure.

Each of the maps represents the instruments of the orchestra according to Fig. 1.

A. All venues

Figures 2(a) and 2(b) and Table II are the sound maps and summary data for all readings taken at all venues between May 2004 and May 2007 inclusive. This includes a great array of venues, repertoire, and orchestral setups but is represented in the orchestra's most common setup.

B. Rehearsal studio (Studio 420)

The greater variability in repertoire played by TQO in its rehearsal Studio 420 (opera, ballet, symphonies, pops programs, small ensemble work, and so on) was reflected in the dBALEQ ranges recorded, with almost all positions having a range in excess of 10 dBALEQ—the largest range of all the venues. Compared with the summarized data from all venues, results specific to the rehearsal studio show a slightly lower mean exposure in all positions across the orchestra [see Figs. 3(a) and 3(b) and Table III].

C. QPAC Concert Hall

In the Concert Hall, average recorded dBALEQ's in the strings are lower than the average when compared with the total recorded data, whereas brass, woodwind, and percussion are generally higher than average. This result in the strings may be due to the size of the concert hall platform, allowing the players to be seated a reasonable distance from the brass and percussion. The result in the brass and woodwinds is most likely due to the repertoire performed in the concert hall by the orchestra and a further indication of the impact of the musician's own sound in exposure assessment. A standard orchestral concert will, more often than not, commence with a lively piece such as an overture followed by a concerto with a soloist and finish with a large-scale symphonic work [see Figs. 4(a) and 4(b) and Table IV].

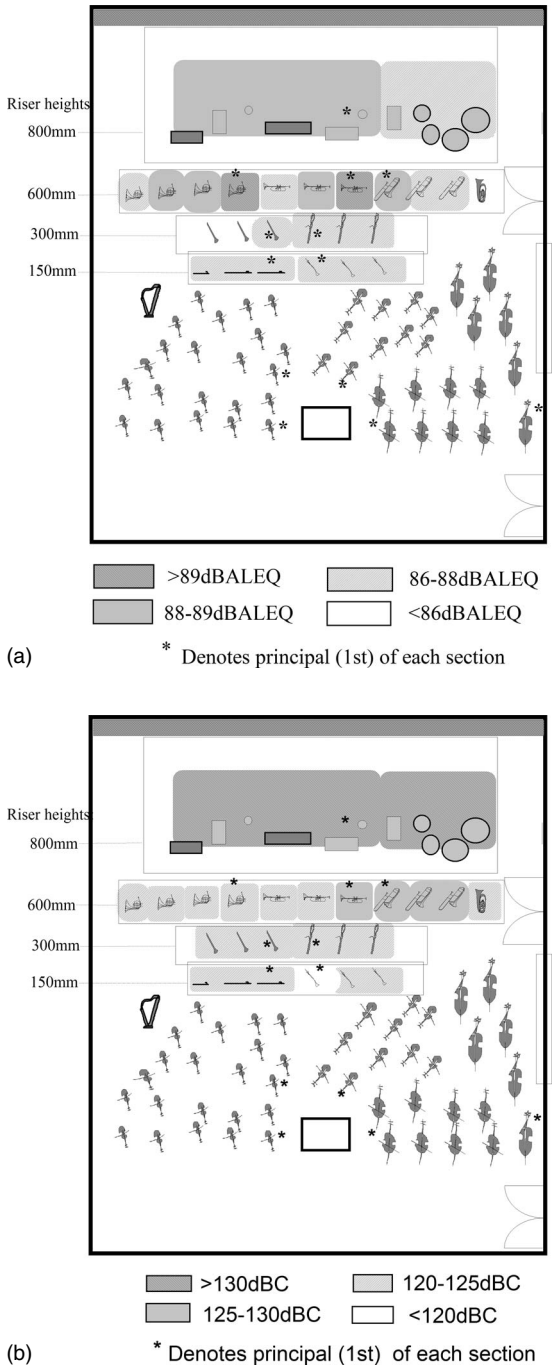


FIG. 3. (a) Mean LEQ—Studio 420, 2004–2007 and (b) median dBC peak—Studio 420, 2004–2007.

D. QPAC Lyric Theatre orchestra pit

Of all the orchestra's venues, it is clear that the orchestra pit is the site of the highest sustained orchestral noise levels. With the exception of percussion and timpani the averaged dBALEQ's of each instrument in the pit is higher than the overall average summarized in the study. This is likely to be due to the nature of opera and ballet music, which uses percussion sparingly, but usually to dramatic effect, whereas many of the other instruments are usually occupied for the entire length of the performance. The depth of the pit and its position partly under a stage may be partially responsible for the increased levels, particularly toward the back of the orchestra [see Figs. 5(a) and 5(b) and Table V].

TABLE III. Studio 420 summarized data.

Position	dBALEQ average	dBALEQ range	dBC peak median	dBC peak range	Number of samples
Violin 1	81.2	77.4–85.5	115.9	107.1–121.2	4
Violin 2	83.8	78.7–88.3	118.3	111.1–126.6	19
Viola	85	77.6–91.2	119.2	103.7–126.4	27
Cello	84	78.6–87.3	119	113–123.1	15
Bass	83.4	78.4–86.7	119.8	115.1–129.6	21
Harp	84.3	82.5–87.6	117.6	115.7–127	9
Flute 1	87.3	78.1–92.1	120.6	106.5–129.5	47
Flute 2/piccolo	87.1	80.2–92.9	120.4	116.7–127.1	23
Oboe 1	87	80.9–90.7	119.8	112.7–128.6	32
Oboe 2/cor	86.6	81.3–90.9	120.3	112.3–127.6	19
Clarinet 1	87.8	80.4–93.2	122.6	112.1–129.8	60
Clarinet 2/bass Cl	85.7	79.8–89.3	120.4	113.6–129.8	22
Bassoon 1	87.4	80.1–92.8	121.6	112.5–132.3	40
Bassoon 2/contra bassoon	86.9	79.7–94.4	123.5	116.1–129.1	24
Trumpet 1	89.4	83.1–94.9	125.6	119.9–136.6	66
Trumpet 2	88.6	81.1–94.5	124	116.2–138.3	26
Trumpet 3	87.2	79.7–91.6	123.2	118–128	10
Horn 1	89.1	81.4–94.5	122.4	107.3–132.9	67
Horn 2	88.2	81.6–94.1	122.5	116.5–128.9	23
Horn 3	88.9	82.7–95.9	122.7	117.4–133	37
Horn 4	87	84.5–88.9	120.6	118.8–121.2	6
Trombone 1	88.7	82.4–93.6	126.5	117.1–135.3	51
Trombone 2	87.7	78.3–91.7	126.4	118.9–134.2	12
Bass trombone	86.3	80.9–93	126.4	119.4–133.7	15
Tuba	84.6	78.5–90.4	124.9	113.6–128.7	10
Percussion	88	82.3–93.1	135.1	125.5–144.1	31
Timpani	86.2	81.2–91	132.9	123.8–141.3	26
				Total samples:	**742**

E. Impact of repertoire

To illustrate the impact of repertoire on an individual position's noise exposure, Fig. 6 illustrates a reading taken from the first trumpet (right ear) position in a standard orchestral program at QPAC's Concert Hall featuring contrasting repertoire.

The program in this instance began with Weber's Overture to *Oberon* (9 min) followed by Prokofiev's *Violin Concerto No. 2* (26 min), followed by a (20-min) interval and then Tchaikovsky's *Symphony No. 4* (44 min). For the duration of the first piece—which is lightly scored but featuring trumpet flourishes and fanfares—the level equivalent (LEQ) is 94.1 dBA; for the violin concerto—which is also lightly scored but a quieter, less dramatic piece—the LEQ is 81.9 dBA; and for the symphony—a heavily scored, dramatic piece with some quiet passages and one quiet movement—the LEQ is 95.7 dBA. This left the first trumpet position with an overall exposure of 93.8 dBALEQ for a performance just short of 2 h inclusive of breaks. Note the sustained nature of the noise levels, particularly around 21.45, during the final movement of the symphony where the brass are called upon to play long, loud lines.

F. Transient nature of percussion/timpani

Percussion and timpani orchestral noise was characterized by extremely high-level transient peaks and medium to high LEQ's. Figure 7 is a typical timpani reading for a standard symphonic program at QPAC's concert hall.

This program was an overture, followed by a Rachmaninoff Piano Concerto and Dvorak's *Symphony No. 9*. The dramatic nature of the overture (to Smetana's opera *The Bartered Bride*) is seen by the dotted peak line, which hits 132.8 dBC and is the loudest point for the program. When compared to the trumpet readings from Fig. 6 the different nature of orchestral noise exposure for percussionists becomes apparent.

G. Assessment of risk according to noise exposure

In order to further quantify the risk faced by orchestral musicians, the findings may be extrapolated to risk levels according to Australian workplace health and safety regulations. According to this legislation, one single noise dose (100% DND) is defined as 85 dBALEQ over 8 h using a 3 dB exchange (AS/NZS, 2005a). Once a mean average dBALEQ has been determined—giving an indication of expected noise exposure per orchestral call for any given position—expected average dose may then be estimated. Common to most orchestras, The Queensland Orchestra works in "calls" of either 2.5 h (usually a rehearsal) or 3 h duration (usually a performance). Table VI gives an indication of how average LEQs as depicted in Figs. 2(a), 3(a), 4(a), and 5(a) may translate to levels of risk for individual

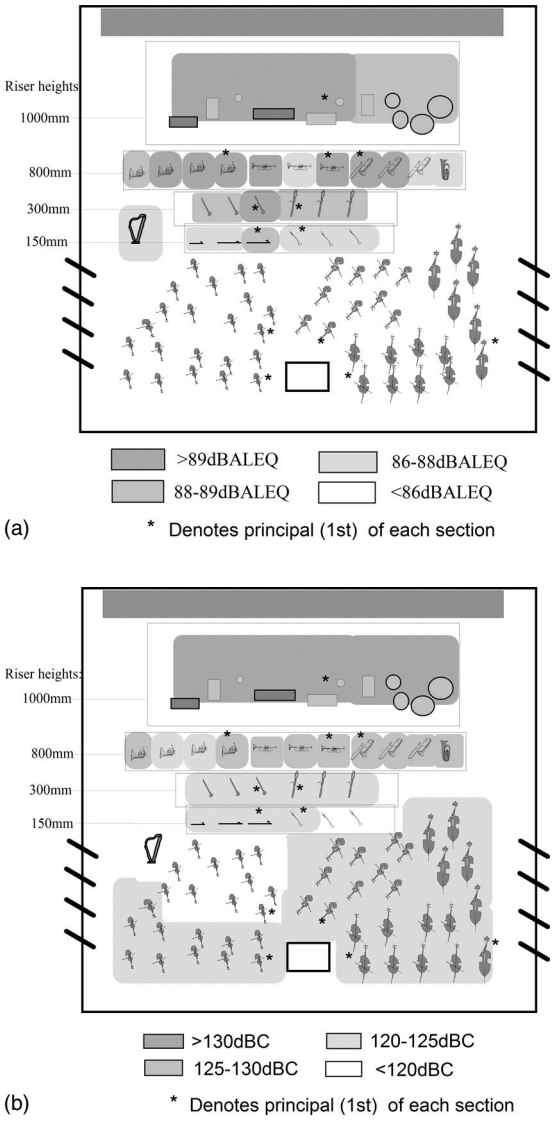


FIG. 4. (a) Mean LEQ—QPAC Concert Hall, 2004–2007 and (b) median dBC peak—QPAC Concert Hall, 2004–2007.

positions. Each of these figures may be viewed using Table VI in order to ascertain an estimate of expected noise dose.

VII. DISCUSSION

This study had three stated aims. To achieve the first aim—to investigate the nature of orchestral noise in greater detail than has previously been accomplished—data from noise readings of three years of orchestral activities have been collated and presented both graphically and as summary tables according to key venues and overall. It is clear from these data that different positions in the orchestra—even those adjacent and playing the same instrument—are exposed to different levels of risk. Although this has been found to some extent in previous studies, no study had yet uncovered the true detail of these variations.

In order to interpret these data presented and to achieve the second and third stated aims of the study—to further guide future research and to provide a basis for planning and education of musicians and their managers—the discussion will proceed according to instrument type.

A. Strings and Harp

String players on the rear desks (i.e., at the back of the various sections) were generally the positions reported on in this study due to their proximity to high sound emitting instruments. In cramped quarters (such as the orchestra pit) the rear desks of the violins were usually positioned directly adjacent to the piccolo, flutes, and the horns in the case of the violins, and in front of the brass and percussion in the case of the cellos, violas, and double basses. In the orchestra pit, noise spikes from the percussion and brass account for the high peak levels amongst the basses and cellos. Although the recorded LEQs of the strings are generally the lowest of the orchestra, string players state that it is the presence of high peaks from other instruments that causes them the most discomfort.

The harp faces similar issues to the rear desks of the violins. The harp was exposed to higher than average levels in the pit—with peaks up to 123.3 dBC and one dBALEQ of 89.5. This was probably due to the proximity of both the horn section and the piccolo.

Apart from these hazards however, the strings and harp were consistently exposed to lower noise levels when compared to the rest of the orchestra. This does not mean that they were never exposed to potentially hazardous noise levels, as can be seen from the upper end of the ranges, but they were not as regularly at risk as their colleagues in the woodwind, percussion, and brass. The highest level recorded in the strings was a dBALEQ of 94.9 at the rear of the viola section in QPAC's Playhouse orchestra pit, where the violas were in very close proximity to the trumpet section. This measurement was repeated with close to identical results.

Violinists and violists hold their instruments very close to their left ear. Due to this, the dosimetry microphone was unable to accurately measure the true exposure to this ear for these players. It is likely that there may be a greater risk to these players from their own instruments than the results of this study indicate. This increased unilateral exposure to violinists and violists is supported by Axelsson and Lindgren (1981), but is beyond the scope of this investigation.

B. Woodwinds

More often than not the woodwinds sit in two rows. In the pit setup the woodwinds were against the rear wall of the pit, but in other venues the brass and percussion were generally directly to their rear. In the back row, the first clarinet and first bassoon typically registered higher exposure than their seconds (a pattern common throughout the winds and brass). As already discussed, this difference can probably be attributed to the higher register of the parts played by the first players.

Levels in the pit for the clarinets were higher than at other venues, despite the pit being the only main venue where no brass or percussion were directly behind them. Bassoons on the other hand, seated next to the clarinets, received their lowest averages in the pit. This emphasizes the extent to which a musician's own instrument contributes to their noise exposure.

TABLE IV. QPAC Concert Hall summarized data.

Position	dBALEQ mean	dBALEQ range	dBC peak median	dBC peak range	Number of samples
Violin 1	85.3	84.2–86.2	124.9	124.7–125.1	4
Violin 2	84.1	80.6–89.3	119.4	115.1–134.4	23
Viola	84.2	76.1–92.6	120.4	106.9–127.5	22
Cello	83.6	76.2–89.3	122.2	108.9–127	7
Bass	84.2	78.1–89.6	120.7	114.4–128.8	23
Harp	86	82.7–90.2	118.9	114.6–127.4	7
Flute 1	88.1	81.6–91.4	121.7	113.9–130.1	21
Flute 2/piccolo	87.6	82.6–91.2	121.7	117.7–127.1	14
Oboe 1	87.4	80.5–91.3	120.1	110.1–130.4	8
Oboe 2/cor	87.8	83.1–92.1	119.7	116.3–127.4	8
Clarinet 1	89.3	86.8–93.9	123.5	119.7–130.8	28
Clarinet 2/bass clarinet	88.1	83.9–90.5	123.4	119–126.8	11
Bassoon 1	88.5	79.9–93.2	123.9	98.7–134.4	22
Bassoon 2/contra bassoon	88.2	82.8–91.6	124.9	120.5–127.9	17
Trumpet 1	90.2	86.2–94	128.75	119.5–137.5	28
Trumpet 2	87.7	80.2–92	126.2	115.3–134.6	13
Trumpet 3	89.9	88.2–93.6	126.5	122.7–133.3	4
Horn 1	89.9	85–93.3	126.7	121.6–133.1	27
Horn 2	89.1	83.5–95	123.9	117.7–131.1	11
Horn 3	89.5	86.6–93.1	124.1	119.9–128.6	17
Horn 4	88.9	82.9–93.5	125.6	122.7–133.8	12
Trombone 1	89.5	84.2–94.1	128.5	122.5–134.2	23
Trombone 2	89.4	86.6–93.9	129.5	125.2–132.5	9
Bass trombone	87.4	81.5–93.6	125.9	118.1–133.6	6
Tuba	87.7	83.9–91.5	129.8	123.3–132.9	5
Percussion	89.7	81.5–96.1	138.9	121.3–146.9	14
Timpani	88.6	80.5–96.3	133.1	128.4–137.7	17
				Total samples:	**401**

In the front row of the woodwinds, flute 2/piccolo readings highlight the importance of register in noise exposure. In this case, the piccolo—which is used a great deal in the pit and plays in a register well above the first flute—pushes the second player's exposure level beyond that of the first player. All flute and piccolo readings indicate that their right ear was the most vulnerable, being the side on which their instrument is held. The oboes have the lowest levels of the woodwind section, and experienced their highest levels when playing standard symphonic programs in the Concert Hall.

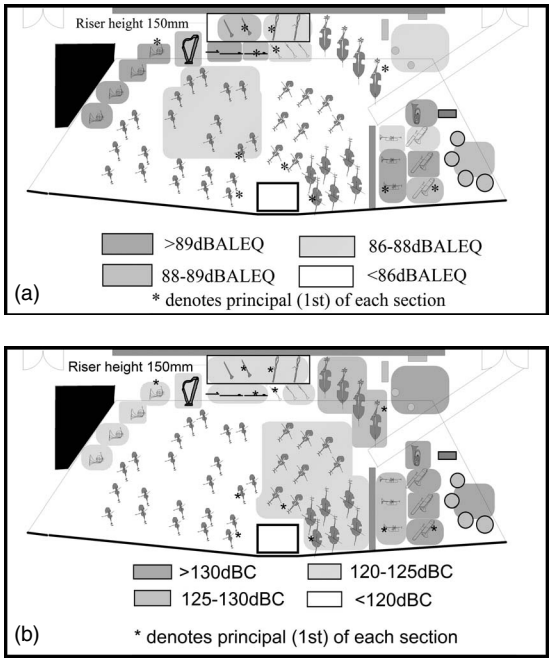


FIG. 5. (a) Mean LEQ—QPAC Lyric Theatre pit, 2004–2007 and (b) median dBC peak—QPAC Lyric Theatre pit, 2004–2007.

C. Brass

Despite the obvious hazard of the percussion directly to the rear, brass levels were consistently high because of the volume and nature of the sound their instruments generate. The impact of the percussion on the brass is generally seen as higher noise peaks rather than increasing their LEQ, due to the sporadic use of percussion in much orchestral music. This can be clearly illustrated by brass readings in a program such as Brahms's *Symphony No. 3*, which has no percussion except timpani, yet still brass readings were 90.2 dBALEQ for first horn and 90.6 dBALEQ for first trumpet. The brass typically create sustained medium to high dBALEQ levels, driving up their noise exposure significantly.

With the exception of the first trumpet, the horns (more specifically, the first horn) were the group most consistently exposed to very high LEQs. The design of the horn (the bell of the instrument faces the right and the rear) invites obvious problems for any horn player's right ear and to a lesser extent for the left ear of the next player in line (players sit at

TABLE V. QPAC Lyric Theatre pit summarized data.

Position	dBALEQ mean	dBALEQ range	dBC peak median	dBC peak range	Number of samples
Violin 1	85.1	80.1–90.6	118.5	113.4–131.2	11
Violin 2	86.7	83.9–90.7	119.9	114.5–130.4	10
Viola	85.3	82.3–87.8	122.5	114.5–130.5	13
Cello	85.2	82.2–89.1	122.4	116.6–135.9	16
Bass	84.3	80–88.5	126	118.5–135	23
Harp	88.8	87.9–89.5	121.3	118–123.3	3
Flute 1	89	84.9–93.1	121.4	117–127.7	10
Flute 2/piccolo	89.4	87.1–93.5	121.4	118–125.8	13
Oboe 1	86.2	83.7–88.7	117.9	114.8–120.6	4
Oboe 2/cor	87.2	84.1–90.2	121.5	118.9–126.9	10
Clarinet 1	88.7	84.7–91.9	122.1	116.1–126.3	10
Clarinet 2/bass clarinet	89	88.1–89.8	120.8	119.3–124.2	3
Bassoon 1	87.9	85.2–90.4	120.1	107.2–130	6
Bassoon 2/contra bassoon	86.3	85–88.5	122.3	118.4–131.7	6
Trumpet 1	90	83.6–94.8	127	120.8–135.8	16
Trumpet 2	89.9	83.3–95.3	127.7	119.6–136.3	14
Trumpet 3	87.2	82.1–92.3	125.1	124.6–125.6	2
Horn 1	89.8	85.4–93.5	123.1	120.4–126.9	11
Horn 2	89.6	85.5–92.6	121.2	119.1–123	10
Horn 3	90.7	88.2–93.3	122.7	119.8–126.9	8
Horn 4	89.3	87.3–91.9	121.2	119.8–123.8	7
Trombone 1	89	85.5–91.2	131.6	125.6–134.6	8
Trombone 2	89.6	85.1–95.4	132.1	126.4–137.1	9
Bass trombone	87	82.6–92.2	130.1	125–135.7	9
Tuba	90.1	89.2–91.7	130.6	129–134	4
Percussion	88.5	86.4–91.2	131.8	124.8–144	8
Timpani	86.7	82–96.1	132.1	125.5–144	9
				Total samples:	**253**

least a metre apart). In a horn section the first and third are the "high" players in terms of register, whereas the second and fourth are the "low" players. This is reflected in the results of the study with uniformly higher exposure levels for the first and third horns over the second and fourth players. The first and second received their highest average levels in the concert hall, whereas the third and fourth received their highest levels in the lyric theatre pit. This may be due to the third and fourth often having to play with their bells facing the sidewall in the pit. Peak levels amongst the horns were generally higher when the percussion was to their rear, such as in the Concert Hall or Studio 420, but peak levels in the pit—where the percussion are on the opposite side (see Fig. 5)—also registered around 123 dBC for the first and third, and around 121 dBC for the second and fourth. Each of the four horns was consistently exposed to higher LEQs at their

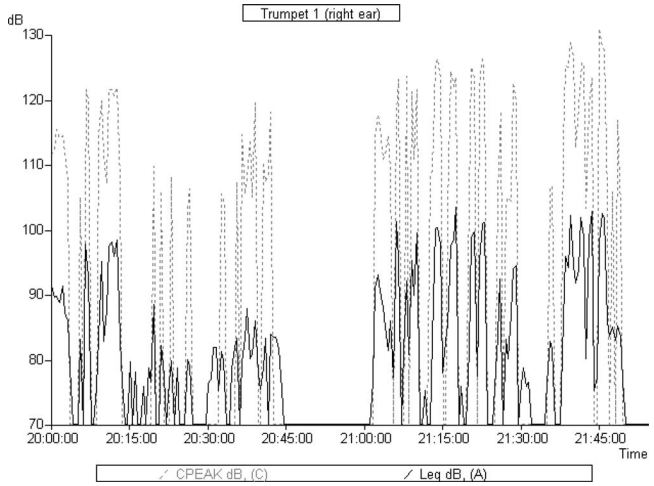


FIG. 6. Exposure to first trumpet during a concert at QPAC Concert Hall.

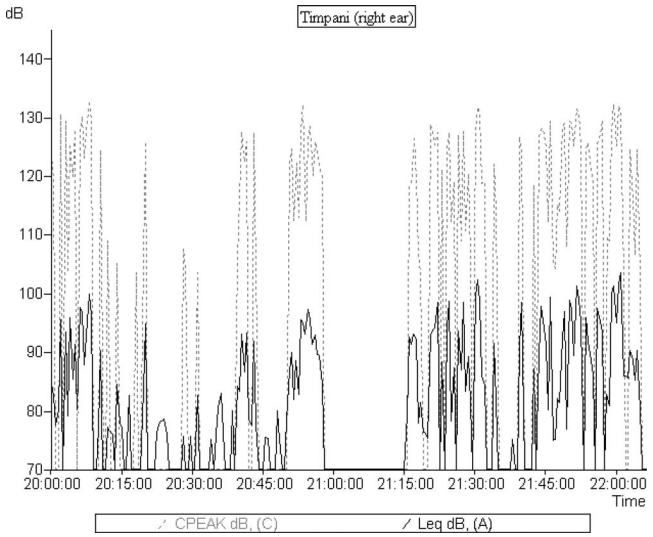


FIG. 7. Exposure to timpani during a concert at QPAC Concert Hall.

TABLE VI. Estimated risk level by exposure.

Risk level	Mean average (dBLEQ)	Average DND	
		3 h call	2.5 h call
High	>89	>100%	>80%
Medium high	88–89	80–100%	60–80%
Medium low	86–88	50–80%	40–60%
Low	<86	<50%	<40%

right ear, even with a very loud player to their left. This was clearly due to the proximity of their own instrument's bell and is yet another example of the degree to which noise exposure amongst orchestral musicians is dependent on their own instrument.

It is a logical assumption that the players directly in front of the trumpets are exposed to a much higher level than the trumpet players themselves; however, this was routinely not the case at The Queensland Orchestra. There are possibly two reasons for this finding. First, players in front of the trumpets were usually either on a lower rostrum and/or had wraparound head screens in place. In the orchestra pit an angled screen was often in place to reflect the trumpet's sound up and out. Second, the music stand directly in front of the trumpet—made from steel with the music itself resting on it—may be responsible for reflecting much of the trumpet players' sound back to them. Of the three trumpet positions, the first trumpet consistently recorded higher noise levels in all venues. Once again this may be attributed to the fact that the first trumpet plays in a higher register than the rest of the brass section and higher than both the second and third trumpets. The result leaves the first trumpet as the player with the highest expected dBALEQ of any musician in the orchestra.

Sustained levels in the trombones and tuba tend to be less than the trumpets, which may be as a result of the lower register in which they play. The first trombone is routinely exposed to higher noise levels than the second trombone, the third (bass) trombone and the tuba with the exception of the Lyric Theatre pit, where the second trombone generally recorded higher levels. This is probably due to the position of the players—with the second trombone literally surrounded by high-level instruments (trombones to either side, timpani directly behind and trumpets in front), whereas the first trombone has a clear area directly to his left.

The trumpets, trombones, and tuba were often exposed to high peak levels due to the proximity of the percussion, but never exceeded the actionable level of 140 dBC. In a symphonic program, peaks in the brass typically registered between 130 and 133 dBC. This is still exceptionally loud and would contribute significantly to a player's discomfort.

D. Percussion/timpani

Percussion and timpani orchestral noise was characterized by extremely high-level transient peaks and medium to high LEQs (see Fig. 6). Comparing the dBC peak readings to the dBA peak readings of the same event tells us a great deal about the nature of these percussive spikes routinely seen in the percussion and timpani. It is common for there to be a very large gap between these two figures. For instance, a percussion peak during Shostakovich's *Symphony No. 11* registered at 143.8 dBC, yet the dBA peak reading for this event was 113.1 dBA. This peak was caused by a snare drum in conjunction with a bass drum and was confirmed by analyzing the time of the peak and determining whether a lesser event registered on both of the other units at an identical time. The dBC peak/dBA peak difference implies a great deal of acoustic energy at high and/or low frequencies as opposed to the part of the frequency spectrum where the dBA filter is most sensitive. This occurrence was typical in percussion and timpani readings and also applied to those in close proximity to these instruments, particularly the bass section when in the orchestra pit.

VIII. CONCLUSION

This study has demonstrated that three key variables—position, venue and repertoire—impact significantly upon the noise exposure of individual musicians. It is clear from the findings that musicians in close proximity do not necessarily share the same risk profile and should approach hearing conservation in subtly different ways. It is also clear from the findings that repertoire is a key determinate in noise exposure and that musicians, in particular, positions, can expect their noise exposure to vary dependent upon the venue in which they are performing or rehearsing. The interaction between the three variables—such as specific repertoire only occurring in specific venues—gives important pointers in predicting any given musicians' exposure for an orchestral call, and will clearly differ from orchestra to orchestra.

Overall, the findings indicate the principal trumpet, first and third horn and principal trombone are at greatest risk of exposure to excessive sustained noise levels, and that the percussion and timpani are at greatest risk of excessive peak noise levels. However, the findings also strongly support the notion that the true nature of orchestral noise is a great deal more complex than this.

Although this study is by no means exhaustive and data gathering is ongoing, no study to date has undertaken such a thorough noise survey of any orchestra, nor has any study systematically attempted to account for the many variables inherent in the noise exposure of orchestral musicians. We believe the findings form a solid basis for future investigations into the conservation of musicians' hearing and into the improvement of working conditions for orchestral musicians generally.

ACKNOWLEDGMENTS

The authors wish to thank The Queensland Orchestra for its assistance and participation in this study. In particular, the hard work of Judy Wood, Sarah Evans, Darryl Keys, and the Operations Team, and the patience and goodwill of each of the musicians has been integral in making this study possible.

Australian Standards/New Zealand Standards (AS/NZS). (**2005a**). "Australian/New Zealand Standard: Occupational noise management," *AS/*

NZS:1269, Standards Australia, Sydney, Australia/Standards New Zealand, Wellington, New Zealand.
Australian Standards/New Zealand Standards (AS/NZS). (**2005b**). "Occupational noise management part 1: Measurement and assessment of noise immission and exposure," *AS/NZS:1269.1*, Standards Australia, Sydney, Australia/Standards New Zealand, Wellington, New Zealand, p. 37.
Axelsson, A., and Lindgren, F. (**1981**). "Hearing in classical musicians," Acta Oto-Laryngol., Suppl. **377**, 3–74.
Camp, J. E., and Horstman, S. W. (**1992**). "Musician sound exposure during performance of Wagner's ring cycle," Med. Prob. Perf. Art. **7**, 37–39.
Chasin, M. (**1996**). *Musicians and the Prevention of Hearing Loss* (San Diego, Singular Publishing Group, Inc., San Diego, CA).
Jansson, E., and Karlsson, K. (**1983**). "Sound levels recorded within the symphony orchestra and risk criteria for hearing loss," Scand. Audiol. **12**, 215–221.
Laitinen, H. M., Toppila, E. M., Olkinoura, P. S., and Kuisma, K. (**2003**). "Sound exposure among the Finnish national opera personnel," Appl. Occup. Environ. Hyg. **18**, 177–182.
Lee, J., Behar, A., Kunov, K., and Wong, W. (**2005**). "Musicians' noise exposure in the orchestra pit," Appl. Acoust. **66**, 919–931.
Mikl, K. (**1995**). "Orchestral music: An assessment of risk," Acoust. Aust. **23**, 51–55.
Royster, J. D., Royster, L. H., and Killion, M. C. (**1991**). "Sound exposure and hearing thresholds of symphony orchestra musicians," J. Acoust. Soc. Am. **89**, 2793–2803.
Sabesky, I. J., and Korczynski, R. E. (**1995**). "Noise exposure of symphony orchestra musicians," Appl. Occup. Environ. Hyg. **10**, 131–135.
Sataloff, R. T., and Sataloff, J. (**2006**). *Occupational Hearing Loss* (CRC Press: Taylor and Francis Group, Boca Raton, FL).
Schacke, G. (**1987**). "Sound pressure levels within an opera orchestra and its meaning for hearing," Abstract of paper delivered to the 22nd International Congress on Occupational Health, 7th Sept. – 2 Oct., 1987, Sydney, Australia.
van Hees, O. S. (**1991**). *Gehoorafwijkingen bij Musici* (Coronel Laboratorium, Universitiet van Amsterdam, Amsterdam) p. 257.
Westmore, G. A., and Eversden, I. D. (**1981**). "Noise induced hearing loss and orchestral musicians," Arch. Otolaryngol. **107**, 761–764.
Woolford, D. H. (**1984**). "Sound pressure levels in symphony orchestras," Audio Engineering Society 1984 Australian Regional Convention, Melbourne, Australia.
Woolford, D. H., Carterette, E. C., and Morgan, D. E. (**1988**). "Hearing impairment among orchestral musicians," Music Percept. **5**, 261–284.

Bottom-up approach for microstructure optimization of sound absorbing materials

Camille Perrot,[a] Fabien Chevillotte, and Raymond Panneton
Groupe d'Acoustique de l'Université de Sherbrooke (GAUS), Department of Mechanical Engineering, Université de Sherbrooke, Quebec J1K 2R1, Canada

(Received 8 January 2008; revised 8 May 2008; accepted 15 May 2008)

Results from a numerical study examining micro-/macrorelations linking local geometry parameters to sound absorption properties are presented. For a hexagonal structure of solid fibers, the porosity ϕ, the thermal characteristic length Λ', the static viscous permeability k_0, the tortuosity α_∞, the viscous characteristic length Λ, and the sound absorption coefficient are computed. Numerical solutions of the steady Stokes and electrical equations are employed to provide k_0, α_∞, and Λ. Hybrid estimates based on direct numerical evaluation of ϕ, Λ', k_0, α_∞, Λ, and the analytical model derived by Johnson, Allard, and Champoux are used to relate varying (i) throat size, (ii) pore size, and (iii) fibers' cross-section shapes to the sound absorption spectrum. The result of this paper tends to demonstrate the important effect of throat size in the sound absorption level, cell size in the sound absorption frequency selectivity, and fibers' cross-section shape in the porous material weight reduction. In a hexagonal porous structure with solid fibers, the sound absorption level will tend to be maximized with a 48 ± 10 μm throat size corresponding to an intermediate resistivity, a 13 ± 8 μm fiber radius associated with relatively small interfiber distances, and convex triangular cross-section shape fibers allowing weight reduction.
 [DOI: 10.1121/1.2945115]

PACS number(s): 43.50.Gf, 43.20.Hq, 43.20.Wd, 43.20.El [KA] Pages: 940–948

I. INTRODUCTION

A major issue in automobile, aeronautical, and building industries concerns the need to increase or adapt the sound absorption spectrum of commonly used sound absorbing materials. However, the most advanced models used to characterize and predict sound absorbing material performances are mainly based on interdependent macroscopic parameters, which do not take explicitly into account the local geometry of the porous media (i.e., its microstructure). For these reasons, optimizing sound absorbing materials from their fabrication remains a difficult task mostly done by trial and error. A strict optimization method would firstly rely on our ability to predict the acoustic properties of porous media from the description of their local geometry. Secondly, it would propose pertinent realistic modifications of their microstructure having predictable impacts on their absorption spectrum. The intent of this paper is to present such an optimization procedure following the bottom-up approach (i.e., from an optimized microstructure to the desired acoustical macroscopic behavior).

How do local geometry parameters relate to sound absorption spectrum in porous media? How do macroscopic flow and thermal properties depend on throat size, cell size, and fibers' cross-section shape? These are two of the many questions that dominate studies of relationships between microstructure and acoustic properties of porous media such as open-cell foams and fibrous materials. Such questions may be addressed in different manners. A common method consists in conducting a lot of laboratory measurements on samples of varying microstructural parameters.[1,2] Alternatively, in a search for a theoretical understanding, one may try to better understand the mathematical and physical basis of the macroscopic equations governing acoustic dissipation phenomena.[3–8] Also, numerical studies based on simulations can be considered.[9–16] Finally, recent studies include hybrid[17] approaches combining numerical predictions of key physical parameters used as input data in empirical models.[18] Each of these ways of considering these questions has advantages and disadvantages. Laboratory measurements are of indisputable value; however, their interpretation may be limited to a specific group of materials. Theoretical studies at the macroscopic scale lead to robust models, but they also require measurements of nonindependent macroscopic parameters. Numerical simulations usually attempt to bridge the gap between theory and experiments. They are nevertheless typically restrained by either the need to simplify geometry, physics, or both. Finally, hybrid approaches suffer from the weakness of empirical models providing poor physical insight and being unable to consider already nonexisting microstructural configurations.

In recent years, another approach to the numerical study of long-wavelength acoustic properties of porous media has gained some interests. The idea is to numerically solve, in a microstructural configuration that consists of a periodic unit cell (PUC), the linearized Navier–Stokes equation in harmonic regime with the local incompressibility condition[11] (dynamic viscous problem) and the linearized heat equation in harmonic regime[8] (dynamic thermal problem), with appropriate boundary conditions, and then to study how volume-averaged properties of the velocity and thermal fields relate

[a]Author to whom correspondence should be addressed. Electronic mail: camille.perrot@usherbrooke.ca

to microscopic details of the geometry. Compared to macroscopic models, such an approach offers the ability to study the microphysical basis of the acoustical macrobehavior.

For the case of the dynamic viscous problem, solutions mainly based on finite element methods (FEMs) have been investigated. Craggs and Hildebrandt[9] solved the viscous problem for specific cross sections of uniform pores. Wang and Lu[10] determined the optimized acoustic properties of polygonal ducts through semianalytical solutions. Zhou and Sheng[11] treated the case of a cylindrical tube with sinusoidal modulation of its cross section, and three-dimensional (3D) fused-spherical-bed and fused-diamond lattices. Firdaouss *et al.*[12] paid attention to a corrugated pore channel. Cortis *et al.*[13] studied the case of two-dimensional (2D) configurations made of a square arrangement of solid cylinders. They were also interested by the corrugated pore channel.[14] Gasser *et al.*[15] treated the 3D case of the face centered cubic sphere packing. An attempt to grasp the viscous dynamic behavior of more complex microstructures, such as a real open-cell aluminum foam sample, has also been carried out recently; thanks to a basic 2D model geometry with a relatively good success.[16]

Alternatively, the random-walker simulation method has been recently proposed by Lafarge[19] to provide an efficient resolution of the dynamic thermal problem. The principle of the method consists in simulating Brownian motion for a large number of the fluid-phase particles, and to link their mean square displacements to the thermal conduction properties of the confined fluid. An important point of the method is that, once the mean square displacements of a large number of particles have been estimated, the dynamic thermal response might be obtained for all frequencies. Contrary to finite element analysis, the solution has not been computed at each frequency. The random-walker simulation method has been implemented in two and three dimensions for computing the trapping constant of a 2D arrangement of overlapping fibers of circular cross sections,[20] and 3D digitalized geometries.[21] However, the trapping constant only provides the asymptotic low frequency behavior of the thermal problem. The first numerical simulations in harmonic regime have recently been proposed for the case of 2D regular and random arrangements of fibers with circular cross sections.[19] This work has been extended to 3D PUC, and applied to the determination of the dynamic thermal characteristics of an open-cell aluminum foam.[22,23]

Starting from these microphysical foundations, the aim of this paper is to illustrate the potential of such a bottom-up approach for microstructure optimization of highly porous open-cell foams and fibrous sound absorbing materials. In the framework of this paper, only the dynamic viscous boundary value problem is considered. The porous structure is a hexagonal lattice of solid fibers in air. For this simple geometry, it is shown how local geometry parameters are related to the sound absorption spectrum of the porous media through the main macroscopic parameters, giving a physical insight for why it was achieved. In particular, we will examine the influence of (i) the throat size, (ii) the cell size, and (iii) the cross-section shape of the fibers (i.e., circle, convex, straight, and concave triangles). An outline of this paper is as follows: In Sec. II, the model geometry and computational method behind the bottom-up approach are introduced. In Sec. III, results on a hexagonal structure of solid fibers are presented.

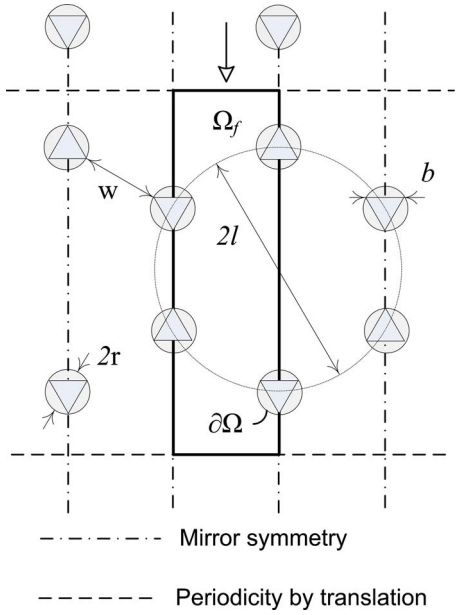


FIG. 1. (Color online) Local geometry model showing solid fibers having circular or triangular cross section shapes arranged in a hexagonal pattern. Triangular cross sections are considered to be inscribed in the corresponding circular ones. Numerical computations are performed at the scale of a vertical identified periodic unit cell (PUC).

II. NUMERICAL CALCULATIONS

A. Model geometry

A typical 2D hexagonal arrangement of fibers having l and r as local characteristic dimensions is depicted in Fig. 1. The fibers (the solid phase) are assumed motionless and diluted in air (the fluid phase). Their cross sections form the nodes of the hexagons. In this illustration, the cross sections of the fibers are circular. This is a typical case for a porosity $\phi \approx 0.85$ (or less). There are experimental evidences[24] that the cross-section shape of a foam ligament evolves from a circle ($\phi \approx 85\%$) for low porosity foams to convex ($\phi \approx 90\%$), straight ($\phi \approx 94\%$), and concave ($\phi \approx 98\%$) triangles for high porosity foams. For this reason, the shape of the cross section will be also considered as a local geometry parameter. See Fig. 2 for an illustration of the different cross-section shapes used in this study.

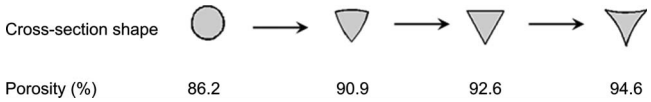


FIG. 2. Cross-section shapes and associated porosity values considered in this study. The porosity values are obtained with $r=32$ μm and $w=70$ μm. Triangular cross-section shapes (convex, straight, and concave) are inscribed in the initial circular cross section.

B. Computational method

The problem is addressed in three main steps: (1) solve numerically the asymptotic low (steady Stokes) and high (electric) frequency viscous boundary value problems using the FEM; (2) compute the static viscous permeability k_0, the viscous characteristic length Λ, and the tortuosity α_∞, as defined by Johnson *et al.*[4] by appropriate volume averaging of the corresponding asymptotic velocity fields; and (3) derive the frequency-dependent viscous and thermal response functions such as the effective density $\rho(\omega)$ and bulk modulus $K(\omega)$ of the entrained fluid in the rigid frame porous medium from the previously computed macroscopic parameters using the analytical models of Johnson *et al.*[4] and Allard and co-workers,[5,6] respectively, from which an approximate but robust description of propagation and absorption phenomena of an acoustic wave through a porous medium is entirely known.[6] Then, modification of the local geometry parameters by an iteration process allows us to maximize the area under the sound absorption coefficient curve in the frequency range of interest (here, from 10 to 10 000 Hz).

1. Asymptotic boundary value problems

At the zero angular frequency ($\omega=0$), the following set of equations is observed:

$$-\nabla\pi+\Delta\mathbf{w}+\mathbf{e}=0 \quad \text{in } \Omega_f, \tag{1}$$

$$\nabla\cdot\mathbf{w}=0 \quad \text{in } \Omega_f, \tag{2}$$

$$\mathbf{w}=0 \quad \text{on } \partial\Omega, \tag{3}$$

with the condition that π is a stationary field, simply describing the viscous fluid motion created in a porous medium in steady state regime. This is the steady Stokes problem in the fluid volume Ω_f for periodic structures (see Fig. 1), where $\mathbf{w}$ is the scaled static velocity field in the pore in m^2, and $\mathbf{e}$ is a unit vector. In what follows, the symbol $\langle\ \rangle$ designates a fluid-phase average. Writing the pressure p in terms of its mean and deviatoric parts, $p=\langle p\rangle+\Pi$ with $\langle\Pi\rangle=0$, the macroscopic pressure gradient is related to $\mathbf{e}$ in Eq. (1) by $\nabla\langle p\rangle=-|\nabla\langle p\rangle|\mathbf{e}$. The small fluctuation Π is related to π by $\Pi=|\nabla\langle p\rangle|\pi$. Finally, the static velocity field $\mathbf{v}$ is related to $\mathbf{w}$ by $\mathbf{v}=|\nabla\langle p\rangle/\eta|\mathbf{w}$, where η is the dynamic viscosity of the fluid.

At the opposite frequency range, when ω becomes very large, the viscous boundary layer becomes negligible and the fluid tends to behave as a perfect one, having no viscosity. In these conditions, the perfect incompressible fluid formally behaves according to the electric problem.[3] This electric problem is relevant to sound propagation as long as the wavelength is large enough for the saturating fluid to behave as an incompressible fluid in volumes of the order of the homogenization volume (a period in the case of periodic structure). $\mathbf{E}$ is the scaled electric field that solves the corresponding electrical conduction problem for a porous medium filled with a conducting fluid and having an insulating solid phase, i.e.,

$$\mathbf{E}=-\nabla\varphi+\mathbf{e} \quad \text{in } \Omega_f, \tag{4}$$

$$\nabla\cdot\mathbf{E}=0 \quad \text{in } \Omega_f, \tag{5}$$

$$\mathbf{E}\cdot\mathbf{n}=0 \quad \text{on } \partial\Omega, \tag{6}$$

and φ is a spatially stationary or periodic scalar field representing the deviatoric part of the electric potential.

The so-called steady Stokes and electric boundary value problems have been solved using a commercial finite element code[25] on the hexagonal 2D porous structure (cellular). No-slip boundary conditions at the pore walls, and periodicity of π were prescribed (then the static velocity field $\mathbf{w}$ is automatically periodic). For the electric problem, Neumann boundary conditions on the fluid-solid interface, and periodicity of φ on the inlet-outlet surfaces were used. Generalized Neumann boundary conditions are set in the remaining borders due to the symmetries of the problems. Calculations were performed with varying grid size to check the convergence of the results. Up to a total of 15 elements were used in the thickness of the viscous boundary layer. The symmetry property[26] of the viscous permeability tensor was also checked to guarantee the variation of the asymptotic solutions to less than a percent.

2. Macroscopic parameters' evaluation

Macroscopic parameters are then derived from the flow field solutions by spatial averaging. The open porosity ϕ was computed from the volume of the mesh, and the thermal characteristic length Λ' defined as the fluid-phase volume to wet surface ratio was obtained by the volume to wet surface ratio of the mesh. The static viscous permeability k_0 is computed from the static velocity field, and the viscous characteristic length Λ and the tortuosity α_∞ are computed from the electric field as defined by Johnson *et al.*[4] The components k_{0ij} defining the static viscous permeability tensor are simply given by[16]

$$k_{0ij}=\phi\langle\mathbf{w}^j\cdot\mathbf{e}^i\rangle, \tag{7}$$

where the superscript j refers to the direction of the imposed pressure gradient, and the components of $\mathbf{e}^i$ are $e^i_j=\delta_{ij}$ (with $\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ if $i\neq j$). In the studied configuration shown in Fig. 1, the gradient is along the vertical axis.

The components $\alpha_{\infty ij}$ defining the tortuosity tensor are derived from[27]

$$\alpha^{-1}_{\infty ij}=\langle\mathbf{E}^j\cdot\mathbf{e}^i\rangle, \tag{8}$$

where $\alpha^{-1}_{\infty ij}$ is the inverse of the tortuosity tensor $\alpha_{\infty ij}$. The viscous characteristic length Λ is computed using the definition of Johnson *et al.*[4] as follows:

$$2/\Lambda=\int_{\partial\Omega}\mathbf{E}^2dS/\int_{\Omega}\mathbf{E}^2dV, \tag{9}$$

who introduced this length-scale parameter Λ as the weighted pore volume to wet surface ratio. In the case of microscopic anisotropy for periodic porous structures such as the one studied here, static viscous permeability k_{0ij} and tortuosity $\alpha_{\infty ij}$ tensors reduce to scalars k_0 and α_∞, respectively.[26] This property makes implicitly reference to a

generalization in harmonic regime of the proof given by Torquato for the symmetry property in static regime in Ref. 26.

3. Analytical models

Equations derived by Johnson *et al.*[4] and Allard and co-workers[5,6] are then used to relate the macroscopic parameters to the effective density and bulk modulus of a fluid filled porous medium. The frequency-dependent absorption coefficient is then expressed from these quantities.[6] Note that the way the problem is addressed can eventually be refined by considering the computation of additional macroscopic parameters such as the static viscous tortuosity α_0, the static thermal permeability k_0', and the static thermal tortuosity α_0' defined by Lafarge *et al.*[8,27,28] as successive improvements of the modeled frequency-dependent viscous and thermal response functions.

Effective density and bulk modulus functions can be conveniently represented by the following approximate models:

$$\rho(\omega) = \rho_0 \alpha_\infty \left[1 + \frac{1}{i\varpi} f(\varpi) \right], \tag{10}$$

$$\frac{1}{K(\omega)} = \frac{1}{K_a} \left\{ \gamma - (\gamma - 1) \left[1 + \frac{1}{i\varpi'} f'(\varpi') \right]^{-1} \right\}, \tag{11}$$

where $K_a = \gamma P_0$ is the adiabatic bulk modulus, γ is the specific heat ratio, and P_0 the atmospheric pressure. ϖ and ϖ' are dimensionless viscous and thermal angular frequencies given by the following expressions:

$$\varpi = \frac{\omega}{\nu} \frac{k_0 \alpha_\infty}{\phi} \tag{12}$$

and

$$\varpi' = \frac{\omega}{\nu'} \frac{k_0'}{\phi}, \tag{13}$$

with $\nu = \eta/\rho_0$, $\nu' = \nu/\mathrm{Pr}$, Pr the Prandtl number, and the following shape functions f and f':

$$f(x) = 1 - P + P\sqrt{1 + \frac{M}{2P^2} ix}, \tag{14}$$

$$f'(x) = 1 - P' + P'\sqrt{1 + \frac{M'}{2P'^2} ix}, \tag{15}$$

where the dimensionless shape factors have been introduced:

$$M = \frac{8 k_0 \alpha_\infty}{\Lambda^2 \phi}, \tag{16}$$

$$M' = \frac{8 k_0'}{\Lambda'^2 \phi}, \tag{17}$$

$$P = \frac{M}{4\left(\frac{\alpha_0}{\alpha_\infty} - 1 \right)}, \tag{18}$$

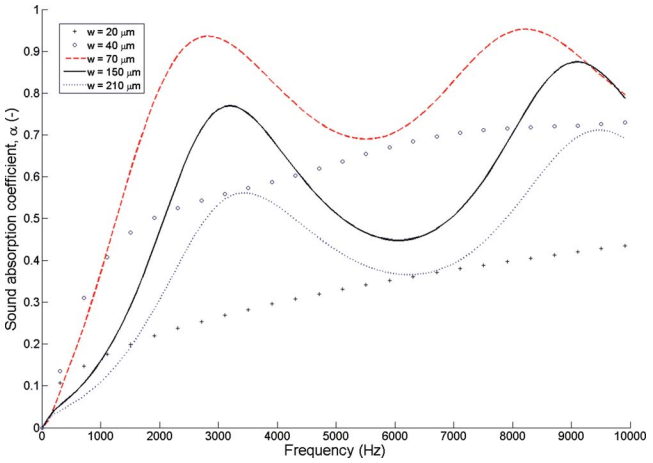


FIG. 3. (Color online) Effect of throat size w in the range 20–210 μm on the sound absorption coefficient. r is fixed at 200 μm.

$$P' = \frac{M'}{4(\alpha_0' - 1)}. \tag{19}$$

It can be seen that for $P = P' = M' = 1$, the model reduces to the so-called Johnson–Allard–Champoux model.

Although additional macroscopic parameters, as well as local geometry parameters can be considered, the focus of this analysis is practical trend exploration of sound absorption due to porous media morphology; how throat size, cell size, and cross-section shape of the ligaments influence sound absorption coefficient through the main macroscopic parameters in an idealized highly porous open-cell foam or fibrous absorber.

III. RESULTS AND DISCUSSION

A number of local geometry calculations are presented to illustrate the use of the computational method in maximizing the absorption coefficient of a hard backed 2D hexagonal porous structure having a total thickness of 25 mm for varying (i) throat sizes, (ii) cell sizes, and (iii) fiber cross-section shapes. To do so, the following sound absorption performance index will be used:

$$I = \frac{1}{N} \sum_{i=1}^{N} \alpha(\omega_i), \tag{20}$$

where $\alpha(\omega_i)$ is the normal incidence sound absorption computed at the ith angular frequency, and N is the number of discrete angular frequencies used in the frequency range of interest—here 10–10 000 Hz. For a given configuration, this indicator has to be maximized to yield the optimized local parameters.

A. Effect of throat size

Consider a plane acoustic wave propagating through the porous structure depicted in Fig. 1. The initial radius value of the circular cross-section fiber was fixed at $r \approx 200$ μm. Then, by an iteration process, the optimal throat size $w_{\mathrm{opt}} = 70$ μm was clearly determined, as shown in the plot of Fig. 3. For the initial r value, l was varied incrementally with a 10 μm step in the range 420–610 μm to find the optimum

TABLE I. Numerical evaluation of the main macroscopic parameters associated with a hexagonal porous structure of solid fibers with circular cross-section shapes for varying throat size values at constant fiber radius $r=200$ μm.

Throat size w (μm)	Porosity ϕ (%)	Thermal characteristic length Λ' (μm)	Viscous characteristic length Λ (μm)	Static airflow resistivity σ (N m^{-4} s)	Tortuosity α_∞ (−)	Performance index I (%)
20	45.16	165	37	1 108 618	2.05	30.47
30	47.68	182	54	408 035	1.77	46.12
40	50.03	200	71	201 463	1.61	58.17
50	52.23	219	87	116 813	1.50	66.46
60	54.28	237	103	74 965	1.43	71.50
70	56.21	257	119	51 591	1.37	73.68
80	58.01	276	135	37 365	1.33	73.58
90	59.71	296	150	28 136	1.30	71.94
100	61.31	317	166	21 845	1.27	69.40
110	62.81	338	181	17 386	1.25	66.41
120	64.22	359	196	14 123	1.23	63.29
130	65.56	381	211	11 670	1.22	60.21
140	66.83	403	226	9 784	1.20	57.26
150	68.02	425	241	8 306	1.19	54.49
160	69.15	448	256	7 129	1.18	51.91
170	70.23	472	271	6 177	1.17	49.53
180	71.24	496	286	5 398	1.16	47.32
190	72.21	520	301	4 752	1.15	45.29
200	73.13	544	316	4 213	1.14	43.40
210	74.00	569	332	3 757	1.14	41.65

$l=470$ μm, which yielded the maximum area under the absorption curve over the maximum possible area under the absorption curve, that is, $I_{opt}=73.67\%$. Note that at $l=480$ μm, similar global performances are obtained with $I=73.58\%$. The corresponding absorption spectrum is characterized by a higher absorption peak. More generally, for $460\ \mu\text{m}\leq l\leq 490\ \mu\text{m}$, that is, $60\ \mu\text{m}\leq w\leq 90\ \mu\text{m}$, the global absorption performances are systematically higher than $95\%\times I_{opt}\approx 70\%$, corresponding to an intermediate static airflow resistivity (defined as $\sigma=\eta/k_0$) range between 28 000 and 75 000 N m^{-4} s, see Table I.

B. Effect of fiber radius

Next, by setting $w=w_{opt}$, r was varied to find the optimum r_{opt}. r was initially varied with ten linearly spaced values in the range 50–350 μm. As index I decreases, a new iterative process was performed with ten linearly spaced values in the lower range 5–45 μm. The optimal fiber radius $r_{opt}=32$ μm was then determined with $I_{opt}=82.84\%$ and a static airflow resistivity equal to 26 758 N m^{-4} s, see Table II. Note that for a relatively large range of fiber radius, $10\ \mu\text{m}\leq r\leq 83\ \mu\text{m}$, $I\geq 95\%\times I_{opt}$, see Fig. 4. More generally, compared to the previous case, significant sound absorption enhancement can be obtained by keeping the optimal throat size and reducing the fiber radius, thus reducing in the mean time the cell sizes, which appear to be an argument in favor of small cell sizes ($2l=268$ μm) rather than the large ones ($2l=940$ μm). This phenomenon can be interpreted in terms of absorption frequency selectivity, where, given an optimal throat size and around the optimum fiber radius, a design choice could be made in terms of a rather large band or a relatively low frequency range absorption. Also interesting is the fact that the optimal fiber radius at constant throat size tends to minimize both viscous and thermal characteristic lengths, see Table II. This can be seen as a new criterion for pore (cell) size optimization at constant window (throat size) dimension.

C. Effect of cross-section shape

For the same geometry as considered in Fig. 1, the effect of the ligament cross-section shape on the sound absorption coefficient was also examined. Holding *a priori* w_{opt} and r_{opt} fixed, the cross-section shape c has to be searched to find the optimum c_{opt}. As previously mentioned in Sec. II A, the evolution of the fiber cross-section shape from a circle to a concave triangle is experimentally associated with an increasing porosity—see Fig. 2. With the aim to reflect this tendency, triangular cross sections were inscribed into the circular cross section, while keeping Λ' constant. However, this last condition, which is important to isolate the cross-section shape effect, also implies that the initial length between two ligaments associated with w_{opt} has to be modified. For the case of a straight triangular cross section, l can be expressed analytically as a function of Λ' in the following form: $l=\sqrt{(6b\Lambda'+\sqrt{3}b^2)/3\sqrt{3}}$, where $b=r_{opt}\sqrt{3}$. The next step is then to express l as a function of Λ' when considering the remaining cross-section shapes, convex and concave triangles. One more local geometry variable has to be taken into account, the convexity/concavity radius of curvature R, but it seems reasonable to state that $R\sim l$ from geometric considerations, see Fig. 1. Starting from macro-/microanalytical relationships linking Λ' to l, i.e.,

TABLE II. Numerical evaluation of the main macroscopic parameters associated with a hexagonal porous structure of solid fibers with circular cross-section shapes for varying fiber radius at constant throat size $w = 70\ \mu$m.

Fiber radius r (μm)	Porosity ϕ (%)	Thermal characteristic length Λ' (μm)	Viscous characteristic length Λ (μm)	Static airflow resistivity σ (N m^{-4} s)	Tortuosity σ_∞ (−)	Performance index I (%)
5	99.06	524	265	16 565	1.00	75.03
9	97.27	336	173	19 562	1.01	79.13
14	95.12	271	142	21 557	1.02	80.98
18	92.86	238	127	23 119	1.03	81.96
23	90.60	220	120	24 453	1.04	82.48
27	88.43	208	115	25 651	1.05	82.74
32	86.36	200	113	26 758	1.06	82.84
36	84.41	196	111	27 796	1.07	82.82
41	82.58	192	110	28 780	1.08	82.73
45	80.87	190	110	29 718	1.09	82.58
50	79.08	189	110	30 725	1.10	82.37
83	70.02	195	112	36 588	1.17	80.42
117	64.22	209	115	41 504	1.23	78.30
150	60.25	227	117	45 832	1.29	76.32
183	57.37	247	118	49 751	1.35	74.52
217	55.19	267	120	53 364	1.40	72.88
250	53.48	287	121	56 734	1.46	71.38
283	52.10	308	122	59 908	1.51	70.01
317	50.98	329	123	62 917	1.56	68.76
350	50.03	350	124	65 784	1.61	67.60

$$\Lambda' = \frac{l\sqrt{3}}{4\arcsin(b/2l)}\phi, \tag{21}$$

with

$$\phi = 1 - \left\{ \frac{b^2\sqrt{3}}{2} \pm 6\left[l^2 \arcsin\left(\frac{b}{2l}\right) - \frac{bl}{2}\sqrt{1-\frac{b^2}{4r^2}}\right]\right\} \Bigg/ \frac{3\sqrt{3}}{2}l^2, \tag{22}$$

a numerical inversion can be obtained after replacing $\arcsin(b/2l)$ and $\sqrt{1-b^2/4r^2}$ with their respective truncated Taylor expansions at order 3, $b/2l + b^3/48l^3$ and $1 - b^2/8l$. It finally yields $w_{opt} \approx 53\ \mu$m whatever the triangular cross-section shape.

As shown in Fig. 5 and Table III, the sound absorption performance index of a stack of fibers or open-cell foams with ligaments of convex, straight, and concave triangular cross-section shapes are slightly enhanced compared to those with a circular cross-section shape. More importantly, the porosity is actually significantly increasing with concavity, which is a compatible feature with foams' fabrication process.[24] Furthermore, the sound absorption enhancement with cross-section shape concavity is still associated with Λ minimization as it can be observed in Table III, which con-

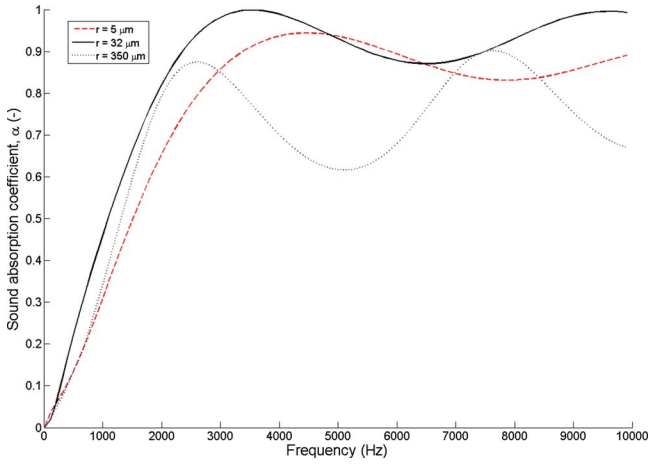


FIG. 4. (Color online) Effect of the fiber radius r in the range 5–350 μm on the sound absorption coefficient. w is fixed at the optimal throat size obtained for $r = 200\ \mu$m, that is, $w = 70\ \mu$m.

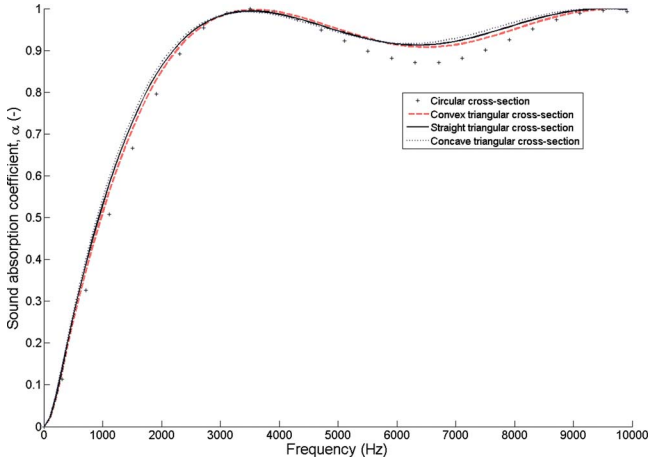


FIG. 5. (Color online) Cross-section shape modification and its effect on the sound absorption coefficient. Triangular cross-section shapes are inscribed in an initial fiber of circular cross-section shape with $r = 32\ \mu$m and $w = 70\ \mu$m.

TABLE III. Numerical evaluation of the main macroscopic parameters associated with a hexagonal porous structure of solid fibers at constant fiber radius $r=32$ μm and thermal characteristic length Λ' for varying inscribed cross-section shapes, i.e., (1) circle, (2) convex, (3) straight, and (4) concave triangle.

Cross-section shape c	Porosity ϕ (%)	Thermal characteristic length length Λ' (μm)	Viscous characteristic length Λ (μm)	Static airflow resistivity σ (N m^{-4} s)	Tortuosity σ_∞ (−)	Performance index I (%)
1	86.21	200	113	26 837	1.07	82.84
2	90.85	201	97	32 661	1.06	84.91
3	92.59	200	82	34 307	1.06	85.39
4	94.56	199	72	35 104	1.06	85.76

firms the potential of Λ minimization as a new sound absorption optimization criterion at constant throat size.

D. Multivariable analysis

The previous analyses focused on specific local geometry parameters and how they individually influence sound absorption. While these analyses provide physical insight, they are not completely rigorous in the sense of a multivariable problem. The relatively small number of calculations performed above does not guarantee that the global optimum has been found in the 3D space of the local parameters (w, r, and c). As commonly used in multivariable optimization problems, the variables identified as having a preeminent role (w and r) on the targeted property (I) can be varied simultaneously to develop a response surface. For example, an identified throat size optimum at constant fiber radius could be different for another different fiber radius. Such a multivariable approach is well appropriate to track the actual optimal (w,r) couples. It also enables extracting practical charts indicating a manufacturer with the aimed fiber radius considering a given controlled throat size, or reciprocally, in order to optimize the produced sound absorber. Furthermore, the global maximum $(w,r)_{\text{opt}}$ is undoubtedly identified, providing an absolute target, with the admissible intervals of variation to stay within 95% of I_{opt}.

A multivariable optimization was performed for different values of w and r in the range $w=[20,200]$ μm and $r=[1,200]$ μm, with c being assigned to circular and triangular cross-section shapes. The results are shown in Fig. 6. For both cross-section shapes, the resulting response surfaces shown by Fig. 6 (top) seem to demonstrate the existence of global maximums. For the case of the circular cross-section shape, $I_{\text{opt}}=85.47\%$ at $(w,r)_{\text{opt}}=(48,6)$ μm. For the case of the triangular cross-section shape, $I_{\text{opt}}=85.91\%$ at $(w,r)_{\text{opt}}=(48,13)$ μm. Furthermore, it appears clearly in Fig. 6 (bottom) that, contrary to the triangular cross-section case for which the optimal throat size is constant for varying radius (straight dotted line), the circular cross-section case shows a more complex relation between optimal throat size and radius (curved dotted line). This different behavior might be interpreted geometrically as follows. Given two triangular cross sections, a fiber radius increase does not significantly modify the solid surface seen by the acoustic wave in the vicinity of the throat. On the contrary, for circular cross sections, a fiber radius increase is associated with a non-neglectable solid surface increase, which is seen by the acoustic wave in the vicinity of the throat, except if the throat size is also increased. Finally, it is of practical interest to mention that, if the fabrication process is advanced enough to produce a fiber pattern keeping the optimal throat size, a relative large tolerance (5–150 μm) is acceptable in terms of fiber radius to stay with 95% of I_{opt}.

E. Comparison with experimental data

Finally, it would be interesting, however not necessary, to compare the results of this bottom-up approach to experimental data. The melamine foam is known to be one of the best acoustic materials in terms of sound absorption. It is often used to design anechoic chambers, and in aeronautic applications. It possesses a highly porous open-cell structure, with very elongated thin ligaments. In some extent, it could be seen as a fibrous structure such as the one described in Fig. 1. The cross-section shape of a melamine foam ligament is made of concave triangles. A scanning electron micrograph of a real melamine foam sample is presented in Fig. 7. A rough estimate of the local geometry parameters can be obtained from such micrographies, yielding $b \simeq 4.3 \pm 0.3$ μm and $l \simeq 46.5 \pm 31.3$ μm (the average length of a ligament on a micrograph is taken as the average length between two ligaments in the model geometry) and $r \simeq 2.5 \pm 0.2$ μm and $w \simeq 41.5 \pm 31.7$ μm. By reporting this couple of local geometry values in the chart of Fig. 6, one can see that this local geometry configuration lies in the best absorption region.

F. Limitations and future works

For practical trend exploration of sound absorption due to porous media morphology, simplifications were made and refined analytical models were not used. However, in the limiting case of fiber radius tending to 0, porosity tends to 1 (i.e., very small fibers concentration), which means that the dimensionless shape factors must tend to zero and that the validity of the results obtained with the Johnson–Allard–Champoux model become questionable. For this reason, a computational check of I was carried out using the refined analytical models in the optimal geometric configuration associated with fibers of circular cross-section shape. The differences for I between Johnson–Allard–Champoux and the refined models were found to be less than 3% (refined mod-

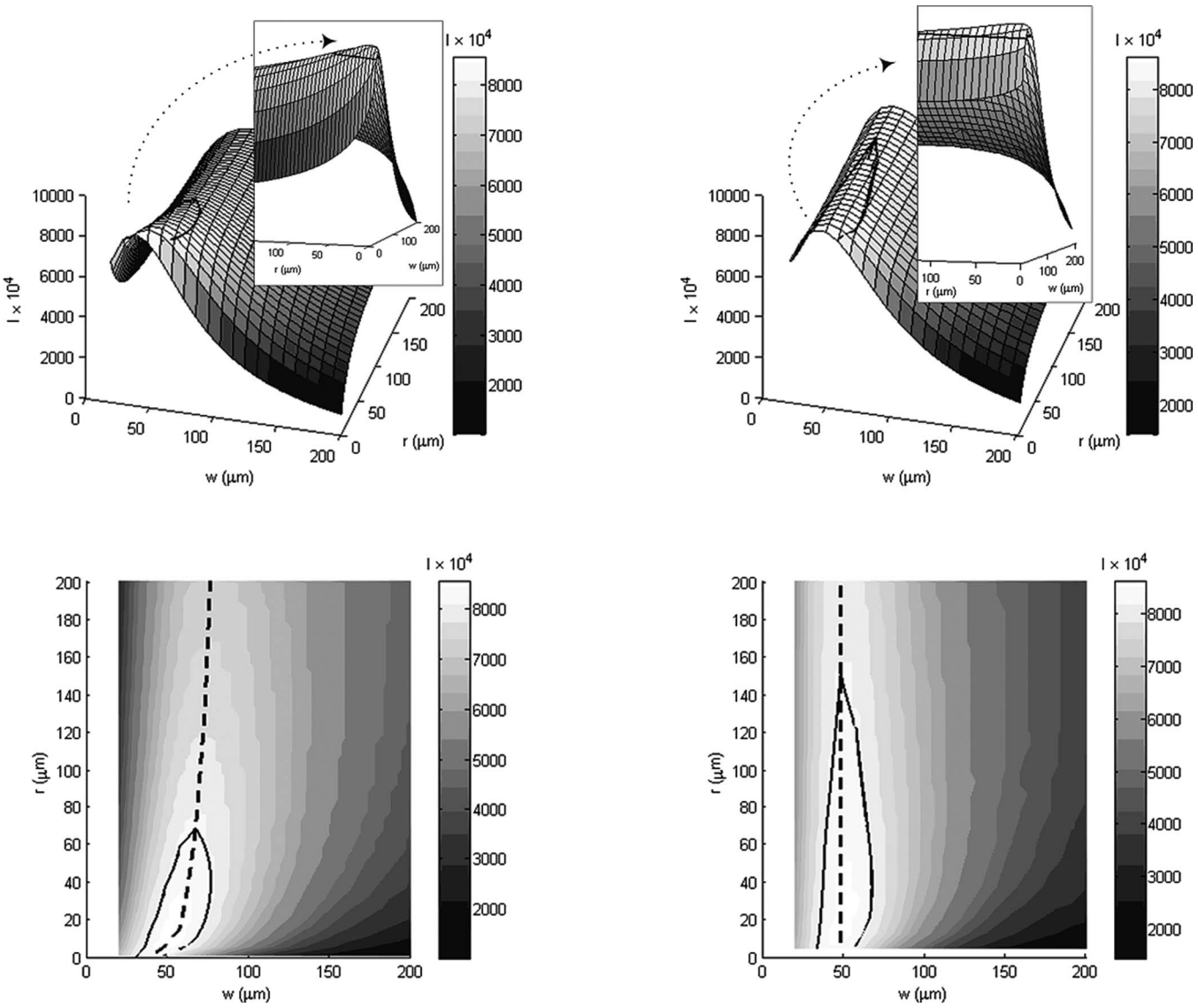


FIG. 6. Response surfaces (top) for the circular (left) and triangular (right) cross-section shape cases and associated 2D map (bottom). Straight lines delimit the $95\% \times I_{\text{opt}}$ zone, and the dotted lines are associated with the optimal (w, r) couples.

els predict better sound absorption performances—notably in the low frequency range). Future works should include (i) a weighting of the low frequency range in the computation of a sound absorption performance index, as well as (ii) a close examination of the acoustical macrobehavior in the limit of small radius using refined models. In particular, the notion of an optimum radius appearing while using the Johnson–Allard–Champoux model needs to be checked. This would also require significant modifications of the fluid-solid boundary conditions.

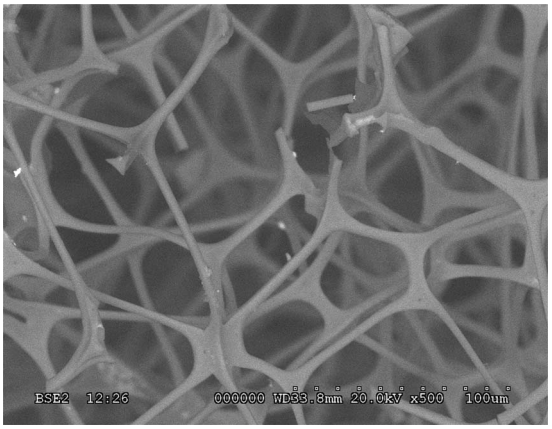


FIG. 7. Scanning electron micrograph of a real efficient sound absorbing melamine foam sample.

IV. CONCLUDING REMARKS

The results of our bottom-up approach for microstructure optimization of sound absorbing materials are summarized in this section. For a given fiber radius, an optimal throat size controlling the sound absorption level can be found, corresponding to an intermediate static airflow resistivity. By contrast, given an optimal throat size, the fiber radius (i.e., cell size) essentially modulates the absorption curve. It is worth mentioning that the optimal absorption curve is the one which minimizes the viscous characteristic length at constant throat size. This property can be used as a

design guide for sound absorption optimization. Another observation is the fact that for a given thermal characteristic length, as the porosity increases from its circular cross-section value to its concave triangular cross section value, the throat size reduces with the viscous characteristic length, enhancing only slightly the sound absorption coefficient of the porous structure but increasing notably its porosity (i.e., reduction of bulk density or weight). Finally, practical investigation charts have been proposed to indicate local geometry parameters tending to maximize the sound absorption coefficient. The validity of these charts was corroborated by comparison with the measured local geometry parameters of a real efficient sound absorbing melamine foam. This confirms the potential of such a bottom-up approach for microstructure optimization of sound absorbing materials.

ACKNOWLEDGMENTS

This work was supported in part by Grants-in-Aid from Alcan, CQRDA, NSERC, and Auto21. A part of the research presented in this paper was also financed by the Fonds Québécois de la Recherche sur la Nature et les Technologies by the intermediary of the Aluminium Research Centre-REGAL. Irène Kelsey from the Material Characterization Center under control of the Materials and Intelligent Systems Institute is gratefully acknowledged for providing technical assistance with scanning electron microscopy.

[1]A. Cummings and S. P. Beadle, "Acoustic properties of reticulated plastic foams," J. Sound Vib. **175**, 115–133 (1994).
[2]R. T. Muehleisen, C. W. Beamer, and B. D. Tinianov, "Measurements and empirical model of the acoustic properties of reticulated vitreous carbon," J. Acoust. Soc. Am. **117**, 536–544 (2005).
[3]R. J. S. Brown, "Connection between formation factor for electrical-resistivity and fluid-solid coupling factor in Biot equations for acoustic waves in fluid-filled porous media," Geophysics **45**, 1269–1275 (1980).
[4]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[5]Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).
[6]J. F. Allard, "Modelling sound absorbing materials: Propagation of Sound in Porous Media," edited by Elsevier Applied Science (Elsevier Science, New York, 1993).
[7]S. R. Pride, F. D. Morgan, and A. F. Gangi, "Drag forces of porous media acoustics," Phys. Rev. B **47**, 4964–4975 (1993).
[8]D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[9]A. Craggs and J. G. Hildebrandt, "Effective densities and resistivities for acoustic propagation in narrow tubes," J. Sound Vib. **92**, 321–331 (1984).
[10]X. Wang and T. J. Lu, "Optimized acoustic properties of cellular solids," J. Acoust. Soc. Am. **106**, 756–765 (1999).
[11]M. Y. Zhou and P. Sheng, "First-principles calculations of dynamic permeability in porous media," Phys. Rev. B **39**, 12027–12039 (1989).
[12]M. Firdaouss, J.-L. Guermond, and D. Lafarge, "Some remarks on the acoustic parameters of sharp-edged porous media," Int. J. Eng. Sci. **36**, 1035–1046 (1998).
[13]A. Cortis, D. M. L. Smeulders, D. Lafarge, M. Firdaouss, and J.-L. Guermond, in *IUTAM Symposium on Theoretical and Numerical Methods in Continuum Mechanics of Porous Materials. Series: Solid Mechanics and Its Applications*, University of stuttgart, Germany, edited by W. Ehlers (Kluwer, Dordrecht, 1999), pp. 187–192.
[14]A. Cortis, D. M. J. Smeulders, J.-L. Guermond, and D. Lafarge, "Influence of pore roughness on high-frequency permeability," Phys. Fluids **15**, 1766–1775 (2003).
[15]S. Gasser, F. Paun, and Y. Brechet, "Absorptive properties of rigid porous media: Application to face centered cubic sphere packing," J. Acoust. Soc. Am. **117**, 2090–2099 (2005).
[16]C. Perrot, F. Chevillotte, and R. Panneton, "Dynamic viscous permeability of an open-cell aluminum foam: Computations vs experiments," J. Appl. Phys. **103**, 024909 (2008).
[17]K. Schladitz, S. Peters, D. Reinel-Bitzer, A. Wiegmann, and J. Ohser, "Design of acoustic trim based on geometric modeling and flow simulation for non-woven," Comput. Mater. Sci. **38**, 56–66 (2006).
[18]F. P. Mechel, "Ausweitung der absorberformel von Delany and Bazley zu tiefen frequenzen," Acustica **35**, 210–213 (1976).
[19]D. Lafarge, in *Poromechanics II: Proceedings of the Second Biot Conference on Poromechanics*, edited by J.-L. Auriault (Swets & Zeitlinger, Grenoble, 2002), pp. 703–708.
[20]S. Torquato, "Efficient simulation technique to compute properties of heterogeneous media," Appl. Phys. Lett. **55**, 1847–1849 (1989).
[21]D. A. Coker and S. Torquato, "Simulation of diffusion and trapping in digitized heterogeneous media," J. Appl. Phys. **77**, 955–964 (1994).
[22]C. Perrot, R. Panneton, and X. Olny, "Periodic unit cell reconstruction of porous media: Application to an open cell aluminum foam," J. Appl. Phys. **101**, 113538 (2007).
[23]C. Perrot, R. Panneton, and X. Olny, "Computation of the dynamic thermal dissipation properties of porous media by Brownian motion simulation: Application to an open-cell aluminum foam," J. Appl. Phys. **102**, 074917 (2007).
[24]A. Bhattacharya, V. V. Calmidi, and R. L. Mahajan, "Thermophysical properties of high porosity metal foams," Int. J. Heat Mass Transfer **45**, 1017 (2002).
[25]COMSOL 3.4, WTC-5 pl. Robert Schuman, Grenoble, France.
[26]S. Torquato, "Relationship between permeability and diffusion-controlled trapping constant of porous media," Phys. Rev. Lett. **64**, 2644 (1990), makes implicitly reference to a generalization in harmonic regime of the proof given by Torquato for the symmetry property in static regime.
[27]D. Lafarge, "Propagation du son dans les matériaux poreux à structure rigide saturés par un fluide viscothermique (Sound propagation in rigid porous media saturated by a viscothermal fluid)," Ph.D. thesis, Université du Maine, 1993.
[28]D. Lafarge, "Modèles linéaires de propagation," in *Milieux Poreux et Poreux Stratifiés (Linear models of propagation in Porous Media and Stratified Porous*, Matériaux et Acoustique (Materials and Acoustics), Vol. **1**, edited by M. Bruneau and C. Potel (Lavoisier, Paris, 2006), pp. 143–188.

Fast affine projections and the regularized modified filtered-error algorithm in multichannel active noise control

J. M. Wesselink[a)]
University of Twente, Faculty EEMCS, P.O. Box 217, 7500AE Enschede, The Netherlands

A. P. Berkhoff[b)]
TNO Science and Industry, MON-Acoustics, P.O. Box 155, 2600AD Delft, The Netherlands

(Received 17 September 2007; revised 9 May 2008; accepted 27 May 2008)

In this paper, real-time results are given for broadband multichannel active noise control using the regularized modified filtered-error algorithm. As compared to the standard filtered-error algorithm, the improved convergence rate and stability of the algorithm are obtained by using an inner–outer factorization of the transfer path between the actuators and the error sensors, combined with a delay compensation technique using double control filters and a regularization technique that preserves the factorization properties. The latter techniques allow the use of relatively simple and efficient adaptation schemes in which filtering of the reference signals is unnecessary. Results are given for a multichannel adaptive feedback implementation based on the internal model control principle. In feedforward systems based on this algorithm, colored reference signals may lead to reduced convergence rates. An adaptive extension based on the use of affine projections is presented, for which real-time results and simulations are given, showing the improved convergence rates of the regularized modified filtered-error algorithm for colored reference signals.
 [DOI: 10.1121/1.2945169]

PACS number(s): 43.50.Ki, 43.60.Mn, 43.40.Vn [KAC] Pages: 949–960

I. INTRODUCTION

Many algorithms used for broadband active noise control are based on the adaptive least-mean-square (LMS) algorithm.[1] The low complexity and the relatively good robustness properties are the major advantages of the LMS algorithm. However, the speed of convergence of the algorithm may be substantially less than desired. There are several possible reasons for a reduced convergence speed. The frequency dependence of the secondary path and the shape of the spectrum of the reference signal may lead to reduced convergence rates. These two influences have been compensated for in the preconditioned filtered-error algorithm.[2] Another algorithm that has similar performance figures is the hybrid filtered error LMS algorithm.[3] These algorithms have the additional advantage that they are more efficient than the filtered-reference algorithm for the case of multiple reference signals. In the hybrid filtered-error algorithm the error filter has been modified to reduce the influence of the delay in the adaptation loop, which in turn improves the convergence properties. In the preconditioned filtered-error LMS (PLMS) algorithm it was proposed to filter the reference signals[4] with the inverse of the minimum-phase process that colors the reference signals and to use a minimum-phase/all-pass decomposition for the secondary path. A method using a whitening of multiple reference signals and single-channel minimum-phase/all-pass decomposition was presented in Ref. 5. Another type of factorization based on singular value decomposition in the frequency domain suitable for multichannel systems was presented in Ref. 6. A disadvantage of the filtered-error algorithm is that the maximum convergence speed is reduced due to the inherent delay in the adaptation path. The negative influence of this delay can be eliminated by using an inner–outer factorization of the transfer path between the actuators and the error sensors, combined with a delay compensation technique using double control filters.[7] The latter algorithm can be combined with a regularization technique that preserves the factorization properties.[7] Also in the latter algorithm, the so-called regularized modified filtered-error (RMFe) algorithm, a requirement for good convergence is that the reference signals are white. In the affine projection (AP) algorithm and the fast affine projection (FAP) algorithms,[8–10] the decolorizing mechanism is included in the adaptation loop. The latter algorithms are able to decolorize autoregressive processes.[11] In Ref. 11, Rupp also described algorithms that are able to decolorize moving average processes. A multichannel modified filtered reference implementation using AP or FAP was introduced by Bouchard.[12] In the AP and the FAP algorithms it is necessary to calculate the inverse of an estimated autocorrelation matrix. In the algorithm introduced by Bouchard this inversion is performed by using an iterative Gauss–Seidel algorithm. In another publication it was proposed to use the conjugate gradient method to find the inverse.[13] Other methods were based on the assumption that the governing matrix is Toeplitz,[14] which holds when the adaptive filter is considerably longer than the affine projection order. In a paper by Heping an overview of methods for finding the inverse matrix is presented.[15] Heping analyzes the complexity for the

[a)]Electronic mail: j.m.wesselink@utwente.nl
[b)]Electronic mail: arthur.berkhoff@tno.nl

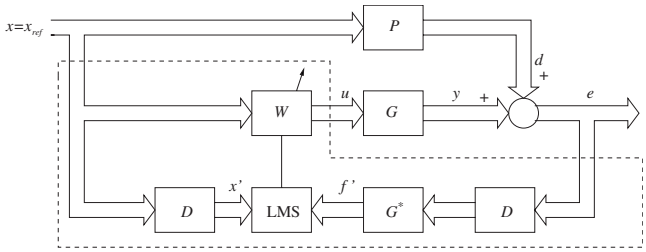


FIG. 1. Filtered-error adaptive control scheme.

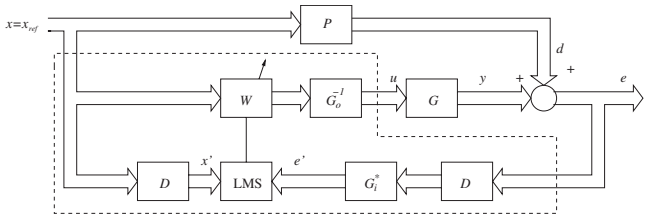


FIG. 2. Preconditioned filtered-error adaptive control scheme.

most widely used methods, and also gives a method for computing the inverse with an LDL^T decomposition. Fast approximate implementations of the FAP algorithm are given by Douglas.[16]

The major contributions of the present paper are as follows. First, results are given of a real-time implementation of the RMFe algorithm of Ref. 7. Real-time results will be given for feedback and feedforward implementations. The steady-state reductions of the algorithms obtained in real time are compared with the optimal, causal Wiener filter solutions. Second, methods are described to improve the convergence speed of the RMFeLMS algorithm for colored reference signals by an extension with a multichannel FAP algorithm, using the possibility for delay-less adaptation as provided by the RMFe algorithm. The paper is organized as follows. Section II will give an introduction to the RMFeLMS algorithm. Section III gives a description of the modifications that are needed to extend the RMFeLMS algorithm with an AP algorithm and its fast derivatives. Real-time experimental results and the optimal Wiener solutions will be presented in Sec. IV. The results will be presented for a feedback architecture as well as for a feedforward architecture. Section V will contain the conclusions.

II. THE REGULARIZED MODIFIED FILTERED-ERROR ALGORITHM

In this section the equations for the regularized modified filtered-error algorithm will be derived. The starting point is the filtered-error algorithm (see Fig. 1).

The controller can be found within the dashed line. The sample moment is denoted by n and the unit-delay forward-shift operator as q. The signal vectors are defined as $x(n)=[x_1(n)\cdots x_K(n)]^T$, $d(n)=[d_1(n)\cdots d_L(n)]^T$, $e(n)=[e_1\cdots e_L(n)]^T$, and $u(n)=[u_1\cdots u_M(n)]^T$, which are the reference signals, the disturbance signals, the error signals, and the actuator driving signals, respectively. The assumption is that noise which is uncorrelated with the reference signal is negligible. The disturbance signals $d(n)$ result from an assumed $L\times K$-dimensional primary path $P(q)$ driven by the reference signals. The goal of the algorithm is to minimize the signals $e(n)=d(n)+y(n)$, where $y(n)$ are the contributions from an $L\times M$-dimensional transfer path $G(q)$, the secondary path, which is driven by actuator driving signals $u(n)$. The actuator driving signals are obtained by feeding the reference signals through an $M\times K$-dimensional control filter $W(q)$. The controller $W(q)$ is assumed to be an $M\times K$-dimensional matrix with finite impulse response (FIR) filters. The ith filter coefficient of this control filter is denoted as the $M\times K$ matrix W_i, where $i=0,\ldots,I-1$, i.e., $W(q)=\Sigma_{i=0}^{I-1}q^{-i}W_i$. The adaptation in this algorithm is obtained with a LMS rule:

$$W_i(n+1)=W_i(n)-\alpha f'(n)x'^T(n-i), \tag{1}$$

where the superscript T denotes matrix transpose and where $x'(n)$ is a delayed version of the reference signal such that

$$x'(n)=D_K(q)x(n), \tag{2}$$

in which D_K is a $K\times K$-dimensional delay operator resulting in a delay of $J-1$ samples:

$$D_K(q)=q^{-J+1}I_K, \tag{3}$$

and in which $f'(n)$ is a filtered and delayed version of the error signal, such that

$$f'(n)=G^*(q)D_L(q)e(n). \tag{4}$$

In Eq. (4), the filtering is done with the adjoint $G^*(q)$, which is the transposed time reverse of the secondary path $G(q)$, i.e., $G^*(q)=G^T(q^{-1})$. The adjoint $G^*(q)$ is anticausal and has the dimensions $M\times L$. To make the adjoint predominantly causal it needs an additional delay operator D_L. The resulting matrix of transfer functions can be implemented as an $M\times L$ matrix of impulse responses with order $J-1$, resulting in J coefficients for each impulse response. The adaptation rate[4] is controlled by α.

The frequency dependence of the secondary path $G(q)$ and the interaction of the different channels will reduce the convergence rate. The convergence can be improved by inserting the inverse of the secondary path in between the control filter $W(q)$ and the secondary plant $G(q)$ (Ref. 4). To ensure stability, only the minimum-phase part of the plant is to be inverted. The secondary path can be written as

$$G(q)=G_i(q)G_o(q), \tag{5}$$

where

$$G^*(q)G(q)=G_o^*(q)G_o(q), \tag{6}$$

since

$$G_i^*(q)G_i(q)=I_M. \tag{7}$$

Assuming $L\geqslant M$, the transfer function $G_i(q)$ has dimensions $L\times M$ and the transfer function $G_o(q)$ has dimensions $M\times M$. The extraction of the minimum-phase part and the all-pass part is performed with a so-called inner–outer factorization.[17] The control scheme in which the inverse $G_o^{-1}(q)$ is used can be found in Fig. 2.

The update rule for this approach is

$$W_i(n+1) = W_i(n) - \alpha e'(n) x'^T(n-i). \tag{8}$$

In Fig. 2, the filtered error is denoted with $e'(n)$, emphasizing that the spectral characteristics of $e'(n)$ closely resemble that of the real error signal $e(n)$. It should be noted, however, that the dimensions are different since $e(n)$ is an $L\times 1$ -dimensional signal while $e'(n)$ is an $M\times 1$ -dimensional signal.

A shortcoming of the algorithm in Fig. 2 is that the convergence rate still suffers from the delay in the adaptation path. The path that reduces the convergence in Fig. 2 will now be analyzed. From this analysis, another algorithm will be derived that improves the behavior of the convergence. Let us start with the error signal $e'(n)$, which can be written as

$$e'(n) = G_i^*(q) D_L(q) [d(n) + G(q) G_o^{-1}(q) W(q) x(n)]. \tag{9}$$

Introducing the $M\times M$ matrix $D_M(q)$ having a delay that is identical to the $L\times L$ matrix $D_L(q)$, and using D_M instead of D_L, Eq. (9) can be rearranged to

$$e'(n) = G_i^*(q) D_L(q) d(n) + D_M(q) G_i^*(q) G(q) G_o^{-1} W(q) x(n). \tag{10}$$

Using Eqs. (5) and (7), $e'(n)$ can be expressed as

$$e'(n) = d'(n) + y'(n), \tag{11}$$

where an auxiliary disturbance signal $d'(n)$ is defined as

$$d'(n) = G_i^*(q) D_L(q) d(n), \tag{12}$$

and where, according to Eq. (7), the delayed preconditioned control output $y'(n)$ is defined by

$$y'(n) = D_M(q) W(q) x(n). \tag{13}$$

From Eq. (13) it can be seen that the transfer function between the output of $W(q)$ and $y'(n)$ is a simple delay $D_M(q)$. Let us now define an auxiliary output $y''(n)=y'(n)$ by

$$y''(n) = W(q) D_K(q) x(n), \tag{14}$$

where $D_K(q)$ is a $K\times K$ dimensional matrix having the same delay as $D_M(q)$. In Eq. (14) there is no delay anymore between the controller $W(q)$ and $y''(n)$. In order to realize this, the signal $e''(n)=e'(n)$ is introduced by noting that $y'(n)=y''(n)$:

$$e''(n) = d'(n) + y''(n). \tag{15}$$

Since $d'(n)$ is not directly available its needs to be reconstructed. This can be done using Eq. (11):

$$d'(n) = e'(n) - y'(n), \tag{16}$$

where, according to Eq. (13), $y'(n)$ can be obtained as the delayed version of the output of $W(q)$. Using $x'(n)=D_K(q)x(n)$, $y''(n)$ can be written as

$$y''(n) = W(q) x'(n). \tag{17}$$

The signal $x'(n)$ is already available as the input to the LMS block in the control schemes of Figs. 1 and 2. Equation (15) can now be written as

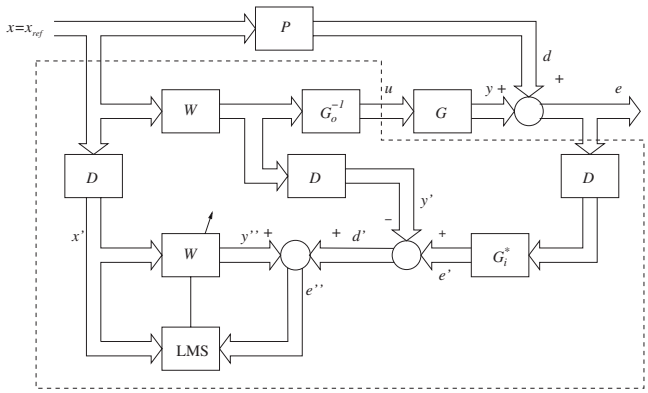


FIG. 3. Modified filtered-error adaptive control scheme.

$$e''(n) = d'(n) + W(q) x'(n). \tag{18}$$

The term $y''(n)$ can be obtained by adding a second set of control filters W operating on the delayed reference signals $x'(n)$. A block diagram based on Eq. (18) can be found in Fig. 3. The update rule for Fig. 3 is

$$W_i(n+1) = W_i(n) - \alpha e''(n) x'^T(n-i). \tag{19}$$

In the control scheme of Fig. 3 (and also Fig. 2), the output $u(n)$ can be very large if the plant G contains zeros or near zeros. In this case the inverse of the outer factor $G_o^{-1}(q)$ can have very high gains, which leads to saturation of the control signal $u(n)$. To prevent this some kind of regularization is required.[7] The required regularization must satisfy Eqs. (5)–(7).[7] This can be realized by using an augmented plant $\bar{G}(q)$:

$$\bar{G}(q) = \begin{bmatrix} G(q) \\ G_{\text{reg}}(q) \end{bmatrix}, \tag{20}$$

in which the regularization is performed by the $M\times M$-dimensional transfer function $G_{\text{reg}}(q)$. For simple weighting of $u(n)$ the regularization function can be set to

$$G_{\text{reg}}(q) = \sqrt{\beta} I_M. \tag{21}$$

The full $(L+M)\times M$-dimensional augmented plant is used in the control scheme, as well as the full $(L+M)\times M$-dimensional inner-factor $\bar{G}_i(q)$ and the $M\times M$-dimensional outer-factor inverse $\bar{G}_o^{-1}$. The modified inner-factor $\bar{G}_i$ is all-pass because $\bar{G}_i \bar{G}_o = \bar{G}$. A control scheme in which this regularization is incorporated can be found in Fig. 4. This method has the advantage that it is consistent with the modified filtered-error scheme, which is dependent on the all-pass property $\bar{G}_i^*(q)\bar{G}_i(q)=I_M$. The scheme in Fig. 4 is a generalized form meaning that different regularization schemes can be used for $G_{\text{reg}}(q)$.

III. AFFINE PROJECTION

The algorithm presented in Fig. 4 has good convergence properties when the reference signals are white. Such a situation also arises when the algorithm is used in a feedback architecture.[7] However, the performance degrades if the reference signals are strongly colored. In the original precondi-

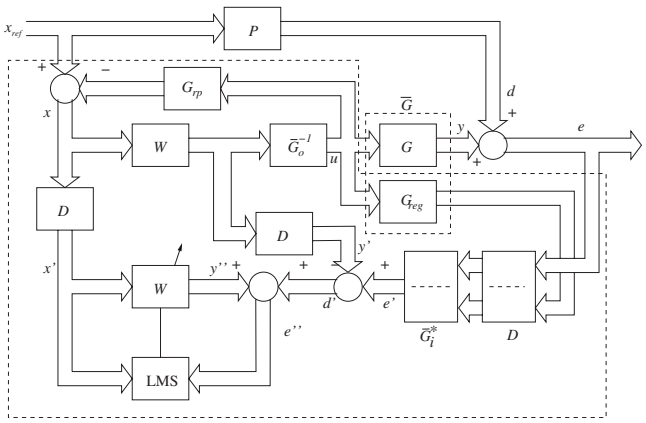


FIG. 4. Regularized modified filtered-error adaptive control scheme with IMC.

tioned filtered-error algorithm[4] the reference signals are prefiltered with the transfer function that colors the noise,[4] assuming *a priori* knowledge of the process that colorizes the noise. The affine projection algorithm is capable of estimating the autoregressive part of the process that colors the noise[11] while adapting the controller parameters. In this section the AP algorithm and the FAP algorithm will be derived, starting from the optimal Wiener filter. The FAP algorithm will be derived especially for the regularized modified filtered-error scheme. The derivation starts by rewriting Eq. (18) as

$$e''(n) = d'(n) + \mathbf{W}^T \underline{x}'(n), \tag{22}$$

in which the delay line $\underline{x}'$ is defined as

$$\underline{x}'(n) = [x'^T(n) \quad \cdots \quad x'^T(n-I+1)]^T, \tag{23}$$

and the filter coefficient matrix $\mathbf{W}$ is defined as

$$\mathbf{W} = [\underline{\mathbf{w}}_1 \quad \cdots \quad \underline{\mathbf{w}}_m \quad \cdots \quad \underline{\mathbf{w}}_M], \tag{24}$$

with

$$\underline{\mathbf{w}}_m = \begin{bmatrix} \mathbf{w}_{1m} \\ \vdots \\ \mathbf{w}_{im} \\ \vdots \\ \mathbf{w}_{Im} \end{bmatrix} \mathbf{w}_{im} = \begin{bmatrix} w_{1im} \\ \vdots \\ w_{kim} \\ \vdots \\ w_{Kim} \end{bmatrix}. \tag{25}$$

The MIMO-FAP algorithm will be derived starting at the MIMO-Wiener filter solution. Using Eq. (22) the cost criterium becomes

$$J = \mathrm{tr}\{E[e''(n)e''^T(n)]\}$$
$$= \mathrm{tr}\, E\{(d'(n) + \mathbf{W}^T\underline{x}'(n))(d'(n) + \mathbf{W}^T\underline{x}'(n))^T\}, \tag{26}$$

where tr denotes the trace operator and E denotes the expected value. Equation (26) can be written as

$$J = \mathrm{tr}\{\mathbf{R}_{d'd'}\} + \mathrm{tr}\{\mathbf{R}_{x'd'}^T\mathbf{W}\} + \mathrm{tr}\{\mathbf{W}^T\mathbf{R}_{x'd'}\}$$
$$+ \mathrm{tr}\{\mathbf{W}^T\mathbf{R}_{x'x'}\mathbf{W}\}, \tag{27}$$

with

$$\mathbf{R}_{d'd'} = E\{d'(n)d'^T(n)\},$$

$$\mathbf{R}_{x'd'} = E\{\underline{x}'(n)d'^T(n)\},$$

$$\mathbf{R}_{x'x'} = E\{\underline{x}'(n)\underline{x}'^T(n)\}. \tag{28}$$

The gradient of J with respect to the controller coefficients is

$$\frac{\partial J}{\partial \mathbf{W}} = 2(\mathbf{R}_{x'd'} + \mathbf{R}_{x'x'}\mathbf{W}). \tag{29}$$

The minimum for J is obtained by setting $\partial J/\partial \mathbf{W}$ to zero, resulting in the optimum MIMO-Wiener filter $\mathbf{W}_{\mathrm{opt}}$,

$$\mathbf{W} = \mathbf{W}_{\mathrm{opt}} = -\mathbf{R}_{x'x'}^{-1}\mathbf{R}_{x'd'}. \tag{30}$$

The derivation of the AP algorithm will start from a Newton iteration,[18] which is written as

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu\hat{\mathbf{R}}_{x'x'}^{-1}[\hat{\mathbf{R}}_{x'd'} + \hat{\mathbf{R}}_{x'x'}\mathbf{W}(n)], \tag{31}$$

where $\hat{\mathbf{R}}_{x'x'}$ and $\hat{\mathbf{R}}_{x'd'}$ are estimates of $\mathbf{R}_{x'x'}$ and $\mathbf{R}_{x'd'}$, respectively. Let K_A be the affine projection order, then the approximations for the correlation matrices are written as

$$\hat{\mathbf{R}}_{x'x'} = \frac{1}{K_A}\sum_{i=0}^{K_A-1}\underline{x}'(n-i)\underline{x}'^T(n-i) = \frac{1}{K_A}\mathbf{X}\mathbf{X}^T, \tag{32}$$

and

$$\hat{\mathbf{R}}_{x'd'} = \frac{1}{K_A}\sum_{i=0}^{K_A-1}\underline{x}'(n-i)d'^T(n-i) = \frac{1}{K_A}\mathbf{X}\mathbf{D}, \tag{33}$$

in which $\mathbf{X}(n)$ is an $IK \times K_A$-dimensional matrix that contains the last K_A vectors of $\underline{x}'(n)$, i.e.,

$$\mathbf{X}(n) = [\underline{x}'(n) \quad \cdots \quad \underline{x}'(n-K_A+1)]. \tag{34}$$

and in which $\mathbf{D}(n)$ is the $K_A \times M$-dimensional matrix containing the last K_A vectors of $d'(n)$, i.e.,

$$\mathbf{D}(n) = \begin{bmatrix} d'^T(n) \\ \vdots \\ d'^T(n-K_A+1) \end{bmatrix}. \tag{35}$$

For notational convenience the time index of $\mathbf{D}(n)$ and $\mathbf{X}(n)$ will be dropped. Substitution of Eqs. (32) and (33) into Eq. (31) results in

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu[\mathbf{X}\mathbf{X}^T]^{-1}[\mathbf{X}\mathbf{X}^T\mathbf{W}(n) + \mathbf{X}\mathbf{D}]$$
$$= \mathbf{W}(n) - \mu[\mathbf{X}\mathbf{X}^T]^{-1}\mathbf{X}[\mathbf{X}^T\mathbf{W}(n) + \mathbf{D}]. \tag{36}$$

In order to prevent problems with the inversion of $\mathbf{X}\mathbf{X}^T$, a small regularization term δ will be added to the diagonal elements. Then, the overall update rule becomes

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu[\mathbf{X}\mathbf{X}^T + \delta\mathbf{I}]^{-1}\mathbf{X}[\mathbf{X}^T\mathbf{W}(n) + \mathbf{D}], \tag{37}$$

which, using the matrix inversion lemma $(\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}]^{-1}\mathbf{D}\mathbf{A}^{-1}$, can be written as

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu\mathbf{X}[\delta\mathbf{I} + \mathbf{X}^T\mathbf{X}]^{-1}[\mathbf{X}^T\mathbf{W}(n) + \mathbf{D}]. \tag{38}$$

Due to the latter step the matrix inversion can be computed more efficiently. Equation (38) defines the AP algorithm.

The computational complexity of AP can be further reduced by the FAP algorithm.[9,10] Starting from the basic AP equation it will be shown how to expand the FAP algorithm to work in combination with the scheme presented in Fig. 4. Equation (38) will be written as

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \mathbf{X}(n)\boldsymbol{\xi}(n), \tag{39}$$

in which

$$\boldsymbol{\xi}(n) = [\delta \mathbf{I} + \mathbf{X}^T(n)\mathbf{X}(n)]^{-1}\mathbf{E}(n). \tag{40}$$

and in which a $K_A \times M$-dimensional matrix $\mathbf{E}(n)$ is defined that contains the last K_A error vectors:

$$\mathbf{E}(n) = \begin{bmatrix} e''^T(n) \\ \vdots \\ e''^T(n-K_A+1) \end{bmatrix}, \tag{41}$$

such that

$$\mathbf{E}(n) = \mathbf{D}(n) + \mathbf{X}^T(n)\mathbf{W}(n). \tag{42}$$

In the original FAP algorithm the three equations, Eqs. (42), (40), and (39), will be reduced in complexity. In the original FAP algorithm[10] it has been proposed to use a fast transversal filter to estimate the inverse autocorrelation matrix. This leads to issues with stability so it has been proposed[13] to calculate the inverse directly. In this paper only Eqs. (42) and (39) will be simplified and the matrix inversion in Eq. (40) will be approximated using a steepest-descent iteration. The error signal Eq. (42) will be calculated in such a way that the complexity will fall from $IK \times K_A$ to $IK \times 1$. The complexity of the update rule Eq. (39) will be reduced from $IK \times K_A \times L$ to $IK \times L$. It can be shown[10] that the error signal matrix can be written in the following form:

$$\mathbf{E}(n) = \begin{bmatrix} e''^T(n) \\ (1-\mu)\bar{\mathbf{E}}(n-1) \end{bmatrix}, \tag{43}$$

in which $\bar{\mathbf{E}}(n)$ are the $K_A - 1$ top rows of $\mathbf{E}(n-1)$.

FAP uses the assumption that the update rule can be rewritten in a form that uses alternative update coefficients. In this paper these alternative update coefficients will be derived. The reason for this is that the algorithm presented in Fig. 4 uses a copy of the filter coefficients in the upper branch. This makes it necessary to also find a new expression for these filter coefficients. The alternative filter coefficients will be written as $\hat{\mathbf{W}}$.

The alternative filter coefficients are defined as

$$\hat{\mathbf{W}}(n+1) = \hat{\mathbf{W}}(n) - \mu \underline{x}'(n-(K_A-1))E_{K_A-1}(n), \tag{44}$$

with $\underline{E}_{K_A-1}(n)$ defined as the bottom row of $\underline{\mathbf{E}}$, which in turn is defined as

$$\underline{\mathbf{E}}(n) = \boldsymbol{\xi}(n) + \begin{bmatrix} \underline{0}_M^T \\ \bar{\underline{\mathbf{E}}}(n-1) \end{bmatrix}. \tag{45}$$

This results in the following update rule for $W(n)$,

$$\mathbf{W}(n+1) = \hat{\mathbf{W}}(n+1) - \mu \bar{\mathbf{X}}(n)\bar{\underline{\mathbf{E}}}(n), \tag{46}$$

in which $\bar{\mathbf{X}}(n)$ is defined as the $K_A - 1$ left-most columns of $\mathbf{X}(n)$. The derivation of Eqs. (45) and (46) can be found in the Appendix.

Let us study the relation between $\mathbf{E}(n)$ and $\mathbf{E}(n-1)$. First an expression for the first row of Eq. (43) will be derived. Since $\mathbf{W}(n)$ is not directly available, we use Eq. (46) and substitute it in the transpose of Eq. (22):

$$e''^T(n) = d'^T(n) + \underline{x}'^T(n)[\hat{\mathbf{W}}(n) - \mu \bar{\mathbf{X}}(n-1)\bar{\underline{\mathbf{E}}}(n-1)]. \tag{47}$$

This can be simplified into

$$\begin{aligned} e''^T(n) &= d'^T(n) + \underline{x}'^T(n)\hat{\mathbf{W}}(n) - \mu \underline{x}'^T(n)\bar{\mathbf{X}}(n-1)\bar{\underline{\mathbf{E}}}(n-1) \\ &= d'^T(n) + \underline{x}'^T(n)\hat{\mathbf{W}}(n) - \mu r_{aa}^T(n)\bar{\underline{\mathbf{E}}}(n-1), \end{aligned} \tag{48}$$

with

$$\begin{aligned} r_{aa}(n) = r_{aa}(n-1) &+ \underline{\mathbf{x}}'^T(n)x'(n) \\ &- \underline{\mathbf{x}}'^T(n-I+1)x'(n-I+1), \end{aligned} \tag{49}$$

and

$$\underline{\mathbf{x}}'(n) = [x'(n-1) \quad \cdots \quad x'(n-K_A+1)]. \tag{50}$$

This results in the update equation for the error signal matrix:

$$\mathbf{E}(n) = \begin{bmatrix} d'^T(n) + \underline{x}'^T\hat{\mathbf{W}}(n) - r_{aa}^T\bar{\underline{\mathbf{E}}}(n-1) \\ (1-\mu)\bar{\mathbf{E}}(n-1) \end{bmatrix}, \tag{51}$$

which has dimensions $K_A \times M$.

The output $y_w(n)$ of the upper filter $\mathbf{W}$ in Fig. 4 equals

$$y_w(n) = \mathbf{W}^T(n)\underline{x}(n), \tag{52}$$

with the delay line $\underline{x}(n)$ defined as

$$\underline{x}(n) = [x^T(n) \quad \cdots \quad x^T(n-I+1)]^T. \tag{53}$$

However, in the FAP routine $\mathbf{W}$ is not directly available. So Eq. (46) is substituted into Eq. (52) resulting in

$$\begin{aligned} y_w(n) &= [\hat{\mathbf{W}}(n) - \mu \bar{\mathbf{X}}(n-1)\bar{\underline{\mathbf{E}}}(n-1)]^T\underline{x}(n) \\ &= \hat{\mathbf{W}}^T(n)\underline{x}(n) - \mu \bar{\underline{\mathbf{E}}}^T(n-1)\bar{\mathbf{X}}^T(n-1)\underline{x}(n) \\ &= \hat{\mathbf{W}}^T(n)\underline{x}(n) - \mu \bar{\underline{\mathbf{E}}}^T(n-1)r_{ax}(n), \end{aligned} \tag{54}$$

with

$$\begin{aligned} r_{ax}(n) = r_{ax}(n-1) &+ \bar{\mathbf{x}}'^T(n)x(n) \\ &- \bar{\mathbf{x}}'^T(n-I+1)x(n-I+1). \end{aligned} \tag{55}$$

The inverse in Eq. (40), in the following denoted as $\mathbf{X}_{pi}$, was computed by taking the inverse of the autocorrelation matrix $\mathbf{X}_p$, where the latter is updated by the following recursive scheme:

$$\mathbf{X}_p(n) = \begin{bmatrix} r_a & r_{aa}^T \\ r_{aa} & \bar{\mathbf{X}}_p(n-1) \end{bmatrix}, \tag{56}$$

in which

$$r_{aa}(n) = r_{aa}(n-1) + \bar{\mathbf{x}}'^{T}(n)x'(n) - \bar{\mathbf{x}}'^{T}(n-I+1)x'(n-I+1), \quad (57)$$

$$r_{a}(n) = r_{a}(n-1) + x'^{T}(n)x'(n) - x'^{T}(n-I+1)x'(n-I+1), \quad (58)$$

and in which $\bar{\mathbf{X}}_p(n-1)$ consists of the $K_A-1 \times K_A-1$-dimensional upper-left matrix of $\mathbf{X}_p(n)$.

The actual inverse is computed with a steepest-descent iteration,[19] as follows. A square matrix $\mathbf{R}$ of size $K_A \times K_A$ is defined in which each column is a row vector r_i of length K_A, in which $i=1,\ldots,K_A$ indexes the column of $\mathbf{R}$:

$$\mathbf{R} = [r_1 \quad \cdots \quad r_{K_A}]. \quad (59)$$

Each column of the inverse $\mathbf{X}_{pi}$ will be denoted by a separate vector $x_{pi,i}$ with $i=1,\ldots,K_A$ the column index:

$$\mathbf{X}_{pi} = [x_{pi,1} \quad \cdots \quad x_{pi,K_A}]. \quad (60)$$

The steepest descent iteration is performed with the following algorithm:

$$\mathbf{R} = \mathbf{I}_{K_A} - \mathbf{X}_p\mathbf{X}_{pi};$$

for $k=1$ until K_A

$$\alpha_k = \frac{r_k^T r_k}{r_k^T \mathbf{X}_p r_k};$$

$$x_{pi,k} = x_{pi,k} + \alpha_k r_k;$$

end for.

The FAP algorithm can be broken down into four steps. In the first step the output of the upper branch filter is calculated, using Eq. (54). In the second part, the autocorrelation matrix is efficiently calculated and updated, using Eq. (56). The third step will calculate the inverse of the autocorrelation matrix using a steepest descent iteration. And in the fourth and final step the error, the decorrelated error and the update rule are calculated by computation of Eqs. (51), (45), (40), and (44). One of the interesting features of the combination of RMFe with FAP is that the disturbance signals $d'(n)$ which are required for FAP are readily available from RMFe. The FAP algorithm starts with the initialization of the different variables:

$$\mathbf{X}_p = \delta\mathbf{I}_{K_A}, \quad \mathbf{X}_{pi} = \frac{1}{\delta}\mathbf{I}_{K_A}, \quad r_a = \delta,$$

$$r_{aa} = \underline{0}_{K_A-1}, \quad r_{ax} = \underline{0}_{K_A-1}.$$

The inverse $\mathbf{X}_{pi}$ at a particular sample n will be initialized with the inverse as computed at a previous sample $n-1$.

The complexity of the different algorithms will be analyzed. First assume that the algorithms will be broken down in some basic operation and then assume that multiplication and additions are counted as one operation. The complexity of the inverse outer-factor is $O(2(N+M)^2)$. A reduction of complexity to $O(6NM+2(N+M)M)$ is possible when an output normal form is used. The complexity of the transposed-conjugate inner factor is $O(2JLM)$ for PLMS and $O(4JLM)$ for RMFeLMS. The control filter W has a complexity of $O(2ILM)$. The LMS update rule needs $O(3ILM)$ computations. The AP/FAP update rule can be broken down into three operations, the first one being the error vector with a complexity of $O(MK_A+2KMK_AI)$ for AP and $O(M+2KMI)$ for FAP. In the second step the inverse matrix and the decorrelated error vector are calculated with a complexity of $O(4K_A^3+2K_A^2+2K^2M)$. In the third step the update rule will be computated with a complexity of $O(2KMI+KMIK_A+MK^2)$ for AP and $O(3KMI+M)$ for FAP. The RM-FeLMS and PLMS algorithm both have a higher complexity than FeLMS. The complexity of RMFeLMS is almost similar to that of PLMS where the extra complexity is due to the copy of W and the regularization part. The regularization ensures a practically feasible and stable algorithm. The performance and convergence properties of the hybrid filtered-error algorithm[3] are very close to those of the PLMS algorithm. However, the convergence properties and the mean square error of RMFeLMS are even better than that of PLMS.[7] It also must be noted that FeLMS and its derivatives are a different kinds of algorithms with different design trade-offs. For instance, FeLMS usually needs longer filters than RM-FeLMS.

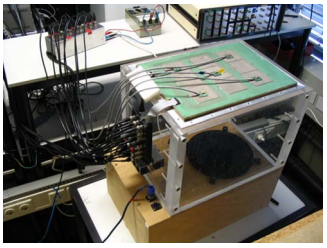


FIG. 5. (Color online) Enclosure with noise source and panel with piezoelectric actuators; the dimensions are given in Fig. 8.

IV. RESULTS

A configuration with five piezoelectric sensors and five piezoelectric actuators on a panel was used for testing of the regularized modified filtered-error algorithm. The control architecture was feedback, using the internal model control (IMC) principle. On the top and the bottom of the panel, nine piezoelectric patches were attached. The piezoelectric patches in the center of the plate and the corners of the plate were used. The panel was mounted on a box made of perspex, as shown in Fig. 5. In the bottom of this box a loudspeaker was mounted that was used to generate the primary source signal. The honeycomb panel as shown in Fig. 5 was built up from epoxy faces of 0.4 mm and an epoxy honeycomb core of 5 mm thickness. The boundaries of the plate were simply supported. The noise had a Gaussian amplitude distribution with a bandwidth of 1 kHz.

The sandwich panels as used for the experiments were designed in such a way that noise reductions could be obtained in the frequency range of interest using a relatively simple control objective based on a direct minimization of the signals from the piezoelectric sensors. Some of the considerations leading to the final design will be mentioned. First, the actuator configuration was selected such that the

structural modes radiating to the far field[20] could be controlled. This led to a five-actuator configuration similar to that of Sors and Elliott.[21] Second, the sensors were placed collocated with respect to the actuators in order to minimize control spillover, taking into account possible in-plane coupling effects. In-plane coupling[22] between piezoelectric patch actuators and piezoelectric patch sensors, which can have negative consequences for the control of the acoustically relevant out-of-plane vibrations, was reduced by the honeycomb core of the sandwich structure. Finite element simulations of the different configurations and a subsequent analysis based on pole-zero distances[22] showed that insertion of such a core between the piezoelectric patch actuator and the piezoelectric patch sensor improved the damping performance of a feedback control system, approximating the performance of a combination of a piezoelectric patch actuator and an accelerometer. Different types of sensors were also evaluated in real-time control experiments. The use of accelerometers led to noise reductions that were approximately 1 dB higher than obtained with the piezoelectric patch sensors. However, the use of piezoelectric patches was found to lead to more robust performance, probably due to the spatial averaging by these sensors[23] and/or the reduced high-frequency content of the sensor signals. In this paper only the results for the piezoelectric patch sensors will be shown.

The controller was implemented on an embedded PC running an RT-Linux operating system. The controller was designed in SIMULINK/MATLAB using custom blocks that were written in the C programming language. The code for the simulation in SIMULINK was identical to the code to generate the real-time code, using the REALTIME WORKSHOP. Analog I/O for this system was implemented on a dedicated module providing 16 analog inputs and 16 analog outputs. This analog-to-digital/digital-to-analog (ADDA) module samples at a higher sample rate than the control algorithm. This high sample rate, which was set to 100 kHz in the experiments, was reduced by means of a decimator on the input and increased by means of an interpolator at the output, using filtering in discrete time. Only a simple filter in the analog domain is needed, reducing the need for complex analog electronic filtering. This setup makes it is possible to use different antialiasing and reconstruction filters with different cut-off frequencies. To reduce the latency for the feedback controller only the interpolation filter was enabled. The sample rate of the controller was set to 2 kHz. In the first test the convergence rate was obtained by measuring the error signal and storing it in a buffer on the platform. This buffer was then read by a host PC in order to perform the necessary postprocessing.

The plant was estimated using a subspace identification technique.[24] These state-space models were converted into the output-normal form.[25] The advantage of the output-normal form is that it can be more efficiently calculated than the underlying state-space system. The state-space model used for the secondary path was of the 50th order. The estimated secondary plant model had a variance accounted for (VAF) of 99.93%. The regularization parameter β was set in such a way that regularization was at a level of 30 dB below the largest frequency-domain peak of G. The length of the time reversed transpose inner-factor was set to $J=80$. The length of the LMS filters was set to $I=40$. The following parameters were set for the LMS algorithm: The step size was set to $\alpha=\frac{1}{20}$ and the leakage was set to $\gamma=1\times10^{-5}$. All parameter settings were first tested by means of simulations.

The convergence curves for the five input channels can be found in Fig. 6. The convergence curve was obtained by averaging 32 realizations of the squared error signals. For the measurements the controller was disabled for 5 s and then switched on for 55 s. The data were normalized to the average disturbance level of input channel 3, thereby defining the 0 dB level in Fig. 6. The control system reached 10 dB reduction within 5 s, i.e., 10 000 samples. The overall reduction for all five channels after 55 s was 12.4 dB for the LMS algorithm and 12.6 dB for a fourth-order FAP algorithm. The overall reduction for the five input channels was obtained from the ratio of the total power for the disturbance signals divided by the total power for the error signals. Increasing the FAP order to higher orders did not improve the performance anymore. It can be seen that there is almost no difference between RMFeLMS and RMFeFAP for this feedback control architecture. This can be expected because a feedback controller tries to make the reference signal white, in which case there is no advantage of FAP over LMS. The spectrum for steady-state operation can be found in Fig. 7. The overall steady-state reduction over all five input channels was 14.1 dB, which is somewhat higher than the 12.6 dB reduction after 55 s. The data used to obtain the steady-state reduction were also used to calculate the optimal Wiener filter solution, which is constrained to be causal. The Wiener filter was computed with the methods described in Ref. 26. The particular computation method for the Wiener filter was based on the use of a single controller W set to reduce the filtered error signals $e'(n)=\bar{G}_i^*(q)D(q)\times[e(n);G_{\mathrm{reg}}(q)u(n)]$. The inputs for the Wiener filter computation were the reference signal $x(n)$, the secondary path $\bar{G}_i^*(q)D(q)\bar{G}(q)\bar{G}_o^{-1}(q)$, and the disturbance signal $\bar{G}_i^*(q)D(q)[d(n);G_{\mathrm{reg}}(q)u(n)]|_{u(n)=0}$, in which the controller is switched off. The model of the secondary path to derive the Wiener filter is the transfer path from the output of the upper block W in Fig. 4 to the signal $e'(n)$ without the part consisting of the single delay block. This delay block was used for improved convergence of the modified adaptive algorithm but should be omitted in the secondary path for computing the optimum Wiener filter, since the optimum Wiener filter as computed here is based on the input signal $x(n)$ instead of $x'(n)$. For feedback control the reference signal was equal to the disturbance signal; in that case only the disturbance signal $d(n)$ was measured. For feedforward control the reference signal $x(n)$ was also measured. The resulting Wiener filter W was then used to compute the error signals $e(n)$. Subsequently, the optimal reduction using the Wiener filter was obtained by comparing $d(n)$ with $e(n)$. In this way it was possible to study the influence of different approximations and parameter settings in the RMFe algorithm on the steady-state performance. The maximum theoretical overall reduction for the five error sensors according to the Wiener filter was 15.5 dB. The real measured overall

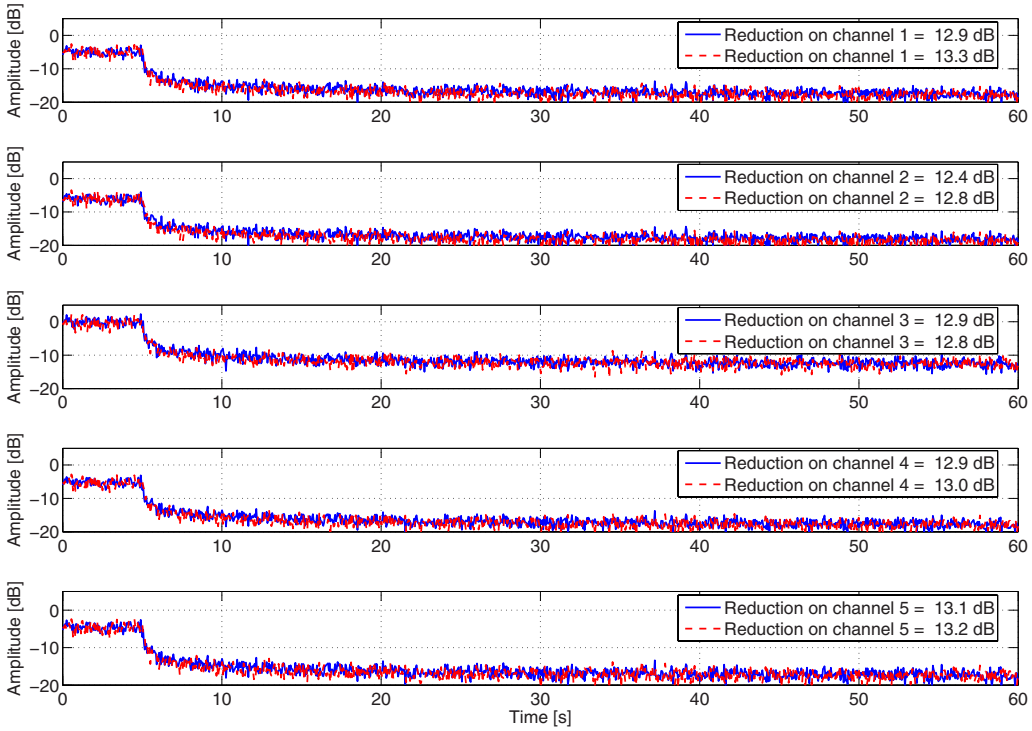


FIG. 6. (Color online) Convergence curve for a real-time five-input five-output feedback control system; the reduction after 55 s for five channels is 12.4 dB (RMFeLMS, solid line) and 12.6 dB (RMFeFAP, dashed line).

reduction on the test setup was 14.1 dB. The difference between the estimated results and the real results can be explained by errors in the IMC model, bias, noncorrelated noise sources, and nonlinearity in the measurement setup, of which the error in the IMC model was the most important. Finally the standard filtered-error algorithm was tested on the feedback test setup, in order to compare it to the RMFe algorithm. Due to the inherent delay in the filtered-error algorithm and the absence of preconditioning the step size had to be set to a relatively small value. The controller already be-

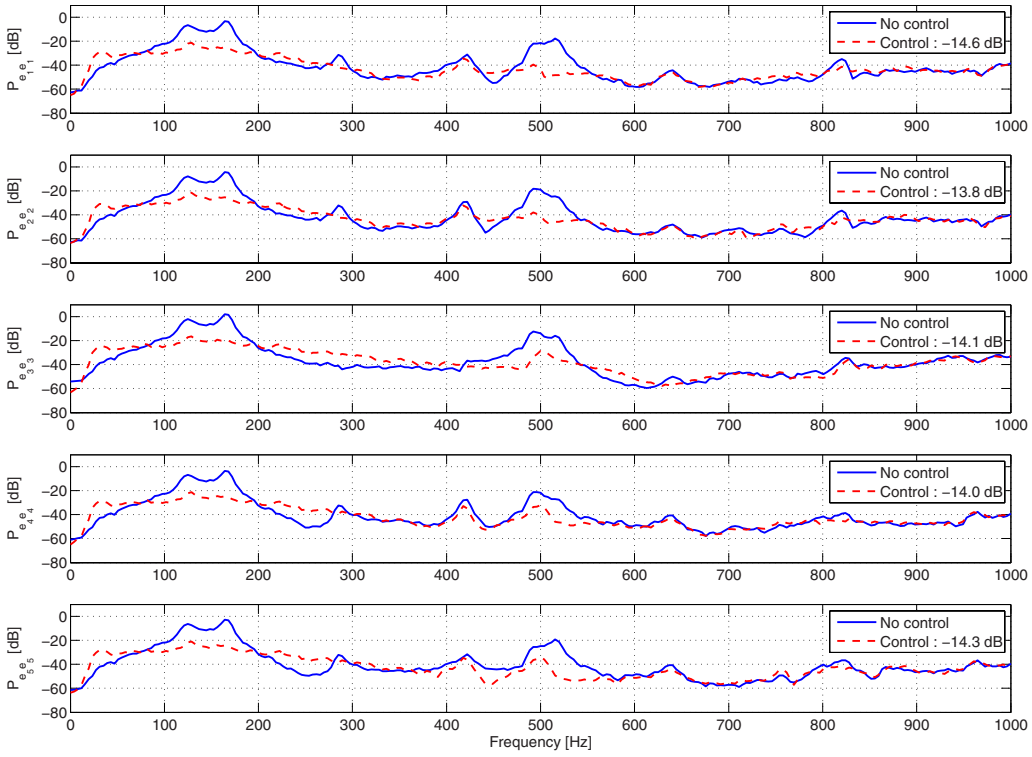


FIG. 7. (Color online) Spectrum of the error signals for a converged real-time feedback controller using the RMFeFAP algorithm; the overall reduction is 14.1 dB at $\alpha=\frac{1}{40}$.

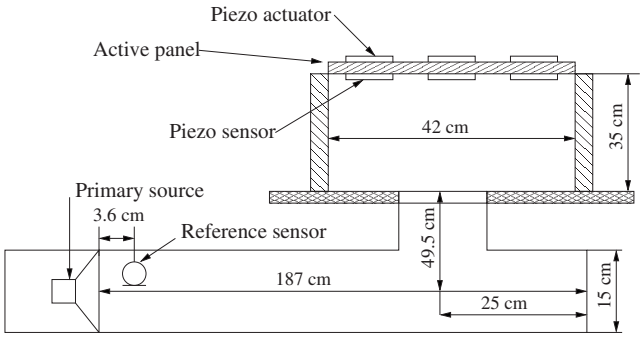


FIG. 8. Schematic overview of the setup used for testing of the feedforward RMFe algorithms consisting of a circular pipe connected to the rectangular perspex box of Fig. 5; the depth of the perspex box is 30 cm.

came unstable with a μ of $\frac{1}{1000}$. For the experiments a value of 5×10^{-4} was used, resulting in approximately 4 dB reduction after 55 s and more than 1 h to reach the steady state reduction.

In order to test the FAP algorithm a modification of the test setup was devised. The modification consisted of a circular pipe connected to the bottom of the perspex box. A schematic overview of the setup can be found in Fig. 8. With this setup the reference signal will be colored due to resonances in the pipe. The real-time results of the RMFeLMS and RMFeFAP implementations can be found in Fig. 9.

For the FAP algorithm, the following settings were used: $\mu=\frac{1}{40}$, $\delta=\frac{1}{4}$, and $K_A=28$. The LMS parameters were $\alpha=\frac{1}{10}$, $\gamma=10^{-5}$. In both algorithms the same regularization parameter β as for the feedback control experiments was used. The length of the time reversed transpose inner-factor was $J=80$ and the length of the adaptive filter was $I=350$ taps. Both the LMS and the FAP algorithm used an IMC algorithm to compensate for the feedback from the actuators to the reference microphone. The IMC model was estimated together with the model from the secondary path using the same stimuli. It was found that IMC was necessary for proper operation, also for this feedforward configuration. The estimation accuracy (VAF) was 99.86% for the secondary path model G and 99.49% for the IMC model G_{rp}. The order of both models was 50. Both models were transformed into the output-normal form. Complexity of the FAP-related computation became dominant for K_A orders of 32 and higher.

From the results in Fig. 10 it can be seen that FAP has better convergence properties than the LMS algorithm. After 55 s, the LMS algorithm reaches a reduction of 11.5 dB, whereas the FAP algorithm reaches a reduction of 13.5 dB. Both algorithms are still not at their maximum reduction after 55 s. The optimal Wiener filter predicted a reduction of 14.0 dB. The real-time implementations of both the LMS algorithm and the FAP algorithm were able to realize this reduction. The spectra for the feedforward controller in steady-state conditions can be found in Fig. 10. Only the result for the RMFeFAP implementation is shown since the result for RMFeLMS is nearly identical. The maximum reduction reached on all five channels is 14.3 dB, which is 0.3 dB higher than the optimal Wiener filter solution. Due to the fact that only one observation is used there was some variance in a single observation. Furthermore it needs to be noted that especially the FAP algorithm can dynamically

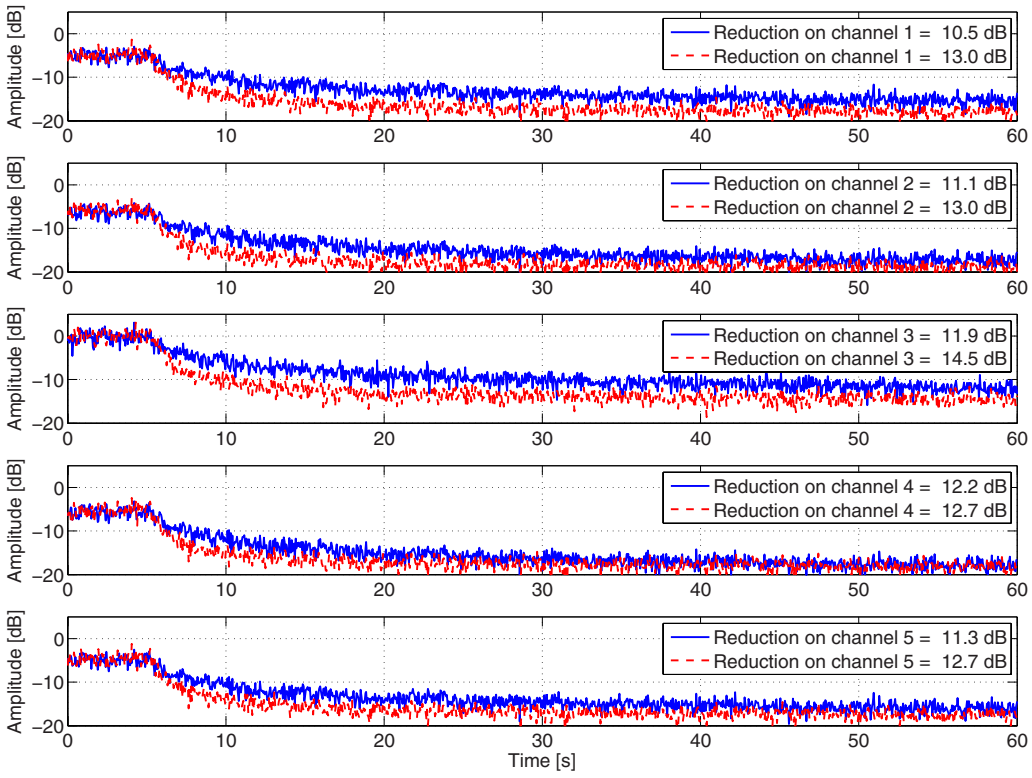


FIG. 9. (Color online) Convergence curve for real-time feedforward control using RMFeFAP (dashed line) and RMFeLMS (solid line); after 55 s, RMFeFAP reached 13.5 dB overall reduction and RMFeLMS reached 11.5 dB overall reduction.

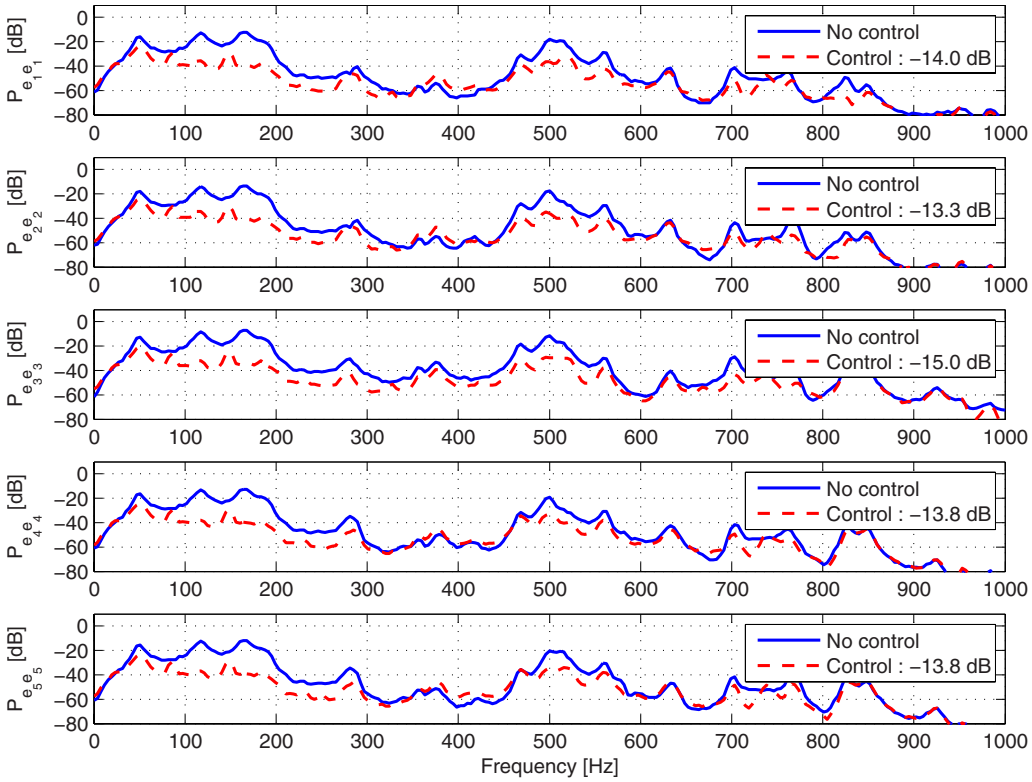


FIG. 10. (Color online) Spectrum of the error signals for a converged real-time feedforward controller using the RMFeFAP algorithm; the overall reduction is 14.3 dB.

track small changes in the statistics of the source. The Wiener filter is only optimal for one observation and cannot track changes in the source statistics.

In addition to the steepest-descent method for matrix inversion, also a conjugate gradient method was implemented. Use of the conjugate gradient method did not lead to significant performance advantages. A simplification of the FAP algorithm as described in the present paper is possible by assuming $\mu=1$. Then also the full matrix inversion can be reduced to solving a single system of equations.[13] However, it was found from experiments that the performance deteriorated because of this simplification; the resulting error signals were approximately 1 to 2 dB higher. Finally, tests were performed in order to verify that the feedforward configuration indeed led to performance advantages as compared to the feedback controller. On the setup of Fig. 8, a feedback controller resulted in a reduction of 7.9 dB, i.e., substantially less than the 14.3 dB reduction using the feedforward controller.

In order to evaluate the noise reductions, additional tests were carried out on a similar panel with integrated lightweight electronics. Apart from the integrated electronics, the resonance frequencies and other properties of this panel were very similar to the other panel. The noise reduction of the panel was measured by positioning microphones at distances of 50 cm from the panel. The average noise reduction for a feedback control system was 8 dB; the reduction on the error sensors was 10.2 dB. For a feedforward control system the average noise reduction was 7 dB; the reduction on the error sensors was 15.5 dB. If higher noise reductions are to be obtained, particularly with feedforward control, then probably the cost function of the control algorithm should be modified in order to include a relationship between the sensor signals and the acoustic radiation from the panel, such as with a radiation mode based technique.[27]

V. CONCLUSIONS

In this paper results were given of a real-time implementation of the regularized modified filtered-error algorithm. The algorithm was extended with the fast affine projection method in order to improve the speed of convergence for colored reference signals. It was shown how to combine the extension with the RMFe algorithm. In both feedback and feedforward control scenarios, the controller was used to reduce the vibrations of a panel by driving piezoelectric patches. The maximum difference of the predicted reduction and the reduction obtained in real-time experiments was 1.4 dB for feedback control and 0.3 dB for feedforward control. It was shown that the RMFeFAP algorithm has better convergence properties than the RMFeLMS algorithm for the case of colored reference signals, which was verified in a feedforward control configuration. For feedback control no significant differences between the algorithms were found, as expected.

ACKNOWLEDGMENTS

Part of the work as described in this paper were performed in the InMAR project of the EU-6th Framework Programme. The authors would like to thank Geert Jan Laanstra and Henny Kuipers for their excellent support and A. Derae-

maecker from ULB for the finite element simulations in order to evaluate the different actuator-sensor combinations.

APPENDIX: DERIVATION OF THE FAP ALGORITHM

Starting with the affine projection update rule Eq. (39), expressing the current filter taps in the original filter estimate and all subsequent $\mathbf{X}(n)$ and $\boldsymbol{\xi}(n)$, and assuming that the system starts at $n=0$ with the delay lines set to zero, results in

$$\mathbf{W}(n+1) = \mathbf{W}(0) - \mu \sum_{i=0}^{n-1} \mathbf{X}(n-i)\boldsymbol{\xi}(n-i). \tag{A1}$$

The vector-matrix product is expanded into

$$\mathbf{W}(n+1) = \mathbf{W}(0) - \mu \sum_{i=0}^{n-1} \sum_{j=0}^{K_A-1} \underline{x}'(n-j-i)\xi_j(n-i), \tag{A2}$$

in which $\xi_j(n)$ is defined as the $j-1$th row of $\boldsymbol{\xi}(n)$, i.e., $\boldsymbol{\xi}(n)=[\xi_0^T(n)\xi_1^T(n)\cdots\xi_{K_A-1}^T(n)]^T$. Exchanging the summations and substitution of $k=j+i$ and $i=k-j$ leads to

$$\mathbf{W}(n+1) = \mathbf{W}(0) - \mu \sum_{j=0}^{K_A-1} \sum_{k=j}^{j+n-1} \underline{x}'(n-k)\xi_j(n-k+j). \tag{A3}$$

The second sum is split into two parts: one from $k=j$ to $k=K_A-1$ and one from $k=K_A$ to $k=j+n-1$, such that

$$\mathbf{W}(n+1) = \mathbf{W}(0) - \mu \sum_{j=0}^{K_A-1} \sum_{k=K_A}^{j+n-1} \underline{x}'(n-k)\xi_j(n-k+j) - \sum_{j=0}^{K_A-1} \sum_{k=j}^{K_A-1} \underline{x}'(n-k)\xi_j(n-k+j). \tag{A4}$$

It can be shown that the j in the upper bound of the second sum of the first double summation in Eq. (A4) can be removed because $\underline{x}'(n)=0$ for $n\leqslant 0$. Substitution of the upper bound in $\underline{x}'(n-k)$ results in $\underline{x}'(-j+1)$, i.e., $\underline{x}'(1)$ for $j=0$ and $\underline{x}'(0)$ for $j=1$ and so on. This means that the terms with $j>0$ will not contribute to the summation and can therefore be removed. The first double summation, denoted as S_1, can thus be written as

$$S_1 = \mu \sum_{k=K_A}^{n-1} \underline{x}'(n-k) \sum_{j=0}^{K_A-1} \xi_j(n-k+j). \tag{A5}$$

The second double sum in Eq. (A4), denoted as S_2, can also be expressed in a form similar to that of Eq. (A5) after reordering of the different terms. Expansion of the second double summation shows that

$$\begin{aligned} S_2 = {} & \underline{x}'(n)\xi_0(n) + \underline{x}'(n-1)\xi_0(n-1) + \ldots + \underline{x}'(n-K_A+1) \\ & \times \xi_0(n-K_A+1)) + \underline{x}'(n-1)\xi_1(n) + \underline{x}'(n-2) \\ & \times \xi_1(n-1) + \ldots + \underline{x}'(n-K_A+1))\xi_1(n-K_A+2) \\ & + \ldots \underline{x}'(n-K_A+1)\xi_{K_A-1}(n). \end{aligned}$$

From the latter expansion it can be seen that the second double summation can be written as

$$S_2 = \mu \sum_{k=0}^{K_A-1} \underline{x}'(n-k) \sum_{j=0}^{k} \xi_j(n-k+j). \tag{A6}$$

Using Eqs. (A5) and (A6), Eq. (A2) can be expressed as

$$\begin{aligned} \mathbf{W}(n+1) = {} & \mathbf{W}(0) - \mu \sum_{k=0}^{K_A-1} \underline{x}'(n-k) \sum_{j=0}^{k} \xi_j(n-k+j) \\ & - \mu \sum_{k=K_A}^{n-1} \underline{x}'(n-k) \sum_{j=0}^{K_A-1} \xi_j(n-k+j). \end{aligned} \tag{A7}$$

Equation (A5) and $\mathbf{W}(0)$ are used to define $\hat{\mathbf{W}}(n)$:

$$\hat{\mathbf{W}}(n) = \mathbf{W}(0) - \mu \sum_{k=K_A}^{n-1} \underline{x}'(n-k) \sum_{j=0}^{K_A-1} \xi_j(n-k+j). \tag{A8}$$

Equation (A6) will be rewritten as

$$S_2 = \mu \sum_{k=0}^{K_A-1} \underline{x}'(n-k)h(n,k), \tag{A9}$$

in which

$$h(n,k) = \sum_{j=0}^{k} \xi_j(n-k+j). \tag{A10}$$

In the right-hand side of Eq. (A9) a simple sum of base vectors $\underline{x}'(n-k)$ multiplied by a scaling coefficient $h(n,k)$ can be recognized. Let us define the following vector of scaling coefficients:

$$\underline{\mathbf{E}}(n) = \begin{bmatrix} h(n,0) \\ \vdots \\ h(n,K_A-1) \end{bmatrix} = \begin{bmatrix} \xi_0(n) \\ \xi_1(n) + \xi_0(n-1) \\ \vdots \\ \xi_{K_A-1}(n) + \ldots + \xi_0(n-(K_A-1)) \end{bmatrix}. \tag{A11}$$

The base is now equal to the $\mathbf{X}(n)$ matrix, so that the right-hand side of Eq. (A9) can be written as the matrix-vector product $\mu\mathbf{X}(n)\underline{\mathbf{E}}(n)$. Then Eq. (A7) can be written as

$$\mathbf{W}(n+1) = \hat{\mathbf{W}}(n) - \mu\mathbf{X}(n)\underline{\mathbf{E}}(n). \tag{A12}$$

Equation (A8) is expressed as an iterative update rule

$$\begin{aligned} \hat{\mathbf{W}}(n+1) &= \hat{\mathbf{W}}(n) - \mu\underline{x}'(n-(K_A-1)) \times \sum_{j=0}^{K_A-1} \xi_j(n-(K_A-1)+j) \\ &= \hat{\mathbf{W}}(n) - \mu\underline{x}'(n-(K_A-1)) \times \underline{E}_{K_A-1}(n), \end{aligned} \tag{A13}$$

in which $\underline{E}_{K_A-1}(n)$ is defined as the bottom row of $\underline{\mathbf{E}}(n)$. Equations (A12) and (A13) are used to express the updated controllers as

$$\mathbf{W}(n+1) = \hat{\mathbf{W}}(n+1) - \mu\bar{\mathbf{X}}(n)\bar{\underline{\mathbf{E}}}(n), \tag{A14}$$

in which $\bar{\underline{\mathbf{E}}}(n)$ are the $K_A - 1$ top-most rows of $\underline{\mathbf{E}}(n)$ and $\bar{\mathbf{X}}(n)$ are the $K_A - 1$ left-most columns of $\mathbf{X}(n)$.

Using Eq. (A11), this can be written recursively as

$$\underline{\mathbf{E}}(n) = \underline{\boldsymbol{\xi}}(n) + \begin{bmatrix} \underline{0}_M^T \\ \bar{\underline{\mathbf{E}}}(n-1) \end{bmatrix}, \tag{A15}$$

in which $\underline{0}_M$ is a column vector filled with M zeros and $\bar{\underline{\mathbf{E}}}$ are the upper $K_A - 1$ rows of $\underline{\mathbf{E}}$.

[1] S. M. Kuo and R. D. Morgan, *Active Noise Control Systems* (Wiley, New York, 1996).
[2] S. J. Elliott, *Signal Processing for Active Control* (Academic, London).
[3] V. E. DeBrunner and Z. Dayong, "Hybrid filtered error LMS algorithm: Another alternative to filtered-x LMS," IEEE Trans. Circuits Syst., I: Regul. Pap. **53**, 653–661 (2006).
[4] S. J. Elliott, "Optimal controllers and adaptive controllers for multichannel feedforward control of stochastic disturbances," IEEE Trans. Signal Process. **48**, 1053–1060 (2000).
[5] S. Ishimitsu and S. J. Elliott, "Improvement of the convergence property of adaptive feedforward controllers and their application to the active control of ship interior noise," Acoust. Sci. & Tech. **25**, 181–187 (2004).
[6] M. R. Bai and S. J. Elliott, "Preconditioning multichannel adaptive filtering algorithms using EVD- and SVD-based signal prewhitening and system decoupling," J. Sound Vib. **270**, 639–655 (2004).
[7] A. P. Berkhoff and G. Nijsse, "A rapidly converging filtered-error algorithm for multichannel active noise control," Int. J. Adapt. Control Signal Process. **21**, 556–569 (2007).
[8] K. Ozeki and T. Umeda, "Adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," Electron. Commun. Jpn. **67**, 19–27 (1984).
[9] S. L. Gay, "Fast projection algorithms with application to voice echo cancellation," Ph.D. thesis, The State University of New Jersey, New Brunswick, NJ, 1994.
[10] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *ICASSP-95, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, New York, May 9–12, 1995, pp. 3023–3026.
[11] M. Rupp, "A family of adaptive filter algorithms with decorrelating properties," IEEE Trans. Signal Process. **46**, 771–775 (1998).
[12] M. Bouchard and F. Albu, "The Gauss-Seidel fast affine projection algorithm for multi-channel active noise control and sound reproduction systems," Int. J. Adapt. Control Signal Process. **19**, 107–123 (2005).
[13] D. Heping, "A stable fast affine projection adaptation algorithm suitable for low-cost processors," in *ICASSP '00, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, New York, June 5–9, 2000, pp. 360–363.
[14] S. Oh, D. Linebarger, B. Priest, and B. Raghothaman, "A fast affine projection algorithm for an acoustic echo canceller using a fixed-point DSP processor," in *ICASSP '97, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, New York, April 21–24, 1997, pp. 4121–4124.
[15] D. Heping, "Fast affine projection adaptation algorithms featuring stable symmetric positive-definite linear system solvers," in *2005, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE, New York, October 16–19, 2005, pp. 166–169.
[16] S. C. Douglas, "Efficient approximate implementations of the fast affine projection algorithm using orthogonal transforms," in *ICASSP, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Piscataway, NJ, May 7–10, 1996, pp. 1656–1659.
[17] M. Vidyasagar, *Control System Synthesis: A Factorization Approach* (MIT, Boston, 1985).
[18] A. H. Sayed, *Fundamentals of Adaptive Filtering* (IEEE Press, New York, 2003).
[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2 ed. (John Hopkins, Baltimore, 1989).
[20] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration*, 2nd ed. (Academic, Orlando, 1997), pp. 24–28.
[21] T. C. Sors and S. J. Elliott, "Modelling and feedback control of sound radiation from a vibrating panel," Smart Mater. Struct. **8**, 301–314 (1999).
[22] A. Preumont, *Vibration Control of Active Structures*, 2nd ed. (Kluwer Academic, Dordrecht, 2002).
[23] A. P. Berkhoff, "Piezoelectric sensor configuration for active structural acoustic control," J. Sound Vib. **246**, 175–183 (2001).
[24] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems* (Kluwer Academic, Dordrecht, 1996).
[25] R. A. Roberts and C. T. Mullis, *Digital Signal Processing* (Addison-Wesley, Reading, MA, 1987).
[26] A. P. Berkhoff, "Control strategies for active noise barriers using near-field error sensing," J. Acoust. Soc. Am. **118**, 1469–1479 (2005).
[27] G. P. Gibbs, R. L. Clark, D. E. Cox, and J. S. Vipperman, "Radiation modal expansion: Application to active structural acoustic control," J. Acoust. Soc. Am. **107**, 332–339 (2000).

Noise reduction in tunnels by hard rough surfaces

Ming Kan Law
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Kai Ming Li[a)]
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 Martin Jischke Drive, West Lafayette, Indiana 47907-2031

Chun Wah Leung
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

(Received 2 January 2008; revised 21 April 2008; accepted 30 April 2008)

This paper examines the feasibility of using two-dimensional hard rough surfaces to reduce noise levels in traffic tunnels with perfectly reflecting boundaries. First, the Twersky boss model is used to estimate the acoustic impedance of a hard rough surface. Second, an image source model is then used to compute the propagation of sound in a long rectangular enclosure with finite impedance. The total sound fields are calculated by summing the contributions from all image sources coherently. Two model tunnels are built to validate the proposed model experimentally. Finally, a case study for a realistic geometrical configuration is presented to explore the use of hard rough surfaces for reducing traffic noise in a tunnel which is constructed with hard boundaries.
 [DOI: 10.1121/1.2932255]

PACS number(s): 43.50.Gf, 43.28.Js, 43.20.Mv, 43.20.El [KA] Pages: 961–972

I. INTRODUCTION

Traffic tunnels are widely constructed to convey road and rail traffic worldwide. Reduction in noise in tunnels is a challenging task. The geometrical designs of road traffic or railway tunnels are essentially unfavorable for implementing noise control measures. Large aspect ratios and hard boundary surfaces have led to unevenly distributions of sound fields and tremendous buildup of noise levels due to the effect of multiple reflections from the tunnel walls. The improvement of the acoustic environments in traffic tunnels is an important issue not only for the benefit of the overall noise reductions but also for improving the acoustic performance of public address systems.

There has been considerable research into the physical phenomena of sound propagation in long enclosures.[1] Numerical models, indoor scale-model experiments and full-scale outdoor field measurements were developed for predicting sound pressure levels, reverberation times and other noise metrics in underground stations,[2,3] corridors,[4] road, and railway traffic tunnels.[5–8] Different geometrical acoustic models, the incoherent and coherent approaches, are frequently used in previous studies because of their solutions are usually cast in a relatively simple form for easy numerical implementation. The incoherent model is also known as the energy approach, which has been extensively used in room acoustics and early studies of sound propagation in urban environments.[2–12] In a recent study for predicting sound fields in a street canyon,[13] the incoherent model was found to be accurate if the width of a street is greater than 10 m. Picat *et al.*[14] also showed experimentally that the incoherent model was adequate if the boundary surfaces of a street canyon reflected sound diffusely. On the other hand, the coherent approach is based on a complex image model. Recent studies[15,16] have shown that the coherent model can lead to more accurate predictions of the sound pressure levels and reverberation times in long enclosures because the model takes into account of the interference effect due to different image sources.

In principle, it is possible to apply conventional sound absorption materials on the tunnel walls for noise reduction.[17] However, their use for noise reduction in tunnels poses a health safety problem because of the known aging effects of fiber glass. In addition, tunnel surfaces are usually designed to reduce potential risks of fire hazards and to minimize the costs associated with cleaning and maintenance for the sound absorption materials.

It has been known that a hard rough surface offers incoherent scattering of incoming sound waves. This phenomenon results in the creation of apparent impedances on the hard rough reflecting surfaces.[18–21] Theoretical predictions according to the boundary element formulation, indoor and outdoor measured data have all confirmed the effect on the propagation of sound above a hard rough surface.[22–24] With this information, we wish to examine the effect of deliberate roughening of an otherwise acoustically hard surface on the propagation of sound in tunnels and to explore their potential use for noise reduction.

In Sec. II, we describe the Twersky's boss model and the complex image source model. They are used to describe the effect of hard rough surface on sound reflections and the propagation of sound in long enclosure, respectively. The classical Weyl-van der Pol formula[25] will be used to compute the propagation of sound at near grazing incidence from an image source over a hard rough surface in the tunnels. Section III presents two sets of experimental results carried out;

[a)]Electronic mail: mmkmli@purdue.edu

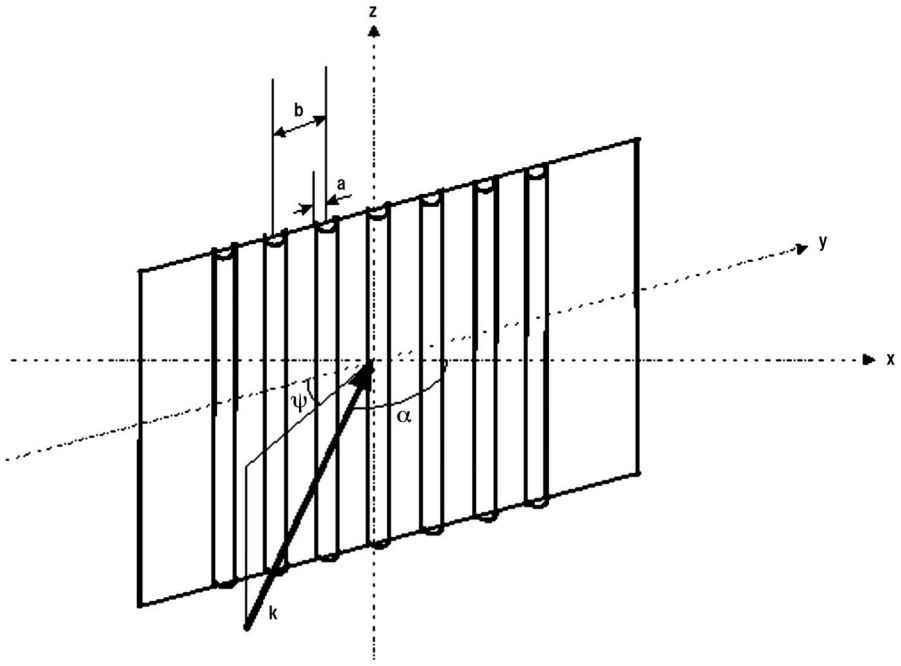


FIG. 1. A schematic diagram to show an incident wave on a rough surface which is made of semicylindrical bosses.

first, in a smaller scale model long enclosure built in an anechoic chamber and, second, a larger one-tenth scale model tunnel mounted outdoors for validating the theoretical formulations. Numerical simulations of the sound fields in a road traffic tunnel and the use of hard rough surfaces for noise reduction are presented in Sec. IV. Concluding remarks and discussions are offered in the last section.

II. THEORY

A. Twersky's boss theory for hard rough planes

Twersky's model describes the coherent reflection and incoherent scattering of sound from a rigid surface embedding with semicylindrical bosses. His mathematical formulation takes into account the interaction of scattering sounds between the neighboring bosses, which enables the use of an effective impedance to predict the coherent reflection from a rough surface. Twersky's model can be applied to calculate the plane wave reflection coefficient of a hard rough surface. Using the notations consistent with the previous publications,[23] the main results for a hard rough surface are stated in the following paragraphs although the details of Twersky's theory can be found elsewhere.[18]

Consider a rough surface that is made by spacing arrays of semicylindrical bosses either periodically or nonperiodically over a rigid plane. A plane wave is incident on this hard rough surface locating at the plane $x=0$ (see Fig. 1 for the schematic diagram of the problem). According to the two-dimensional Twersky boss model, the effective admittance β_e (normalized with air) of the hard rough surface can be evaluated by

$$\beta_e = \eta_{2D} - i\xi_{2D}, \tag{1}$$

where

$$\eta_{2D}(\alpha,\varphi) = \frac{mk^3\pi^2a^4}{8}(1-W^2)\left\{(1-\sin^2\alpha\sin^2\varphi)\left[1 + \left(\frac{\delta^2}{2}\cos^2\varphi - \sin^2\varphi\right)\sin^2\alpha\right]\right\}, \tag{2}$$

$$\xi_{2D}(\alpha,\varphi) = kV[-1 + (\delta\cos^2\varphi + \sin^2\varphi)\sin^2\alpha], \tag{3}$$

where $k(\equiv\omega/c)$ is the wave number, with ω and c as the angular frequency and speed of sound in air, respectively. The parameters η_{2D} and ξ_{2D} are the incoherent scattering loss and multipole coupling terms, respectively, α is the angle of incidence measured from the normal of the rigid plane, and φ is the azimuthal angle between the wave vector and the axes of semicylinders. $V=m\pi a^2/2$ represents the total raised volume per unit area of the bosses, where m is the number of semicylinders per unit length ($m=1/b$) and $\delta=2/(1+I)$ is a measure of the coupling effects between the semicylinders. The parameters a and b are the radius and mean center-to-center spacing of the semi-cylinders, respectively. If the semicylindrical bosses are periodically arranged, then I can be approximated by $(\pi a)^2/3b^2$. Otherwise, it is given by

$$I \cong \begin{cases} 2W(1+0.307W+0.137W^2)(a^2/b^2) & \text{for } W<0.8 \\ \dfrac{\pi^2}{3}\left[1-\dfrac{2(1-W)}{W}\right] + 6\dfrac{(1-W)^2}{W^2}\left[\dfrac{\pi^2}{6}+1.202\right](a^2/b^2) & \text{for } W\geqslant 0.8, \end{cases} \tag{4}$$

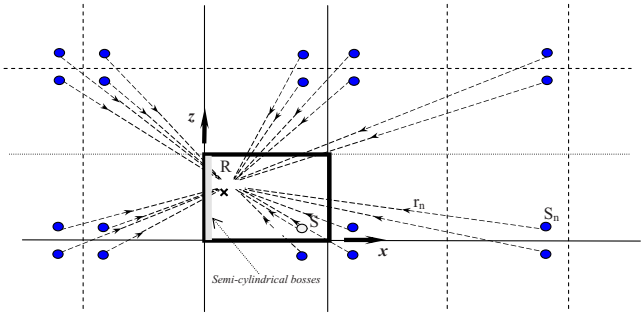


FIG. 2. (Color online) Typical image sources formed in a rectangular enclosure by multiple reflections on the boundary surfaces.

where $W=mb^*=b^*/b$, b^* is the minimum center-to-center separation between two adjacent semicylinders. The term $(1-W)^2$, which is also known as a packing factor, is introduced to account for random distributions of semicylinders. The Twersky theory is based on the assumptions that the product of roughness size and wave number is very much less than unity, and the product of roughness separation distance and wave number is less than unity.

B. Sound propagation in long enclosures with a hard rough surface

Consider the situation for calculating the sound field due to a point source of unit strength radiating sound inside a rectangular enclosure. The sound field is composed of a direct contribution from the source and the contributions from image sources due to multiple reflections inside the enclosure (see Fig. 2). The total sound field at a receiver point can be computed by summing all contributions to give,[15]

$$P_{\text{tot}} = \frac{e^{ikr_d}}{4\pi r_d} + \frac{1}{4\pi}\sum_{n=0}^{N} Q_{sn}\frac{e^{ikr_n}}{r_n}, \tag{5}$$

where N is the maximum number terms used in the ray series and r_d and r_n are the distances of direct path and specularly reflected path with image source n from the receiver, respectively. Unless otherwise stated, the subscript n denotes the corresponding parameters for the image source n. In Eq. (5), a factor known as the combined complex wave reflection coefficient Q_{sn} is used. At each interaction with a boundary plane, the complex reflection coefficient Q_n is determined according to

$$Q_n = R_n + (1-R_n)F(w_n), \tag{6a}$$

where the plane wave reflection coefficient R_n is given by

$$R_n = (\cos\alpha_n - \beta_n)/(\cos\alpha_n + \beta_n), \tag{6b}$$

the boundary loss factor $F(w_n)$ is calculated by

$$F(w_n) = 1 + i\sqrt{\pi}w_n e^{-w_n^2}\,\text{erfc}(-iw_n), \tag{7}$$

w_n is the numerical distance determined by

$$w_n = \sqrt{\tfrac{1}{2}ikr_n}(\cos\alpha_n + \beta_n), \tag{8}$$

β_n is the specific normalized admittance of the reflecting plane, and α_n is the incident angle of the reflected wave. The successive reflections of a sound wave is then modeled as the multiplication product for all the associated Q_n of the image source n, i.e., the complex wave reflection coefficient is multiplied each time when the reflected wave interacts with a boundary surface. The combined complex wave reflection factor is denoted by Q_{sn} in Eq. (5).

The specific normalized admittance β_n is zero if the reflecting surface is a hard smooth plane. On the other hand, if the reflected wave hits the rough surface, then β_n is calculated according to Eqs. (1)–(4). With the use of Eq. (5), the sound field for the propagation of sound due to the presence of rough surfaces inside a long enclosure can be computed.

C. The effect of semicylindrical bosses on sound propagation in tunnels

To investigate the effect of the hard rough surfaces inside a tunnel, it is useful to examine the spherical wave reflection coefficients Q_n for each interaction with the rough surface. It is noted that the modified admittance β_n is different because of the fact the incident and azimuthal angles are different for different image sources. With different values of β_n, the plane wave reflection coefficients R_n and the spherical wave reflection coefficients Q_n are all different. However, both the plane wave reflection and spherical wave reflection coefficient become 1 when β_n tends to zero (i.e., for a hard smooth surface).

According Eq. (1), the Twersky's boss model predicts a nonzero real part η_{2D} for the effective admittance of a hard rough surface. Physically, this "loss" term may be attributed to the infinitesimal viscous losses due to heat convection occurred on the layer between the boss' surface and the atmosphere upon the multiple incoherent scattering of waves between neighboring rigid bosses and the rigid plane. The hard rough surfaces can also be interpreted as the presence of diffusely reflecting surfaces and Kang[1] has demonstrated the effectiveness of using diffusely reflecting surfaces (without increasing absorption) to reduce noise in long spaces.

The apparent "absorptiveness" of a rough surface can be examined by investigating the dependency of the magnitudes of plane wave reflection coefficient with the angle of incident at various azimuthal angles for the reflected waves. Figures 3(a)–3(d) show variations of $|R_p|$ with α for different azimuthal angles (0°, 15°, 30°, 45°, 60°, 75°, and 90°, respectively) with the radius of the semicylindrical bosses a of 0.08 m for cases (a) and (b) and of 0.04 m for cases (c) and (d). The average center-to-center spacing between semicylindrical bosses b varies from $4a$ to $5a$. In Figs. 3(a) and 3(c), the predicted $|R_p|$ are shown for the minimum center-to-center spacing between two adjacent semicylindrical bosses of $3.5a$ (dotted line) and of $3a$ (solid line). In Figs. 3(a) and 3(b), the predicted $|R_p|$ are shown for the minimum center-to-center spacing between two adjacent semicylindrical bosses of $4.5a$ (dotted line) and of $4a$ (solid line).

The source frequency used in the graphs is 500 Hz which means that the plots with $ka=0.74$ are shown in Figs. 3(a) and 3(b) and with $ka=0.37$ are shown in Figs. 3(c) and 3(d).

The magnitude of the reflection coefficient is equal to 1 when the reflected wave is propagated at the grazing angle, $\alpha=\pi/2$. The magnitude of the reflection coefficient is also

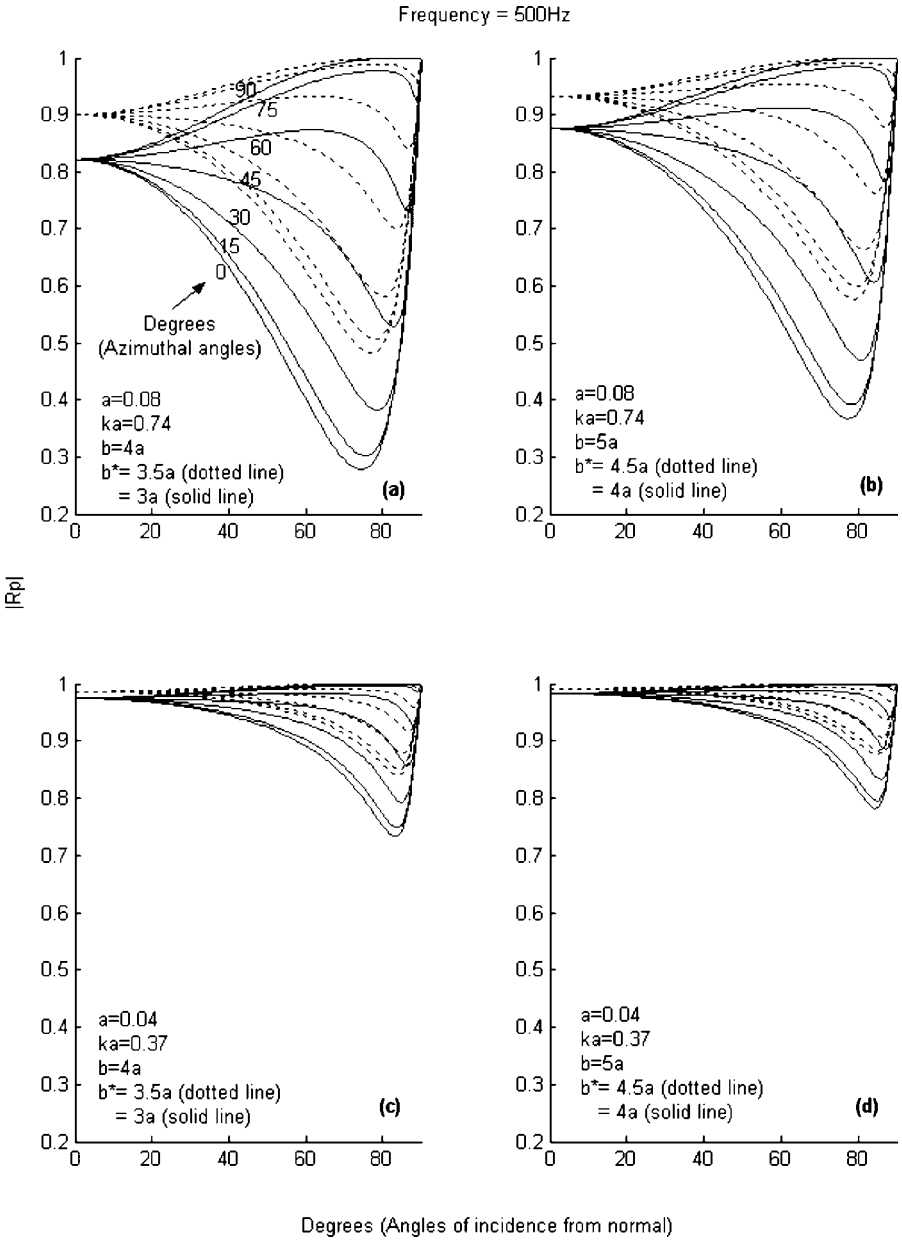


FIG. 3. Predicted variations of $|R_p|$ with the angle of normal incidence, the azimuthal angles of the incident wave, and different roughness parameters.

independent of the azimuthal angle when the reflected wave approaches the rough surface at the normal incidence, i.e., $\alpha=0$. Substituting $\alpha=0$ into Eqs. (1)–(3) and, in turn, into Eq. (6b), $|R_p|$ can be determined as follows:

$$|R_p| = \frac{\sqrt{[1 - mk^3\pi^2a^4(1 - W^2)/8]^2 + [kV]^2}}{\sqrt{[1 + mk^3\pi^2a^4(1 - W^2)/8]^2 + [kV]^2}}. \quad (9)$$

From Fig. 3(a)–3(d), we see that $|R_p|$ decreases when α increases from 0 until it reaches a minimum value before it increases to 1 when $\alpha=\pi/2$. We can see from Fig. 3(a) that $|R_p|$ has a value of as low as about 0.3 for a azimuthal angle of 0° and an incident angle of around 75°.

Comparison of predictions using different roughnesses with $a=0.08$ m in Fig. 3(a) and with $a=0.004$ m in Fig. 3(c) show that more absorption is expected for the surface with higher value of a, i.e., surface with more roughness. In addition, Figs. 3(a) and 3(b) compare $|R_p|$ with two average center-to-center distances ($b=4a$ and $5a$). It can be seen from these two plots that the closer the separations between two adjacent bosses, the larger the predicted effect of the modified admittance, i.e., more apparent "absorption." The degree of nonperiodicity distributions of semicylindrical bosses leads to a larger absorptive effect due to the modified admittance. This effect is shown in all plots of Fig. 3. We can see that use of surface roughness can lead to a certain degree of absorption when the sound waves are incident on a rough surface. Hence, we expect that the deployment of rough surfaces can lead to the reduction in the total sound field in long enclosures.

III. MODEL VALIDATIONS

It is remarkable that an incoherent scattering effect is expected by distributing semicylindrical rods nonperiodically [i.e., $W \neq 1$ in Eqs. (2) and (4)]. These semicylindrical rods were used to simulate a two-dimensional (2D) hard rough surface. Noting the approximation of small ka, the size of

roughness chosen for the current experimental means that the theoretical model for a rough surface should give valid predictions for frequencies up to about 4 kHz. A series of measurements with different source locations were conducted to confirm the Twersky model although extensive studies have been carried out elsewhere.[23]

Measurements were conducted at different receiver locations over a rough surface in the anechoic chamber with size of $6\times6\times4$ m^3 (high). The measured data were obtained to validate the Twersky boss model. The experimental setup is shown in Fig. 3. The rough surface is made of wooden semicylindrical rods embedded on a smooth wooden board. The radius of each semicylindrical rod is 0.015 m, which was fixed onto a 1.25-cm-thick plywood board with dimensions of 2.4×4.8 m^2. Prior measurements were conducted to measure the acoustic characteristic of the plywood board. We found that the plywood board could generally be treated as a perfectly reflecting plane. The semicylindrical rods were assumed to be acoustically hard because they were also made by the same material as the plywood board. The semicylindrical rods and the plywood board were varnished to prevent the leakage of sound. The semicylindrical rods were spaced pseudorandomly with an average distance and a minimum center-to-center separation between two adjacent semicylinders of 0.06 and 0.0525 m, respectively.

A Tannoy speaker, which was fitted on a long brass pipe of 1 m long and 25 mm in diameter, was used to simulate an omnidirectional point source. Preliminary measurements were conducted to examine the directional characteristic of the point source. The measured result, not shown here for brevity, suggests that the deviation in the directivity pattern for all directions is within 1 dB for all frequencies above 250 Hz.

For the measurements, the point source and receiver were located at various positions over the rough surface. A Bruel & Kjaer type 4942 microphone (prepolarized, diffuse field), fitted with a Bruel & Kjaer type 2671 preamplifier and a Bruel & Kjaer NEXUS conditional amplifier, was used as a receiver. The microphone was placed at various heights above the floor by mounting it on an adjustable stand. A special type of test signal called a maximum length sequence (MLS) was employed to obtain the experimental data. The deterministic nature of the MLS provides an excellent signal-to-noise ratio, which was ideal for the current indoor experiments. The MLS signals were generated by the maximum length sequence signal analyzer (MLSSA) 2000 card, transferred via the built-in digital-to-analog-converter and boosted by a Bruel & Kjaer 2713 power amplifier. The MLS signals were then connected to the Tannoy speaker, which emitted sound for experiments. The measurements were recorded in the time domain. Manipulations of time-domain measured signals were possible for eliminating unwanted reflections.

We introduce a term known as the excess attenuation (EA) to facilitate the presentation of the measured results. EA is the difference in sound pressure levels with and without the presence of the rough surface, i.e.,

$$\mathrm{EA} = 20\log_{10}(P_{\mathrm{tot}}/P_{\mathrm{ff}}), \tag{10}$$

where P_{tot} is the total sound field above the rough surface and P_{ff} is the corresponding free field sound pressure at the same receiver location without the presence of any reflecting surface.

The results predicted by the Twersky boss model are subsequently compared with that evaluated by a wave-based numerical formulation. As the geometrical configuration of the current situation is an external problem, the boundary element method[28] (BEM) is an appropriate approach for the purpose of validation for both indoor measurements and the boss model. The BEM formulation has been extensively used to study the physical phenomenon of outdoor sound propagation in an irregular terrain. In the BEM study, the rough surface is partitioned with at least ten elements per wavelength, and each individual semicylindrical boss was represented by seven elements. In the BEM formulation, a FORTRAN program was used to solve the set of simultaneous equations by using the standard matrix method. The computational time for the BEM formulation will increase exponentially for higher source frequency and for a larger source/receiver separation. On the other hand, a simple MATLAB program was developed for the Twersky's boss model which was then used to predict the propagation of sound above the hard rough surface.

Three receiver locations are chosen for comparison. Both source and receiver are located 0.2 m above the bottom of rough surface and with horizontal distances of (a) 1 m, (b) 2 m, and (c) 3 m. In each plot [Figs. 4(a)–4(c)], comparisons of the measurements (dotted lines with symbols) and predictions using the BEM formulation (dashed lines) and the image source model (solid lines) are shown. The effect of atmospheric absorption is ignored in all the numerical predictions.

Comparing the smooth and rough reflecting planes, Figs. 4(a)–4(c) show that there are significant shifts of main ground effect dips to the low frequency for a rough surface. From the measured data and the numerical predictions, it is shown that the plywood board can be treated essentially as a rigid reflecting plane. For the hard rough cases, predictions according to Twersky's model show reasonably good agreement with the measured data and with the predictions using BEM formulation at various locations. It is reassuring to confirm that Twersky's boss model can provide an accurate prediction of the scattering effects due to a hard rough surface.

A model rough surface, which was now built for the next experimental study, was placed either in an indoor enclosure or in a larger outdoor-modeled tunnel. The measured data were obtained to validate the image source model for the propagation of sound over a rough surface in a tunnel.

For the indoor case, we conducted a series of measurements with the use of an enclosure of 4.8 m long which was placed in the anechoic chamber. The enclosure was

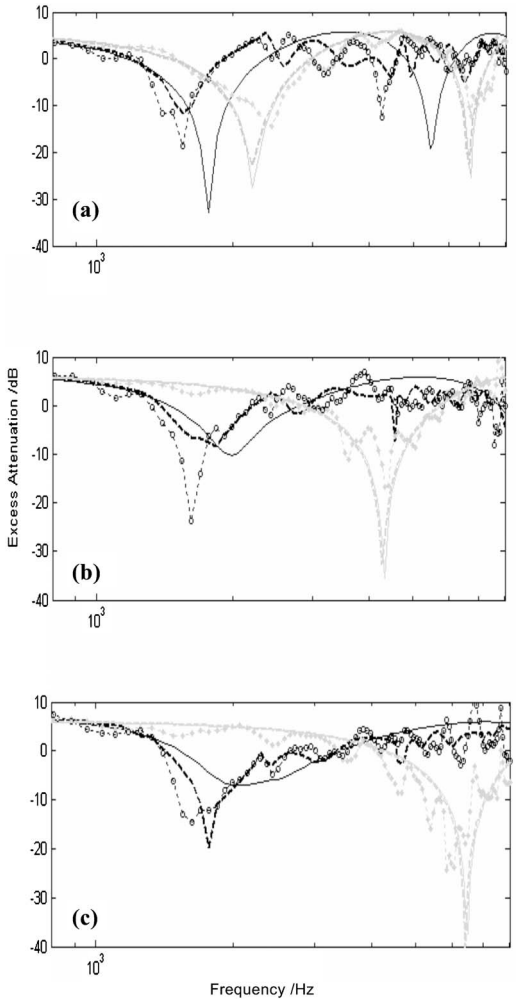


FIG. 4. Comparisons of measured (symbols) and predicted (lines) excess attenuation spectra from a point source over a plywood board with and without semicylindrical bosses. Both source and receiver are located 0.2 m above the plane and with horizontal distances of (a) 1 m, (b) 2 m, and (c) 3 m. Dotted line with circles, measurements (rough), dotted line with stars, measurements (smooth); dashed lines (black), BEM predictions (rough); dashed lines (gray), BEM predictions (smooth); solid lines (black), Twersky's boss model; and solid lines (gray), coherent image source method.

constructed with a rectangular cross section of 1 m wide and 1.2 m high. The four boundary surfaces of the enclosure were made of plywood with flat surfaces. To model the scattering effect due to the roughness of the enclosure's surface, the entire length of the left vertical wall was installed with semicylindrical wooden rods of 0.015 m radius. These semicylindrical rods were distributed pseudorandomly with a minimum center-to-center separation of 0.0525 m. This arrangement ensures that the assumption of small *ka* can be met that, in turn, allows valid numerical predictions for frequencies up to about 4 kHz. The experimental setup is shown in Fig. 5.

Measurements of sound propagation along the enclosure using the same set of measuring equipment are described earlier. However, we present the experimental data in term of the relative sound pressure level (RSL), which is defined as

FIG. 5. An experimental setup in an anechoic chamber for measurement of sound propagation in a long enclosure with a roughening surface on the left vertical wall. Semicylindrical wooden rods of 0.015 m radius were taped securely on the wooden board. The rods were distributed pseudorandomly with a minimum center-to-center separation between two adjacent semicylinders b^* of 0.0525 m.

$$\mathrm{RSL} = 20 \log_{10} \left| \frac{P_{\mathrm{tot}}}{P_{\mathrm{ref}}} \right|, \tag{11}$$

where P_{ref} is the reference free field sound pressure measured at a horizontal distance of 1 m in front of the source. It is more convenient to use the relative sound pressure level in favor of the excess attenuation because it is generally more difficult to obtain free field measured level for the receiver placed over 2 m from the source.

Both the source and receiver were located along the center line of the enclosure and 0.6 m above the floor in the first set of measurements. They were then placed at two offset points: 0.75 and 0.25 m from left vertical wall (installed with semicylindrical rods) with the respective heights of 0.3 and 0.6 m. The experimental data were taken at various horizontal distances in front of the source. The setup is shown in Fig. 5. A separate measurement of the reference sound pressure level was carried out at a horizontal distance of 1 m from the source in the anechoic chamber without the presence of any reflecting surfaces. The background noise level was sufficiently low to provide a reference signal for use in subsequent experimental measurements.

In the computations, as can be seen in Fig. 2, the determination of image source location is followed by using the image source model according to Li and Iu.[15] For the case of hard surface, it can be seen from Eq. (5) that the rectangular enclosure contains an infinite number of image sources due to the presences of ceiling, floor, and two wall surfaces on the left and right sides. Preliminary numerical analyses were conducted. It is found that the series given in Eq. (5) normally converge for about 50 terms. The number of terms N

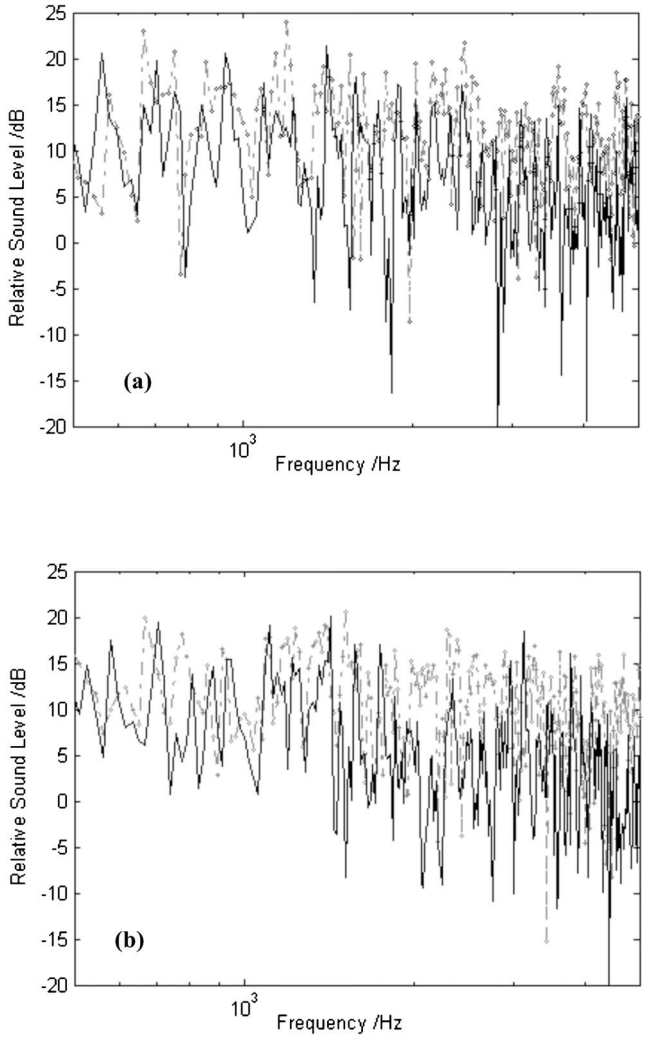


FIG. 6. Comparison of measured (solid line) and predicted (dashed-dotted line with diamonds) relative sound pressure level in a scale model enclosure built in an anechoic chamber with hard boundaries. Semicylindrical rods (a=0.015 m, length=1.2 m, b=0.06 m, b^*=0.0525 m) were taped on to the left vertical surface to form a 2D hard roughness surface. The source and receiver were both located at the center line (0.5 m from either vertical walls) and their heights were 0.6 m above the floor. The horizontal separations between source and receiver were (a) 1 m, and (b) 2 m.

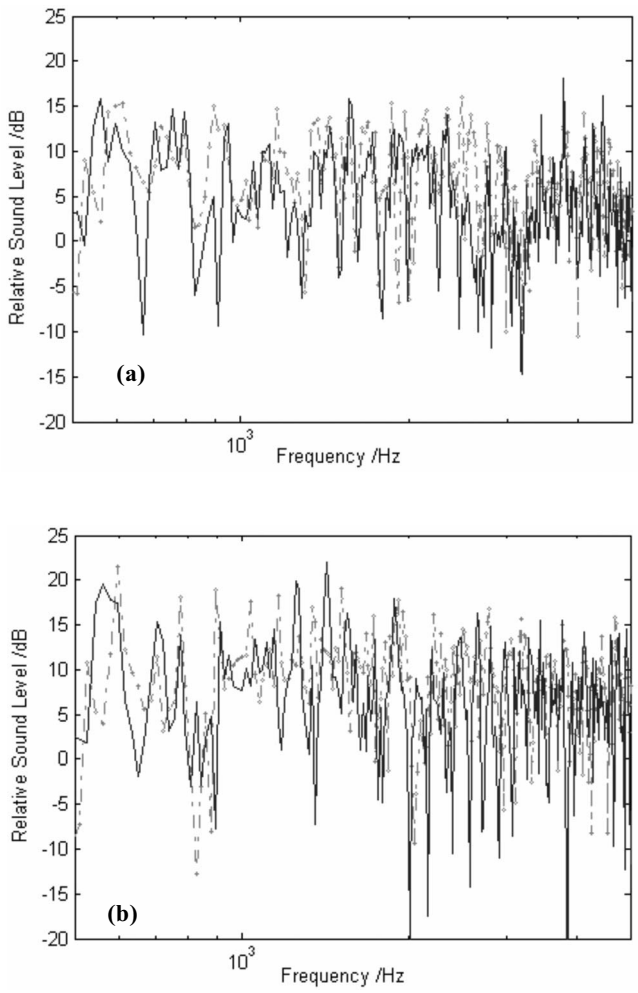


FIG. 7. Same as Fig. 6, except that the source was located at an offset position (0.75 m from the left vertical wall) and the height at 0.3 m. The receiver was also located at an offset position (0.25 m from the left vertical wall) and the height at 0.6 m. The same horizontal separations between source and receiver were used with (a) 1 m and (b) 2 m.

required to ensure the convergence of the ray series depends on the geometrical configuration and the effective impedance of the tunnel surfaces. We do not attempt to optimize the choice of N but simply choose the variation of the magnitude of the reflected waves for two consecutive terms is less than 1%. This condition has proven to be adequate for the image source model to give satisfactory numerical results.

In Figs. 6 and 7, the predicted RSL are shown in which the Twersky model is used in conjunction with Eq. (5) for the numerical models. Both measured and predicted results show a regular pattern of "dips" and "peaks" in the frequency spectra. Both spectra fluctuate considerably as the frequency increases. This phenomenon corresponds to destructive and constructive interferences of the contributory rays. Although there are noticeable discrepancies in the magnitudes of peaks and dips between the measured and predicted RSL, the trends in the frequency spectra agree reasonably well with each other. The interference pattern of the RSL spectra suggests that the phase information of each ray plays an important role in predicting the total sound field in the enclosures. As mentioned in our early studies,[15,16] the energy-based incoherent model will not be adequate to predict the sound fields in the tunnels.

We further remark that the locations of dips and peaks in the spectrum are sensitive to the small change in the source/receiver geometry especially at high frequencies. According to Fig. 6(b), there are several distinct peaks predicted above 2000 Hz but these peaks are not matched with the measured results. Indeed, the patterns of fluctuations in the frequency spectra are quite different for the measured and predicted results. These apparent discrepancies are largely due to the imprecise location of the receiver relative to the source in the tunnel. A small error in the position can lead to predictions of a significant phase shift for the contribution from a ray. This is particularly noticeable for high frequencies and large distances as was shown in comparable results of our previous studies in long enclosures[15–17].

Next, we illustrate the effect of a hard rough surface in tunnels by comparing the sound fields for a hard "smooth" tunnel. The insertion loss (IL) is used in the following presentation, where

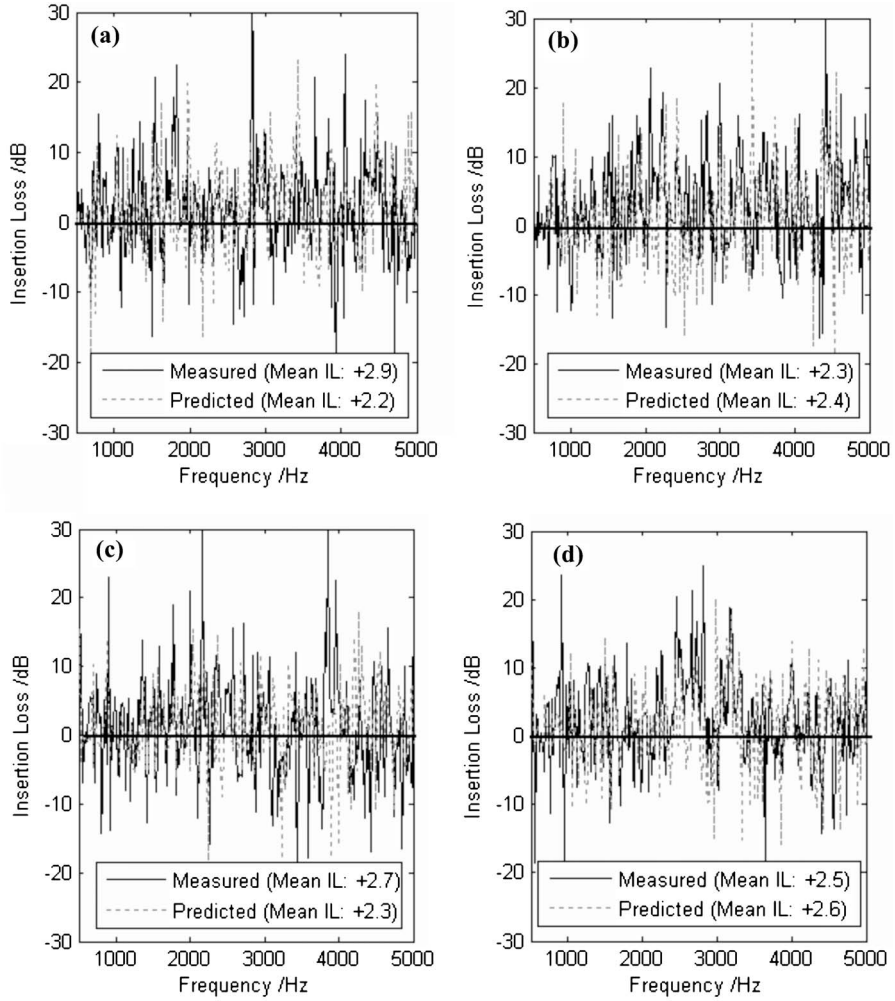


FIG. 8. The insertion loss (IL) spectra at different source/receiver locations with left vertical surface having 2D hard semicylindrical rods in the scale model enclosure over surface with smooth hard condition. The dotted line represents predictions by the complex image source model. The solid line represents results from experimental measurement: [(a) and (b)] source and receiver locations correspond to Figs. 6(a) and 6(b); [(c) and (d)] source and receiver locations correspond to Figs. 7(a) and 7(b), respectively. The mean values of measured and predicted IL over the frequency spectrum are also indicated.

$$\mathrm{IL} = 20 \log_{10}\left(\left| \frac{p_{\mathrm{w/o}}}{p_{\mathrm{w}}} \right| \right), \tag{12}$$

p_{w} and $p_{\mathrm{w/o}}$ are the total sound fields with or without the rough surface in the tunnel, respectively. Figure 8 shows the comparisons of the predicted ILs with experimental data for four receiver locations. Both the predicted and measured ILs fluctuate considerably over the frequency range with consistent agreements on the peak and dip locations. The mean value of measured IL ranges between 2.3 and 2.9 dB. The mean value of the predicted IL ranges between 2.2 and 2.6 dB. Introducing a hard rough surface on one of the vertical walls of the scale model tunnel can lead to an average reduction in the noise levels of about 3 dB over the frequency range from 500 to 5000 Hz.

Another series of measurements have been conducted in a larger model tunnel with a scale of 1:10. The tunnel, which was placed outdoors, was made of gypsum boards with an internal cross sectional area of $1.16 \times 1.46\ \mathrm{m}^2$. The tunnel was 27 m long with an anechoic termination constructed at one end. The MLSSA measurement system was again used. A Renkus-Heinz PN61 loudspeaker was used as a point source in this case to provide a higher sound power. Plywood boards were covered along the left vertical wall of the model tunnel. Varnished wooden rods of semicylindrical shape of 1.4 m long with radius 0.015 m were taped securely onto these plywood boards to form a 2D hard rough surface.

The semicylindrical rods were pseudorandomly spaced and distributed vertically on the left side wall with a minimum center-to-center separation of 0.04 m. To investigate the spatial variations of the sound field, two sets of source/receiver geometries were selected for measurements. For the first set of measurements, the source was located along the center line of the outdoor model tunnel. It was then placed at an offset position of 0.86 m from the left vertical wall for the second set of measurements. Two separate source heights of 0.8 and 0.4 m above the floor were chosen for these two sets of experimental measurements. The receivers were located at different positions along the center line with the horizontal separations of 2, 5, and 10 m from the source.

Since the gypsum boards were used for the construction of this outdoor model tunnel walls, the assumption of perfectly reflecting surfaces led to inaccurate predictions of the total sound fields in the enclosure. Hence, the characterization of the acoustical impedance of the gypsum boards was essential because of contributions from the reflected sound fields are one of the major components for the present situation. An impedance deduction method[26] was used to determine the acoustical characteristics of the gypsum boards. It was also found that a two-parameter impedance model[27] was

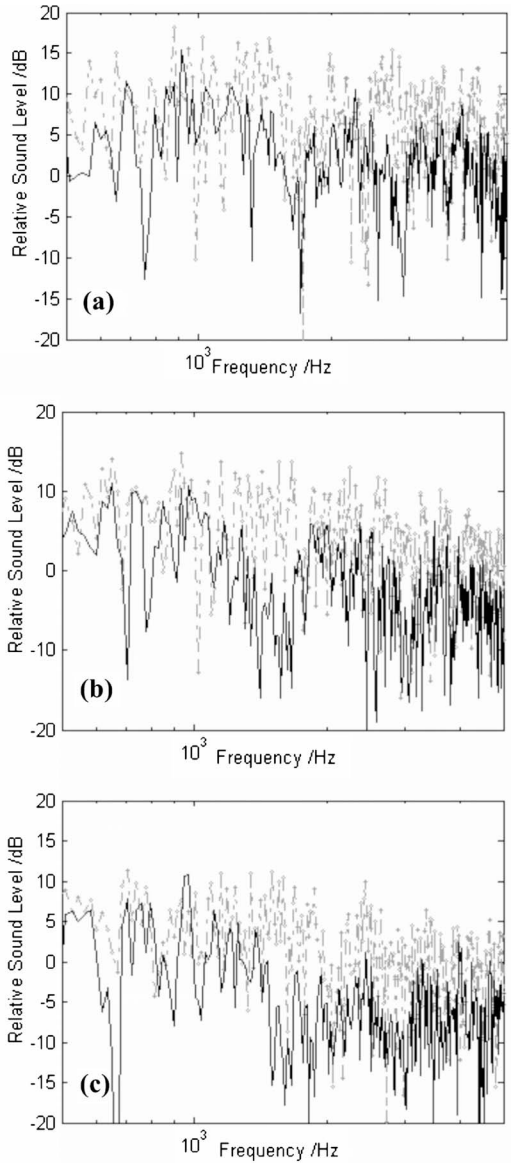


FIG. 9. (Color online) Comparison of measured (solid line) and predicted (dashed-dotted line with diamonds) relative sound pressure level in the outdoor model tunnel. Semicylindrical rods (a=0.015 m, length=1.4 m, b =0.045 m, b^*=0.04 m) were taped securely on the left vertical wall of the tunnel to form a 2D hard roughness surface. The source and receiver were both located at the center line (0.58 m from either wall) and both heights were 0.8 m. The horizontal separations between the source and receiver were (a) 2 m, (b) 5 m, and (c) 10 m.

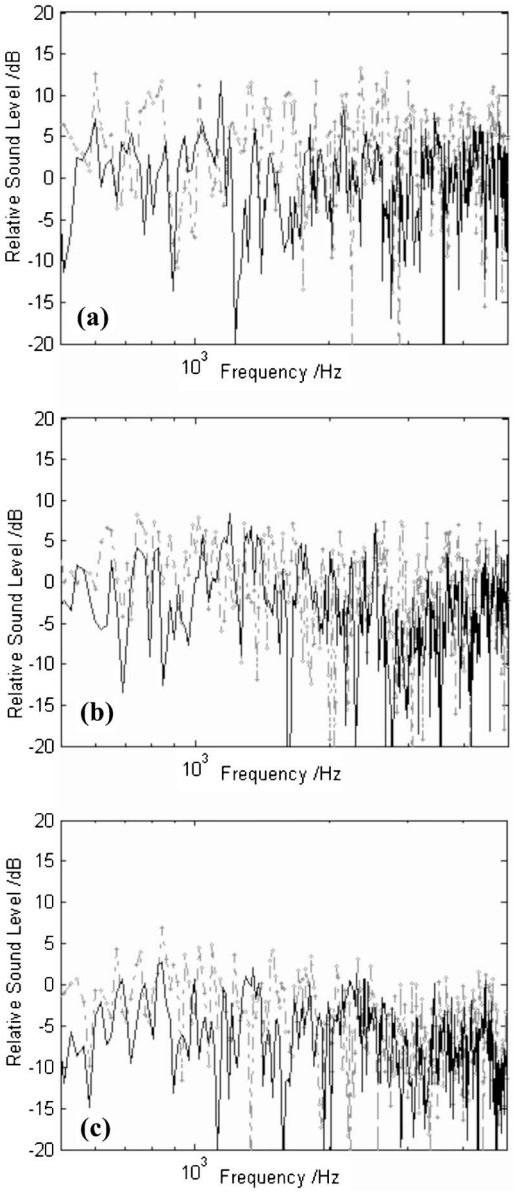


FIG. 10. Same as Fig. 9, except the source and receiver were located at the offset position (0.86 m from the left vertical wall) and the centre line, respectively. The height of the source was 0.4 m and it was 0.8 m above the floor for the receiver. The horizontal separations between the source and receiver were the same at (a) 2 m, (b) 5 m, and (c) 10 m.

sufficiently accurate to represent its acoustical impedance. The effective flow resistivity of 6000 kPa s m^{-2} and the effective rate of change in porosity with depth of 100 m^{-1} were deduced from the measurements in the present study. These parametric values are then used in the subsequent prediction of the sound field in the model tunnel.

Next, we show comparisons of the frequency spectra of the measured and predicted RSL in the model tunnel. Figures 9 and 10 display the measured results for the source and receiver located at the center and offset positions respectively. A clear pattern of dips and peaks, which are similar to the earlier indoor measurements, is evinced in these two plots. In comparing both measured and predicted results, it can be seen that general agreements between measurements and predictions at different receiver positions are obtained.

However, as shown in Fig. 9, the patterns of the predicted and measured frequency spectra show some notable discrepancies when the source and receiver were located along the center line of the model tunnel. Again, these apparent discrepancies are possibly due to the phase errors of contributory rays as discussed earlier for the results of the indoor scale model enclosure. On the other hand, when the source and receiver were located at the offset positions (see Fig. 10), the predicted spectra show better agreement with the measured results.

We remark that the measured and predicted results for a tunnel with smooth hard walls were presented elsewhere.[15] Their results showed that the complex image source method agrees reasonably well with experimental data. Hence, in

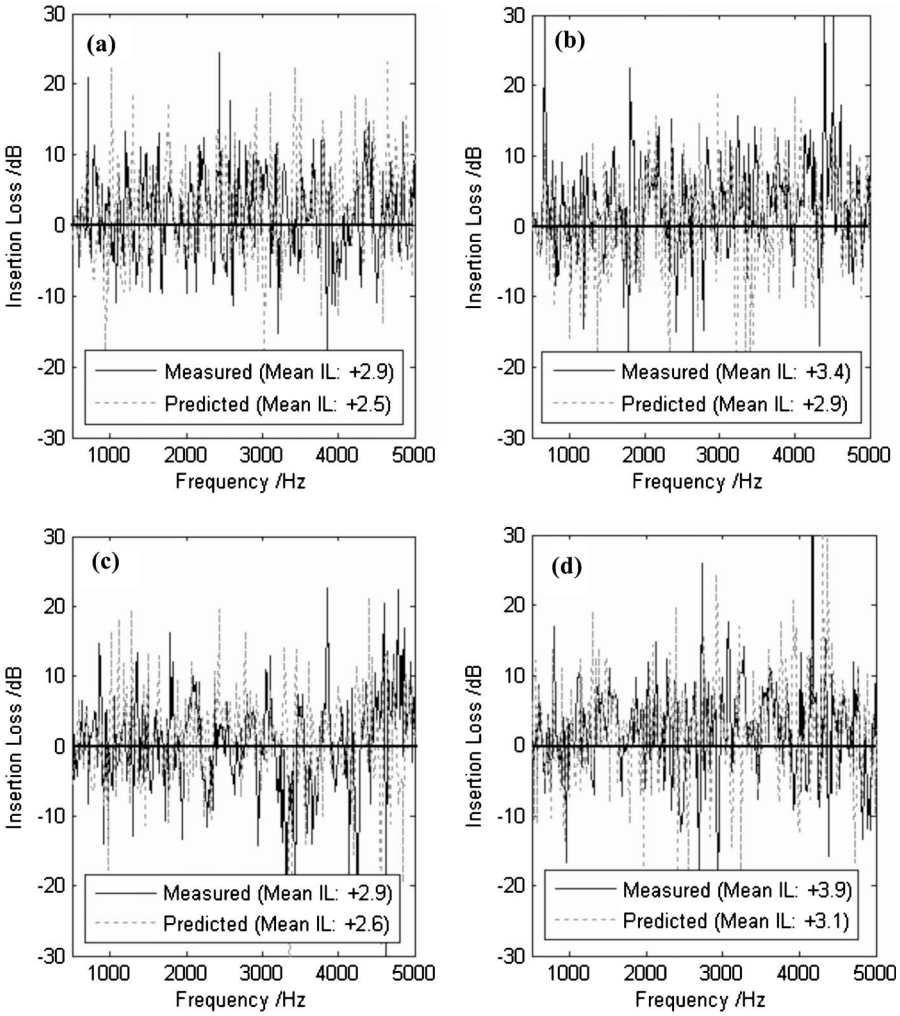


FIG. 11. The insertion loss (IL) spectra at different source/receiver locations with left vertical surface having 2D hard semicylindrical rods in the outdoor model tunnel over surface with smooth hard condition. The dotted line represents predictions by the complex image source model. The solid line represents results from experimental measurement: [(a) and (b)] source and receiver locations correspond to Figs. 9(a) and 9(b); [(c) and (d)] source and receiver locations correspond to Figs. 10(a) and 10(b), respectively. The mean values of measured and predicted IL over the frequency spectrum are also indicated.

Fig. 11, we only compare the predicted insertion loss (IL) with experimental data for four receiver locations. The predicted IL and measured IL, which agree reasonably well for the peak and dip locations, fluctuate considerably over the frequency range of interest. The average value of measured IL ranges between 2.9 and 3.9 dB and it is between 2.5 and 3.1 dB for the predicted IL.

With the measured data at the two model tunnels, it is possible to conclude that the application of the Twersky model for a rough surface in the coherent model is capable to predict the influence of surface roughness on the propagation of sound in long enclosures. Introducing a hard rough surface on one of the vertical walls of the model tunnels can lead to an average reduction in the noise levels of about 3 dB over the frequency range from 500 to 5000 Hz.

IV. APPLICATION—NOISE REDUCTION IN LONG ENCLOSURES BY ROUGH SURFACES

We have validated experimentally in an indoor enclosure and a model tunnel that effective impedance can be introduced on the wall by embedding arrays of semicylindrical bosses on an otherwise smooth and hard boundary surface. The coherent model can be used to predict the propagation of sound in tunnels. It is apparent that the surface roughness can lead to the reduction in the reverberant sound fields by introducing effective impedance on the tunnel walls. This implies that roughening of an otherwise acoustically hard surface may help to provide a passive noise control in tunnel environments.

In this section, we consider an example for noise reduction in a tunnel due to multiple sources. We endeavor to simulate the situations of using a hard rough surface to reduce the overall noise levels in the tunnel. The case chosen in this study emulates a common urban situation where there are many vehicular sources in a tunnel for road traffic. It is noted that the multiple sources were considered in this simulation instead of a simple line source. It is because a relatively large separation distance between adjacent vehicles is expected in most practical situations.

Here, the total sound pressure level L_T due to a total of N_s vehicular sources can be determined straightforwardly as follows:

$$L_T = 10 \log_{10} \sum_{j=1}^{N_S} 10^{\mathrm{SPL}_j/10}, \qquad (12')$$

where SPL_j is the contribution from the jth image source. We remark that the total sound pressure level at a receiver point is obtained by first summing the contributions from all images for each source coherently. The overall L_T is then calculated by adding the contributions from all individual sources incoherently.

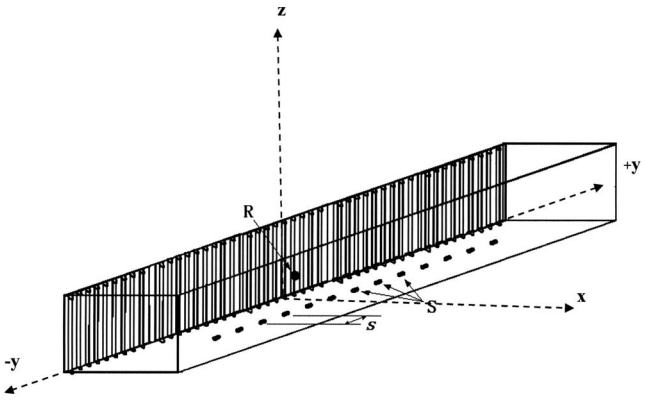


FIG. 12. A model road traffic tunnel use in the numerical simulations.

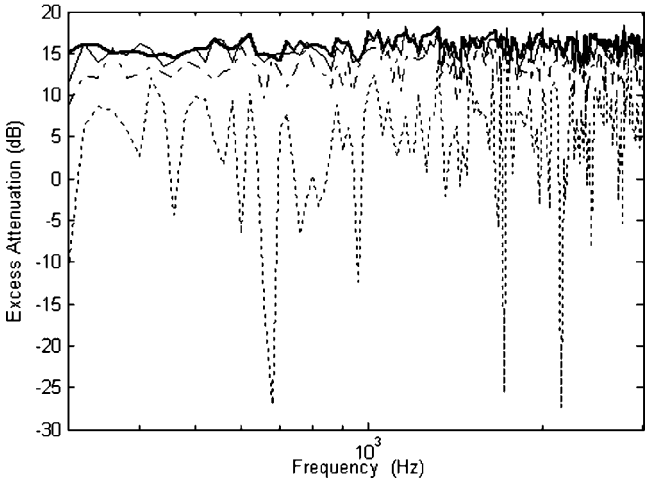


FIG. 13. Predicted excess attenuation in a road traffic tunnel with hard surfaces. (Thick solid line, $s=5$ m; thin solid line, $s=10$ m; dash-dotted line, $s=20$ m; and dotted line, single source).

The objective of this example is to quantify the sound amplification due to the effect of multiple reflections and to assess the sound reduction by deploying a hard rough surface on one side of the tunnel walls. Modeling this situation, we assume that the source and receiver are located inside an infinitely long tunnel with a constant cross-sectional area of rectangular in shape. To facilitate the numerical analysis, a rectangular coordinate system is used, where the origin is located at the bottom left corner of the tunnel (see Fig. 12). The cross section of the tunnel has a dimension of 8 m wide (x direction) and 10 m high (z direction). The tunnel is extended to infinity at both directions along the y axis and the receiver is located at the plane of $y=0$. Suppose a number of concrete blocks of semicylindrical shape with radius of 0.075 m were constructed to form a 2D hard rough surface on one side of the tunnel walls. The concrete blocks are spaced and distributed on the left vertical wall with average separation of 0.3 m. The minimum center-to-center separation between two adjacent semicylinders is 0.2 m. We treat all other boundaries, i.e., the ground, the ceiling, and the right side vertical wall, as perfectly reflecting surfaces. This assumption is justifiable because most surfaces in a tunnel environment are acoustically hard.

The cases of a single source and multiple sources are considered in the following numerical simulations with the receiver placed at (2, 0, 4). For the first simulation, a single source is located at (4, 0, 0.5). For the second case, a series of multiple sources are placed at a height of 0.5 m above the ground and 4 m along the center line of the tunnel. Three separations between two adjacent sources (labeled as s in Fig. 12) are considered with the respective distances of 5, 10, and 20 m along the y axis. In our preliminary analysis, it is found that contributions from any sources which are located at a horizontal distance more than 100 m from the receiver are negligible when their contributed noise levels are compared with those due to the nearer sources.

Figure 13 displays numerical predictions of the sound pressure levels due to the single source and multiple sources. In the plots, the excess attenuation EA is used to present the predicted results where it is defined as the difference in the total sound pressure levels at a receiver point with (L_T) and without ($\bar{L}_T$) the presence of the tunnel, i.e.,

$$\mathrm{EA} = L_T - \bar{L}_T. \tag{13}$$

The predicted EA spectra for the two cases are presented in Fig. 13 for the source frequency ranging from 300 to 3000 Hz. For the case of a single noise source, it can be seen that the overall average sound pressure level is about 8 dB higher than the average free field sound pressure levels. For the case of multiple noise sources with 5 m separation, the overall average sound pressure level increase to about 16 dB above the average free field sound pressure level due to the multiple noise sources. The predicted results for other two separations (10 and 20 m) are rather similar to the results of the 5 m separation with the average EA of about 15 and 13 dB, respectively. For a single source, the predicted results confirm that there is a significant amplification of the overall sound pressure levels because of the effect of multiple reflections of the tunnel walls. The effect of multiple reflections becomes even more acute when there are multiple sources in the tunnel.

Finally, we show that the overall sound pressure levels can be reduced by deploying a 2D hard rough surface on one side of vertical wall inside the road tunnel. This is illustrated in the following numerical simulation. For the case of multiple sources with 5 m separation in the tunnel, Fig. 14 displays the predicted excess attenuation in the tunnel with and without the use of hard rough surface on one side of vertical wall inside the traffic tunnel. We can see that there is an average of about 12 dB above the free field sound pressure levels. This represents a reduction of 4 dB for the case without the presence of the rough surface. For the source and receivers at other locations, calculations show comparable levels of noise reduction. These numerical simulations are not shown here for brevity.

V. CONCLUSIONS

The Twersky boss formulation has been applied to the image source model for the calculation of the sound pressure levels in a rigid tunnel with the presence of a 2D hard rough surface on one of its vertical wall. The proposed theoretical model was validated by comparing with experimental measurements conducted in two model tunnels. The size of the

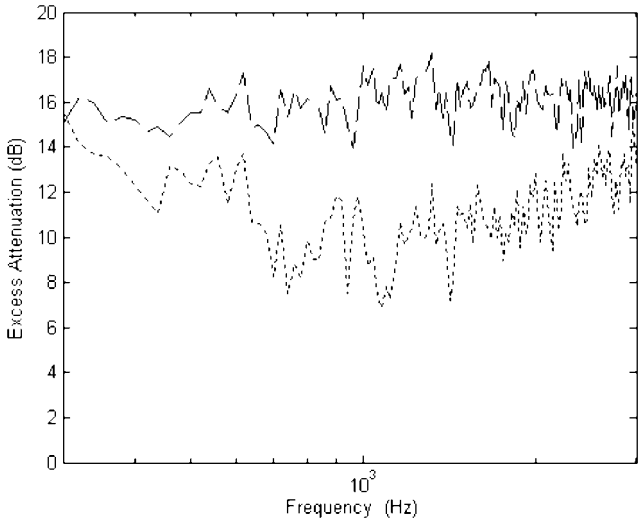


FIG. 14. Comparison of excess attenuation spectra in a road traffic tunnel with and without a hard rough surface: Dashes line, without roughness; dotted line, with roughness.

first model enclosure, which was placed in an anechoic chamber, was 1 m wide, 1.2 m high, and 4.8 meter long. The second tunnel was built with a scale of 1:10, which had a dimension of 1.16 m wide, 1.46 m high, and 27 m long. Comparing with experimental data, it was demonstrated that the presence of a hard rough surface on one of the vertical walls in the tunnel introduced effective impedance on the boundary. The presence of apparent impedance on the boundary surfaces led to the reduction in the average noise levels in an otherwise hard tunnel. In the scale model studies, it is found that the introduction of a hard rough surface on one of the vertical walls of the model tunnels can lead to an average reduction in the noise level of about 3 dB over the frequency range from 500 to 5000 Hz.

Numerical simulations have also been conducted for a traffic tunnel of a realistic size with a cross-sectional area of 6×8 m^2. Numerical simulations show that the sound pressure levels in the tunnel due a single noise source is about 8 dB above the free field sound pressure levels. For multiple noise sources operated in a tunnel, the sound pressure levels are predicted to be 16 dB higher than that of the free field sound pressure levels. The introduction of a hard rough surface on one of the vertical walls can lead to an average reduction in the noise level of 4 dB over the frequency range from 300 to 3000 Hz.

ACKNOWLEDGMENTS

The research described in this paper was financed jointly by the Innovation and Technology Commission of the Hong Kong Special Administrative Region and the Mass Transit Railway Corporation Limited, through the award of an Innovation and Technology Fund grant under the category of the University-Industry Collaboration Programme (Project No. UIM/39). The authors are grateful to William Fung, Chenly Lai, P M Lam, S T So, and T L Yip for their help in conducting the experiments. Part of the manuscript was prepared while one of the authors (M.K.L.) was a visiting scholar at Ray W. Herrick Laboratories, Purdue University.

[1] J. Kang, *Acoustics of Long Spaces: Theory Design and Practice* (Thomas Telford, London, 2002).
[2] J. Kang, "Modelling of train noise in underground stations," J. Sound Vib. **195**, 241–255 (1996).
[3] L. Yang and B. M. Shield, "The prediction of speech intelligibility in underground stations of rectangular cross section," J. Acoust. Soc. Am. **109**, 266–273 (2001).
[4] T. L. Redmore, "A theoretical analysis and experimental study of the behaviour of sound in corridors," Appl. Acoust. **15**, 161–170 (1982).
[5] J. Kang, "Acoustics in long enclosures with multiple sources," J. Acoust. Soc. Am. **99**, 985–989 (1996).
[6] S. J. van Wijngaarden and J. A. Verhave, "Prediction of speech intelligibility for public address systems in traffic tunnels," Appl. Acoust. **67**, 306–323 (2006).
[7] Research Committee of Road Traffic Noise in Acoustical Society of Japan, "ASJ prediction model 1998 for road traffic noise," J. Acoust. Soc. Jpn. **55**, 281–324 (1999).
[8] H. Imaizumi, S. Kunimatsu, and T. Isei, "Sound propagation and speech transmission in a branching underground tunnel," J. Acoust. Soc. Am. **108**, 632–642 (2000).
[9] J. Kang, "Reverberation in rectangular long enclosures with geometrically reflecting boundaries," Acust. Acta Acust. **82**, 509–516 (1996).
[10] M. V. Sergeev, "Scattered sound and reverberation on city streets and in tunnels," Sov. Phys. Acoust. **25**, 248–252 (1979).
[11] R. N. Miles, "Sound field in a rectangular enclosure with diffusely reflecting boundaries," J. Sound Vib. **92**, 203–226 (1984).
[12] J. Kang, "Reverberation in rectangular long enclosures with diffusely reflecting boundaries," Acust. Acta Acust. **88**, 77–87 (2002).
[13] K. K. Iu and K. M. Li, "The propagation of sound in street canyons," J. Acoust. Soc. Am. **112**, 537–550 (2002).
[14] J. Picaut, T. Le Pillès, P. L'Hermite, and V. Gary, "Experimental study of sound propagation in a street," Appl. Acoust. **66**, 149–173 (2005).
[15] K. M. Li and K. K. Iu, "Propagation of sound in long enclosures," J. Acoust. Soc. Am. **116**, 2759–2770 (2004).
[16] K. M. Li and P. M. Lam, "Prediction of reverberation time and speech transmission index in long enclosures," J. Acoust. Soc. Am. **117**, 3716–3726 (2005).
[17] P. M. Lam and K. M. Li, "A coherent model for predicting noise reduction in long enclosures with impedance discontinuities," J. Sound Vib. **209**, 559–574 (2007).
[18] V. Twersky, "Reflection and scattering of sound by correlated rough surfaces," J. Acoust. Soc. Am. **73**, 85–94 (1983).
[19] J. P. Chambers and Y. Berthelot, "Utilizing a modified impedance analogy on sound propagation past a hard, curved, rough surface," J. Acoust. Soc. Am. **1120**, 1186–1189 (2006).
[20] I. Tolstoy, "Coherent sound scatter from a rough interface between arbitrary fluids with particular reference to roughness element shapes and corrugated surfaces," J. Acoust. Soc. Am. **72**, 960–972 (1982).
[21] R. J. Lucas and V. Twersky, "Coherent response to a point source irradiating a rough plane," J. Acoust. Soc. Am. **76**, 1847–1863 (1984).
[22] K. Attenborough and S. Taherzadeh, "Propagation from a point source over a rough finite impedance boundary," J. Acoust. Soc. Am. **98**, 1717–1722 (1995).
[23] P. Boulanger, K. Attenborough, S. Taherzadeh, T. Water-Fuller, and K. M. Li, "Ground effect over hard rough surfaces," J. Acoust. Soc. Am. **104**, 1474–1482 (1998).
[24] P. Boulanger, K. Attenborough, and Q. Qin, "Effective impedance of surfaces with porous roughness: Model and data," J. Acoust. Soc. Am. **117**, 1146–1156 (2005).
[25] C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).
[26] S. Taherzadeh and K. Attenborough, "Deduction of ground impedance from measurements of relative SPL spectra," J. Bone Jt. Surg., Am. Vol. **105**, 2039–2042 (1999).
[27] K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular materials," J. Bone Jt. Surg., Am. Vol. **73**, 785–799 (1983).
[28] S. N. Chandle-Wilde and D. C. Hothersall, "Efficient calculation of the green function for acoustic propagation above a homogeneous impedance plane," J. Sound Vib. **180**, 705–724 (1995).

Verifying the attenuation of earplugs in situ: Method validation using artificial head and numerical simulations

Annelies Bockstael[a)]
Department of Oto-Rhino-Laryngology and Department of Information Technology, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium

Bram de Greve, Timothy Van Renterghem, and Dick Botteldooren
Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

Wendy D'Haenens, Hannah Keppler, Leen Maes, Birgit Philips, Freya Swinnen, and Bart Vinck
Department of Oto-Rhino-Laryngology, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium

(Received 28 March 2008; accepted 27 May 2008)

The use of *in situ* measurements of hearing protectors' (HPD's) attenuation following the microphone in real ear (MIRE) protocol is increasing. The attenuation is hereby calculated from the difference in sound levels outside the ear and inside the ear canal behind the HPD. Custom-made earplugs have been designed with an inner bore that allows inserting a miniature microphone. A thorough understanding of the difference, henceforth called transfer function, between the sound pressure of interest at the eardrum and the one measured at the inner bore of the HPD is indispensable for optimizing the MIRE technique and extending its field of application. This issue was addressed by measurements on a head-and-torso-simulator and finite difference time domain numerical simulations of the outer ear canal occluded by an earplug. Both approaches are in good agreement and reveal a clear distinction between the sound pressure at the MIRE microphone and at eardrum, but the measured transfer functions appear to be stable and reproducible. Moreover, the most striking features of the transfer functions can be traced down to the geometrical and morphological characteristics of the earplug and ear canal.
 [DOI: 10.1121/1.2945709]

PACS number(s): 43.50.Hg, 43.66.Vt [BSF] Pages: 973–981

I. INTRODUCTION

Several studies clearly demonstrate that attenuation values of hearing protectors determined in laboratory conditions significantly exceed the actual protection offered to the individual user (Berger *et al.*, 1998; Franks, 2003). In contrast, the European Noise Directive (2003) on exposure limit values stipulates that the worker's effective exposure must take account of the attenuation provided by the individual hearing protectors. Therefore, the performances of the hearing protection devices should also be verified *in situ*.

Different measurement techniques have been proposed (Franks *et al.*, 2003); the microphone in real ear (MIRE) approach, for instance, offers a quick and objective way to evaluate the attenuation (Berger, 2005). Testing may be carried out with one or two microphones. In the single microphone technique, the receiver is placed in the ear canal during separate, consecutive measurements with and without a hearing protector. Using two microphones, one is placed inside the ear canal underneath the hearing protector, the other measures simultaneously the sound level outside the ear. Both methods have proved to be successful with earmuffs, but the application with earplugs often requires extra adaptations (Hager, 2006; Pääkkönen *et al.*, 2000; Toivonen *et al.*, 2002). By contrast, Voix (2006) and others (Berger, 2007; Nélisse *et al.*, 2007) describe a custom-made earplug with an inner bore that allows insertion of a miniature microphone registering sound pressure levels inside the ear canal behind the hearing protector. In practice, this microphone is mounted in a probe that also contains a reference microphone measuring the sound pressure outside the ear canal (see Fig. 1).

As this test design becomes more widespread (Berger, 2007; Burks and Michael, 2003; Neitzel *et al.*, 2006), a more thorough investigation of the underlying acoustical mechanisms is required, especially with regard to the spatial variation of sound pressure levels in the subject's outer ear canal and the earplug's inner bore. Of particular interest is the relation, henceforth called transfer function, between the sound level measured by the MIRE microphone at the inner bore and the level at the eardrum, as the latter is predicted from the former. Between these two points, an apparent difference is expected at certain frequencies as several authors report substantial pressure fluctuations even in an unoccluded ear canal (Gan *et al.*, 2006; Hammershøi and Møller, 1996; Hellstrom and Axelsson, 1993).

This issue can be addressed by tests with a head-and-torso-simulator (HATS) consisting of a torso and artificial head equipped with pinna, ear canal, and ear simulator mimicking the impedance of the eardrum (Parmentier *et al.*, 2000). The main reason for working with these simulators is

[a)]Electronic mail: annelies.bockstael@ugent.be

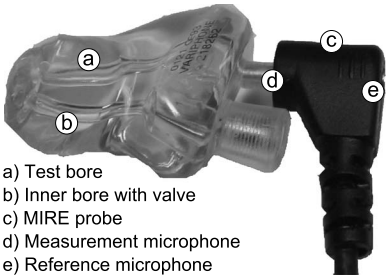


FIG. 1. Earplug with two inner bores; one to adjust the attenuation (with filter or valve) and the test bore for insertion of the MIRE probe with measurement and reference microphone. The measurement microphone measures the sound level in the ear canal behind the hearing protector whereas the reference microphone registers the incoming sound level.

their ability to facilitate certain measurements more difficult to perform with human subjects. Hence, they allow testing in a very stable and controlled setup. The HATS appears to give reproducible results (Schroeter, 1985), close to those obtained with other methods verifying the attenuation of hearing protectors (Parmentier *et al.*, 2000), despite its impossibility to simulate all features of the human head and auditory system (Berger, 2005).

In addition, numerical simulations may be performed to increase the understanding of the HATS measurements and to allow optimizing the MIRE method in a later phase. Here the finite difference time domain (FDTD) technique is chosen as this time domain method is efficient and allows to calculate the spectral transfer functions over the whole auditory spectrum at once. This method has previously been used successfully to model sound propagation in and around the human outer ear (Tian and Qing, 2003; Nakazawa and Nishikata, 2005).

The aim of this research project is to gain insight into the transfer function between the sound level measured at the inner bore of the hearing protector and the sound pressure of interest at the eardrum by combining the results from measurements with the HATS and FDTD simulations.

II. MATERIAL AND METHODS

A. MIRE measurement: General setup

1. Hearing protectors

The tested hearing protectors are manufactured especially for the HATS in acrylic with different attenuation characteristics. Each hearing protector has two inner bores, one allowing the insertion of the MIRE probe (the test bore), the other containing a filter or an adjustable valve determining the attenuation (see Fig. 1). The attenuation of the earplugs with valves may be varied, ranging from an open valve to a completely closed one. Hence, measurements are performed for different conditions: open valve, closed valve, and a valve offering an estimated attenuation of respectively 20, 25, and 30 dB. Conversely, the attenuation offered by the filters is fixed at, respectively, 20, 35, and 65 Lohm. The unit "Lohm" is used by The LEE Company (2006) to reflect flow resistance of gasses and is calculated by the following equation:

$$\text{Lohms} = \frac{Kf_T P}{Q} \tag{1}$$

with Q representing the gas flow, K the gas units constant, f_T a temperature correction factor, and P the upstream absolute pressure.

2. MIRE probe

As stated previously, the MIRE measurements are performed with a probe containing two Knowles low noise FG-3652 microphones; the reference microphone measuring the incoming sound level and the measurement microphone registering the sound pressure in the ear canal behind the hearing protector (Fig. 1).

As the focus of this project lies in the transfer function between the measurement microphone and the sound pressure at the eardrum, the response of the reference microphone is not taken into account. Therefore, the concept "MIRE microphone" henceforth refers to the measurement microphone of the probe.

Correct measurements of the attenuation require a perfect seal of the test bore by the MIRE probe. To asses this requirement, the attenuation of earplugs is measured on the Brüel & Kjær HATS type 4128 C as elaborated in the following, once with a completely closed test bore and once with the test bore ended by the MIRE probe. The similar attenuation values thus obtained confirm that the condition of perfect sealing is fulfilled.

B. Measurements with the HATS

1. Measurement system

Measurements are performed with a laptop PC connected to a four input channel data acquisition front-end of Brüel & Kjær (type 3560-C) linking all sound equipment. Recording equipment consists of two prepolarized free-field 1/2 in. microphones type 4189 (Brüel & Kjær) with preamplifier (type 2669C, Brüel & Kjær), one Knowles low noise FG-3652 microphone (the MIRE microphone) connected to a 9 V preamplifier and the head-and-torso-simulator type 4128 C of Brüel & Kjær with a dual microphone preamplifier (Brüel & Kjær, type 5935). The test stimulus is low pass filtered pink noise with a cutoff frequency of 12.8 kHz generated on the PC using Brüel & Kjær's PULSE LABSHOP version 7.0. The signal is then transmitted via the front-end and a Pioneer A-607 R direct energy metal–oxide–semiconductor (MOS) amplifier through a Renkus-Heinz (model CM 81) loudspeaker. The quality of the sound generation system is not critical as the sound signal will be calibrated out in all measurements.

2. Setup

Testing takes place in an anechoic room to prevent disturbances from sound reflection and background noise, therefore the PC is placed outside and the room is only entered between two successive stimuli. The HATS and one of the free-field microphones are symmetrically placed in front of the loudspeaker at 1.61 m, see Fig. 2. The right test ear of the HATS is oriented toward the loudspeaker.

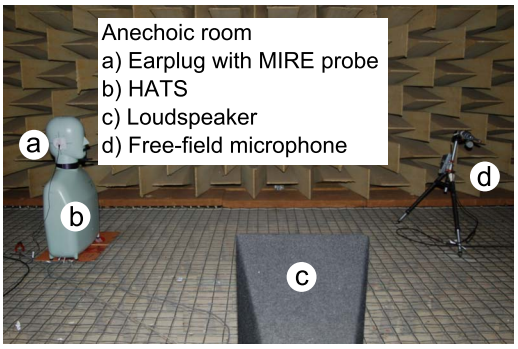


FIG. 2. (Color online) Test setup in the anechoic room with loudspeaker, HATS and free-field microphone.

The aim of the measurements is the determination of transfer functions between the microphone at the ear simulator of the HATS and the MIRE microphone. However, these results should not be influenced by the typical characteristics of the microphones, nor by the features of the test environment and the test signal. Therefore, the responses of the microphones are compared with the simultaneously registered responses of one free-field microphone. Mounting the free-field microphone at approximately the same place as the HATS is impossible as reflections at the HATS's body disturb the reference signal. This resulted in the test setup described earlier. An extra measurement is carried out with the second free-field microphone replacing the HATS to calculate the transfer function between the two measurement points.

3. Measurement sequences and processing

To eliminate possible inaccuracies of the MIRE microphone, its output is first directly calibrated in the free-field over the frequency range of interest by means of the second free-field microphone. Calibrations over the different test days reveal very stable responses. Further, all other microphones are calibrated before each measurement session using the pistonphone 4228 from Brüel & Kjær.

Thereafter, the following steps are carried out for each hearing protector. The earplug of interest is placed in the HATS's ear canal and the MIRE microphone is inserted at the earplug's appropriate bore at a fixed depth so that the probe does not touch the pinna of the HATS. Subsequently, the position of the HATS is checked, the investigator leaves the room and the door is carefully shut. Each measurement is completely repeated in order to verify the reproducibility and to detect possible errors.

The signals from the microphones are registered by the Pulse Labshop software mentioned earlier. Linear averaging is carried out over 3000 samples and overloads are rejected. In the frequency range between 0 Hz and 10 kHz the responses are spectrally analyzed using fast Fourier transform (FFT) (6400 points). To eliminate possible artifacts described previously, the frequency response function is, on the one hand, calculated between the MIRE microphone and the free-field microphone (H_{mf}) and, on the other hand, between the HATS microphone and the free-field microphone (H_{hf}) based on the following equation:

$$H_{xy} = \sqrt{\frac{G_{xy}(k)}{G_{xx}(k)}\frac{G_{yy}(k)}{G_{xy}^{*}(k)}} \tag{2}$$

where x and y are M, H, and F respectively, H_{xy} is the frequency response, $G_{xx}(k)$ and $G_{yy}(k)$ are the autospectra, $G_{xy}(k)$ is the cross spectrum and $G_{xy}^{*}(k)$ is its complex conjugate. Afterward, the transfer function between the MIRE microphone and the HATS microphone (H_{mh}) may be derived by dividing H_{mf} by H_{hf}.

C. Numerical simulations

1. Key factors of the simulations

The key factors and choices made for the numerical FDTD simulations are briefly repeated. Both pressure p and particle velocity $\mathbf{u}$ are discretized in Cartesian grids that are staggered by shifting the grid for discretizing u_α over half of a grid step, $d\alpha/2$, in direction α with respect to the grid chosen for discretizing p. In time, staggering is obtained by calculating p at $t=ldt$ and u at $t=(l+\frac{1}{2})dt$. The resulting equations:

$$u_\alpha^{l+1/2}\left(\alpha+\frac{1}{2}\right) = u_\alpha^{l-1/2}\left(\alpha+\frac{1}{2}\right) - \frac{dt}{\rho_0 d\alpha}\{p^l(\alpha+1) - p^l\} \tag{3}$$

$$p^{l+1} = p^l - \sum_{\beta=x,y,z}\frac{\rho_0 c^2 dt}{d\beta}\left\{u_\beta^{l+1/2}\left(\beta+\frac{1}{2}\right) - u_\beta^{l+1/2}\left(\beta-\frac{1}{2}\right)\right\} \tag{4}$$

with c the speed of sound and ρ_0 the density of air, allow to step in time replacing old values by newly calculated ones without much memory overhead. The brief notation $(\alpha+q)$ is used to indicate that the value is taken at a point shifted by q spatial steps $d\alpha$ in the α direction with respect to the reference location referred to by indices (i,j,k). The first equation is repeated for $\alpha=x,y$, and z.

Boundary impedance of the form

$$Z = j\omega Z_1 + Z_0 + \frac{Z_{-1}}{j\omega} \tag{5}$$

can easily be implemented in the FDTD method (Bottel-doorden, 1994). Such boundary conditions will be used to model the earplug's and ear canal's material.

2. Modeling the impedance of the outer, middle, and inner ear

a. Impedance of the outer ear. For the propagation of sound in the outer ear canal, the impedance of bone (6.12 kg/m^2 s) (Wit *et al.*, 1987) is included as boundary condition. This approximates the real-life situation, where the hearing protector fills the cartilageous part of the outer ear canal and hence relevant sound propagation effects occur mainly in the ossicular part.

b. Impedance of middle and inner ear. The acoustics of the middle and inner ear is represented by the impedance at the eardrum, $Z_d=p/u_n$, where u_n is the orthogonal component of $\mathbf{u}$, as explicitly modeling the middle ear is outside the scope of this work. The impedance at the eardrum is included because different authors have shown that, especially

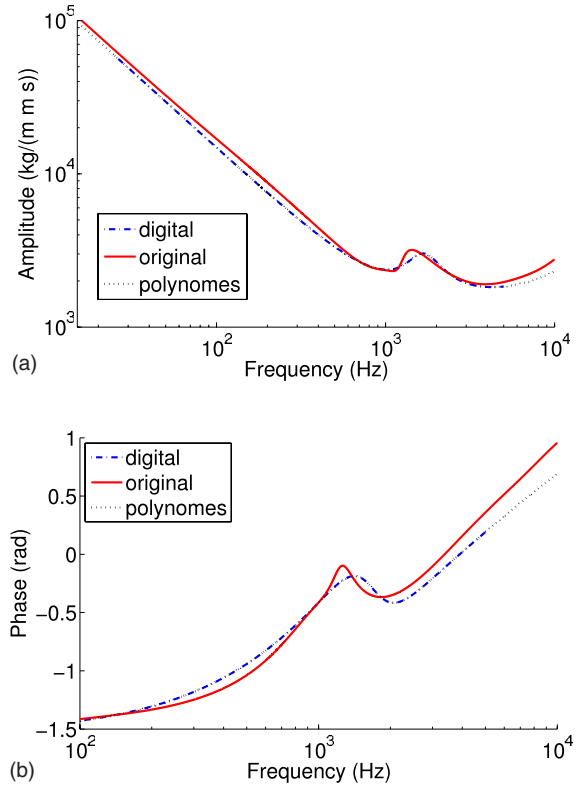


FIG. 3. (Color online) Amplitude and phase of the complex surface impedance of the eardrum with digital approximation.

for intraaural devices, the eardrum impedance needs to be simulated (Schroeter and Poesselt, 1986; Hammershøi and Møller, 1996).

Various eardrum impedance models are suggested in literature. Here the terminating impedance is based on one-dimensional circuit models of the middle and inner ear as the cross-sectional variation is of little importance for the problem at hand. These models allow the approximation of the effective impedance by compounding networks of acoustical and mechanical compliances, masses, frictional resistances and transformers (Gan *et al.*, 2006) representing the relevant physical parts of the auditory system (Kringlebotn, 1988).

Different circuit models by Hudde and Engel (1998), Kringlebotn (1988), Pascal *et al.* (1998), and Shaw and Stinson (1981) are investigated. For each approach, the overall eardrum impedance is calculated and compared with the eardrum impedances for human subjects as reported by different authors (Farmer-Fedor and Rabbitt, 2002; Keefe *et al.*, 1993; Margolis and Keefe, 1999; Voss and Allen, 1994). Further, an unoccluded ear canal is modeled using numerical FDTD simulations with various eardrum impedances according to the different network models. The outcome of these simulations is compared to measurements made by Hammershøi and Møller (1996).

From all these comparisons it can be seen that pressure patterns derived from the model proposed by Kringlebotn (1988) resemble the experimental results most closely. The difference in peak frequency between the FDTD model and experimental results (Hammershøi and Møller, 1996) may be caused by the different canal lengths in experimental subjects and the model (Gan *et al.*, 2006). Hence, the eardrum impedance resulting from Kringlebotn's network is included in the FDTD model as was done by Hiselius (2004) in his two-port model of an occluded ear canal.

To include the complex impedance in the FDTD model a ratio of polynomials in $j\omega$ is first fitted. A bilinear transformation is then used to transform this analog filter to a digital equivalent:

$$p = \frac{\Sigma_{i=0}^{n} d_i z^{-i}}{\Sigma_{k=0}^{m} c_k z^{-k}} u_n. \tag{6}$$

The impedance of the eardrum does not represent locally reacting material. Therefore, it is implemented on the average field:

$$u_n^{l+1/2} = \frac{\Sigma_{k=0}^{m} c_k z^{-k}}{\Sigma_{i=0}^{n} d_i z^{-i}} \langle p^l \rangle, \tag{7}$$

where the angular brackets denote spatial averaging over the eardrum. The temporal and spatial mismatch due to the use of staggered grids turn out to have little influence. Hence the coordinate system is chosen such that a coordinate plane $x = \frac{1}{2}dx$, where u_n is discretized, coincides with the eardrum. The pressure half a grid step away at $x = dx$ is assumed to be a good approximation for the pressure on the eardrum. The single digital filter representing the middle and inner ear is implemented using standard filter routines in MATLAB.

Due to the complicated frequency dependence of the impedance, $m = n = 5$ is needed. Figure 3 shows amplitude and phase of the complex surface impedance of the eardrum and the digital approximation that is used.

3. Modeling the hearing protector

a. Dimensions of the hearing protector. The relevant dimensions of the hearing protector in total and of the two inner bores are measured with a digital caliper accurate to 0.01 mm. However, the length of the inner bores is slightly adapted in the simulations to enhance similarity with the measurements. This modification will be discussed in Sec. IV.

b. Modeling the sound field in the narrow channels in the hearing protector. The channels in the hearing protector are very narrow. Therefore, the effect of viscosity and heat conduction becomes potentially important. To avoid having to compute numerically the strong spatial dependence of the field close to the boundaries, a subgrid scale approximation for the boundary is used. This approach is based on an analytical description of the vorticity and entropy layer close to a flat boundary (Pierce, 1997). It is previously introduced as a subgrid scale approximation in FDTD simulation by Botteldooren (1997). The time domain approximation of the square root is refined in comparison to this earlier publication to be applicable over a wider frequency range. The derivation for the viscosity effect starts from the equation for conservation of impulse for the parallel component u_β averaged over a grid cell orthogonal to the surface

$$j\omega\rho_0 \bar{u}_\beta = -\frac{\partial \bar{p}}{\partial \beta} - \frac{1}{d\alpha}\tau_\beta \tag{8}$$

where $d\alpha$ is the grid cell size orthogonal to the surface, the over stripe indicates the averaging, and

$$-\frac{1}{d\alpha}\tau_\beta = \frac{\mu}{d\alpha}\int_0^{d\alpha}\frac{\partial^2 u_\beta}{\partial\alpha^2}d\alpha. \tag{9}$$

The latter term accounts for the influence of viscosity (μ is the dynamic viscosity). Within the boundary layer approximation, the second order derivative is dominated by the boundary layer field. It vanishes at the edge of the boundary cell ($\alpha = d\alpha$) as long as the boundary layer thickness is small compared to the grid cell size: $\delta = \sqrt{2\nu/\omega} < d\alpha$ where $\nu = \mu/\rho_0$. Thus we obtain

$$\tau_\beta = -\mu \left.\frac{\partial u_\beta}{\partial\alpha}\right|_B. \tag{10}$$

Introducing the exponential decay of the boundary layer results in

$$\tau_\beta = \rho_0(1+j)\sqrt{\frac{\nu\omega}{2}}\bar{u}_\beta. \tag{11}$$

Equations (8) and (11) are now transformed back to time domain. For Eq. (11) this is not trivial. Direct transformation of the product of $\sqrt{\omega}$ with the field value results in a convolution involving a $\sqrt{t}$. Numerical approximation of this convolution can result in very long calculation times. Therefore, the $\sqrt{\omega}$ term is first approximated by a ratio of polynomials in $j\omega$. This methodology is inspired by classical electronic filter design. For digital filter design, a bilinear transformation is traditionally used to transform continuous to discrete time, as it unconditionally results in a stable digital filter that is also a ratio of polynomials. The digital filter approximation that is obtained using standard approximation techniques focusing on a frequency range [0 Hz, 5000 Hz] is written as

$$\tau_\beta = \sqrt{\mu\rho_0}\frac{\sum_{i=0}^n b_i z^{-i}}{\sum_{k=0}^m a_k z^{-k}}\bar{u}_\beta, \tag{12}$$

where z^{-1} is the equivalent in Z domain of a single time step delay. This expansion is used to discretize Eq. (8) in the staggered grid both in space and in time. As usual, this equation is evaluated at $t = ldt$ (where p is known), whereas the components of $\mathbf{u}$ are discretized at $t = (l+\frac{1}{2})dt$. Thus an additional $(1+z^{-1})/2$ is introduced to resolve this mismatch. This eventually leads to the adapted FDTD update equation:

$$\left(\frac{\rho_0 a_0}{dt} + \frac{\sqrt{\mu\rho_0}}{2d\alpha}\right)u_\beta^{l+1/2} = -\sum_{k=0}^m a_k \left.\frac{\partial p}{\partial\beta}\right|^{l-k} - \frac{\rho_0}{dt}\sum_{k=1}^{m+1}(a_k - a_{k-1})u_\beta^{l-k+1/2} - \frac{\sqrt{\mu\rho_0}}{2d\alpha}\sum_{i=1}^{n+1}(b_i + b_{i-1})u_\beta^{l-i+1/2}, \tag{13}$$

where we introduced $a_{m+1} = b_{n+1} = 0$ to simplify notations. Spatial discretization is not explicitly denoted.

Note that the approach used to include boundary layer effects is rather memory extensive. It requires storage of m old values of the spatial derivative of p and $\max(m,n)-1$ additional old values of u_β. Fortunately, this additional storage is required at the boundaries only. For the simulations reported in this article, m and n are chosen equal to 2. This results in the coefficients $a_0 = 1$, $a_1 = -1.95$, $a_2 = 0.95$, $b_0 = 403.73$, $b_1 = -802.77$, and $b_2 = 399.04$, and a reasonably accurate approximation over the frequency range of interest.

The influence of heat conduction on sound propagation can also be introduced using a subgrid scale approximation based on boundary layer theory (Botteldooren, 1997; Howe, 1998). The equation that describes the evolution of the small acoustic pressure fluctuation is usually derived from the conservation of energy and the conservation of mass. By keeping terms that describe heat flux near the flat surface and by assuming that this flux is largest within a small boundary layer close to that surface, the grid cell averaged equation can be written as

$$j\omega\bar{p} = -\rho_0 c^2 \nabla\cdot\bar{u} - \frac{\gamma-1}{d\alpha}q_s \tag{14}$$

where $d\alpha$ is again the grid cell size orthogonal to the surface, the overbar indicates the averaging, and

$$-\frac{1}{d\alpha}q_s = \frac{\kappa}{d\alpha}\int_0^{d\alpha}\frac{\partial^2 T}{\partial\alpha^2}d\alpha \tag{15}$$

where κ is the heat conductivity and T is the acoustic temperature fluctuation. Within the boundary layer approximation, the second order derivative is again dominated by the boundary layer field. It vanishes at the edge of the boundary cell ($\alpha = d\alpha$) as long as the thermal boundary layer thickness is small compared to the grid cell size: $\sqrt{2(\gamma-1)\kappa/\omega\gamma\rho_0 R} < d\alpha$. Thus we obtain

$$q_s = -\left.\kappa\frac{\partial T}{\partial\alpha}\right|_B. \tag{16}$$

Introducing the exponential decay of the boundary layer results in

$$q_s = (1+j)\sqrt{\frac{\kappa\rho_0\gamma R\omega}{2(\gamma-1)}}\bar{T}. \tag{17}$$

The cell averaged acoustical temperature fluctuation $\bar{T}$ is dominated by isentropic acoustic propagation. Thus, the known relationship that relates acoustic pressure to acoustic temperature $p = T$ can be used to obtain

$$q_s = (1+j)\sqrt{\frac{\kappa(\gamma-1)\omega}{2\rho_0\gamma R}}\bar{p} = (1+j)\sqrt{\frac{\kappa\omega}{2\rho_0 c_p}}\bar{p}, \tag{18}$$

where c_p is the specific heat at constant pressure for the gas (air). Exactly the same way as for the viscous boundary layer effect, Eqs. (14) and (18) are now transformed back to time domain and discretized following the FDTD scheme as described earlier.

Note that the approach used to include boundary layer effects is rather memory extensive. It requires storage of m old values of the divergence of u and $\max(m,n)-1$ additional old values of p for implementing the thermal conduction effects. Fortunately, this additional storage is required at the boundaries only. For the simulations reported in this ar-

ticle, m and n are chosen equal to 2. This results in the coefficients $a_0=1$, $a_1=-1.95$, $a_2=0.95$, $b_0=403.73$, $b_1=-802.77$, and $b_2=399.01$, and a reasonably accurate approximation over the frequency range of interest.

c. Modeling the impedance of the hearing protector's material. The impedance of acrylic is calculated based on measurements of the complex Young's modulus E (Hillström *et al.*, 2000, 2003) and Poisson's ratio σ (Hillström *et al.*, 2003). From these data, the complex longitudinal sound speed v may be calculated using the following expression (Jarzynski *et al.*, 2003; Pierce, 1997):

$$v = \sqrt{\frac{\frac{E}{2(1-2\sigma)} + \frac{4}{3}\frac{E}{2(1+\sigma)}}{\rho}} \tag{19}$$

with the density of acrylic $\rho=1183\ \mathrm{kg/m^3}$ (Hillström *et al.*, 2003).

The characteristic impedance of acrylic can easily be calculated based on ρ and the complex wave number $\kappa_\delta = \omega/v$ with ω the angular frequency. Considering the reflection at an infinitely thick and long layer, this yields to a merely constant and real impedance in the frequency range of interest, $Z \approx 3.1\times10^6\ \mathrm{kg/s\,m^2}$.

Impedance of the entity earplug–ear canal. The surface impedance of the earplug terminating the residual part of ear canal between hearing protector and eardrum does not simply equal its material impedance. By contrast, the combination of earplug and resilient ear canal is believed to influence the resulting surface impedance (Hiselius, 2005; Schroeter, 1985).

This surface impedance is measured by tightly mounting the HATS's pinna with hearing protector on an impedance tube. The reflection coefficient and normal surface impedance are determined by measuring the two microphone transfer function according to the ISO standard (1998). Because the impedance of the earplug's material is so high, viscous damping in the measurement tube has to be calibrated out. Within the frequency range of interest, the surface impedance can be approximated by an expression of the form written in Eq. (5) with constants $Z_0=0.50\times10^4\ \mathrm{kg/s\,m^2}$, $Z_1=0.41\ \mathrm{kg/m^2}$ and $Z_{-1}=1.69\times10^7\ \mathrm{kg/s^2\,m^2}$.

e. Modeling the impedance of the miniature microphone. The acoustic impedance of the microphone's diagram system is modeled by a very high impedance. This is based on the fact that a small diameter increases the diagram's impedance substantially (Brüel & Kjær, 1996), yielding to a very high impedance for 1/10 in. microphones.

III. RESULTS

A. Transfer function between MIRE microphone and HATS microphone

Examples of the amplitude of the transfer functions (H_{mh}) between the MIRE microphone and the HATS microphone as measured and simulated are depicted in, respectively, Figs. 4 and 5. The H_{mh} for 25 dB attenuation is omitted from Fig. 4 to enhance clarity of the graph and because this H_{mh} is almost identical to the H_{mh} for 20 dB attenuation. In general, little difference is seen between the different transfer functions.

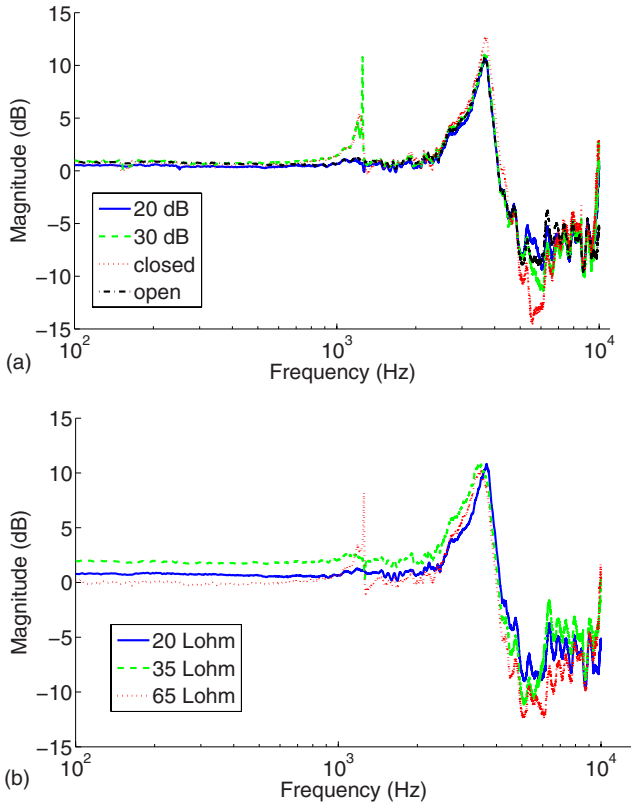


FIG. 4. (Color online) Amplitude of the measured transfer functions between the MIRE microphone and the HATS's microphone for earplugs with valve and filter. (Top) Transfer functions are depicted for a completely closed valve (closed), a completely open valve (open) and a valve offering an attenuation of 20 and 30 dB. (Bottom) Transfer functions are shown for earplugs with, respectively, a 20, 35, and 65 Lohm filter.

For most earplugs and for the numerical simulations, the transfer function between the MIRE microphone and the microphone of the HATS is found to be flat in the lower and middle frequency region (up to 2500 Hz). However, the spectrum of the transfer functions shows a distinct peak around 1000 Hz for some earplugs in all sets of hearing pro-

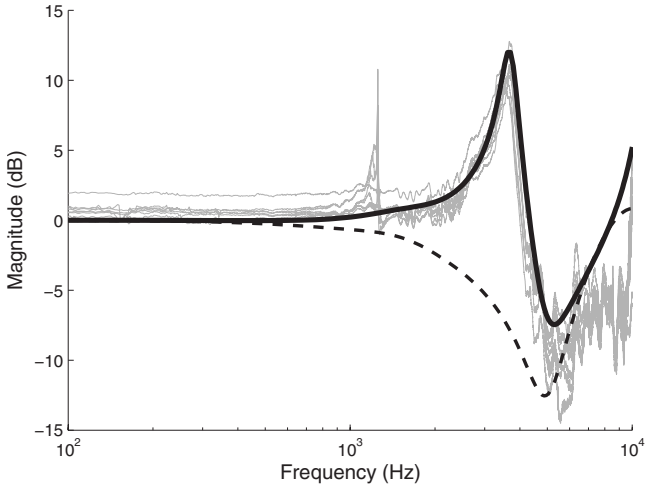


FIG. 5. Amplitude of the simulated transfer function (thick black line) between the sound pressure at the MIRE microphone's position, i.e., at end of the test bore, and at the eardrum. In the background, the measured transfer functions are plotted in gray. The dashed line represents the simulated transfer function with closed test bore.

tection devices. This peak is never seen in the numerical approach with parameters described in Sec. II C. The hypothesis of accidental measurement errors is somewhat inconsistent with the fact that repeated tests show the same aberration. An alternative line of thinking is the assumption that this peak corresponds to an extra mode caused by the artificial character of the HATS.

The transfer functions' amplitude increases above 2500 Hz with a clearly distinguishable peak around 3600 Hz. The FDTD simulations show that merely adaptations in the characteristics of the test bore and its boundaries affect this maximum. Hence, this structure is held responsible for the observed resonance. Measurements with the MIRE microphone attached to the hearing protector in free-field conditions (i.e., outside the HATS' ear canal) confirm this assumption by showing a resonance peak at the same frequency.

In the frequency range between 4000 and 7000 Hz, the spectral behavior of the transfer function is dominated by the resonance of the residual volume of the ear canal. The simulation with closed test bore shown by the dashed line in Fig. 5 already indicates that the residual volume is responsible for this broad dip in the transfer function. From data reported by Tonndorf (1988) the length of the residual volume between the hearing protector and the eardrum is estimated at 16 mm, yielding to a resonance frequency consistent with our measured and simulated transfer functions. The resonance of the test bore thus compensates for the dip produced by the resonance of the residual volume. Unfortunately the compensation does not result in a overall flat transfer function.

At still higher frequencies, the measurements show an increased complex and peaky behavior. This could be due to different higher order modes involving amongst others the vibration of the eardrum. It is unlikely that the HATS measurements are representative for measurements in real humans in this upper frequency range as the inner ear model of the HATS probably fails at these frequencies.

The phase of the transfer functions reveals that the MIRE microphone and the HATS microphone are in phase. Fluctuations in phase are observed at the resonance frequencies described earlier, these findings are consistent with the Kramer–Kronig relations.

B. Reproducibility

The reproducibility is verified by comparing the results from repeated measurements. The responses of the MIRE microphone appear to be stable if the microphone is carefully placed. The reproducibility is within a 1 dB range in the frequency range from 300 Hz to 10 kHz. Outside this range, no variations greater than 2 dB are found.

C. Differences between transfer function

Although only little differences are observed between transfer functions measured within the groups of hearing protection devices, it is investigated whether this variance is caused by possible (small) variations in the design of the different earplugs or by their attenuation. This is done by performing multiple measurements using the acrylic hearing protectors with their adjustable valve put in its different positions. The variation between transfer functions obtained from the same earplug with different attenuation levels nor the differences between hearing protectors with the same attenuation do exceed the outer limits of the reproducibility found earlier (2 dB). Thus these results do not provide evidence for substantial influence of the attenuation level or the earplug's design on the variances between transfer functions. However, the latter finding is somewhat obvious as all protectors are manufactured based on the same impression of the HATS's ear canal.

IV. DISCUSSION

Research on techniques to measure the attenuation of hearing protectors *in situ* is vital because the selection of hearing protectors should be based on their efficiency considering the real use and not only the laboratory attenuation data (Franks, 2003). Among the different hearing protection methods, custom-made earplugs deserve extra attention because they tend to be more positively rated in terms of comfort and usability (Hsu *et al.*, 2004), but nevertheless they also merit individual field attention measurements (Berger *et al.*, 1996), for instance by the MIRE technique investigated in this study.

To apply this approach in practice, the transfer function between the sound pressure at the MIRE microphone and at the eardrum has to be as close to unity as possible or at least, it has to be known and electronically compensated for. This issue is addressed by performing MIRE measurements on the HATS 4128 C. The free-field responses of the HATS in this study's setup is very similar to those reported by the manufacturer for similar sound incident directions. This indicates that the HATS's measurements are free from influences by the test signal and the test environment.

All measurements are performed with fixed sound incident directions because Hammershøi and Møller (1996) have shown that the transmission to the eardrum from any point between the eardrum and the point 6 mm outside the ear canal can be considered directional independent. Moreover they have proved that this directional independence is also valid for a blocked entrance.

In the FDTD simulations, the sound pressure level at the eardrum is based on the sound pressure level at one central point in the ear canal because the sound pressure is constant across the ear canal below 10 kHz and thus only the longitudinal mode is present (Hammershøi and Møller, 1996), given that any nonplanar modes in this frequency range are strongly attenuated (Voss and Allen, 1994).

The backradiation impedance, i.e., the radiation impedance seen outward from the ear entrance (Hudde and Engel, 1998), is not included in this model because Hiselius (2004) has shown that inclusion or exclusion of this parameter has little impact on the results for occluded ears.

The measurement results reveal stable and reproducible transfer functions which are independent of the earplug's attenuation. Moreover, FDTD simulations of the hearing protector and outer ear canal come close to the results of most earplugs, except for the maximum around 1000 Hz visible in

the spectra of some transfer functions. This maximum does never occur in the numerical approach, even when the nominal values of the different parameters are modified. Hence, the origin of the peak is most likely to be located elsewhere, for instance in the artificial character of the middle ear model of the HATS. In the simulations the impedance of the middle and inner ear have been taken from literature data on biological ears and might therefore not show the resonance around 1000 Hz as strongly as the HATS middle ear model.

Further, the test bore was slightly lengthened in the simulations to enhance similarity with its measured resonance frequency. This adaptation accounts for loss mechanisms for which exact quantification is outside the scope of this work. For instance, the viscothermal damping effects at the (sharp) edges at the end of the test bore are not accounted for. Also, the test bore's termination toward the eardrum consists in reality of a very small cavity with rounded, inclined walls. The simulation includes a cavity with comparable dimensions but with straight walls. The more gradual transition between the air in the ear canal and the test bore might in reality enhance the radiation of sound toward the residual part of the ear canal. Hence, the resonance frequency of the test bore might be lowered.

Anyway, the lower measured resonance frequency is thought to be due to characteristics of the test bore only. As measurements with the hearing protector in free-field show a comparable resonance frequency, the outer ear nor the middle may influence this maximum (see Sec. III A). Moreover, hearing protectors differing in the design of the second inner bore all have a similar maximum (see Fig. 4). These findings strongly support the hypothesis of the test bore determining the resonance frequency. Hence, it seems justified to include the most likely effects of both test bore's terminations in the simulations by slightly lengthening the test bore itself.

The combination of measurements and simulations makes it obvious that the most striking features of the transfer functions may be traced down to the test bore and the ear canal's residual cavity between hearing protector and eardrum. As the main differences between particular acrylic earplugs lie in the dimensions of these structures, it is most likely that only limited adaptations based on physical dimensions of a particular earplug are needed to extend the suitability of the numerical model as a correction for the MIRE measurement microphone's response.

Hence, the model obtained in this research project will be further investigated for human subjects. If the transfer functions can be predicted for an individual human, the worker's effective exposure defined by the European Directive (2003) may be calculated from the sound level measured by the MIRE microphone.

V. CONCLUSION

Measurements carried out in this project clearly reveal an apparent difference between the sound pressure registered by the MIRE measurement microphone and the sound pressure of interest at the eardrum. Hence, the responses of the MIRE microphone need to be corrected by a certain transfer function to allow prediction of the individual's effective exposure.

The transfer functions measured in this study on a HATS appear to be very stable and reproducible with in general a maximum around 3600 Hz and multiple negative deviations in the higher frequencies. FDTD simulations of the earplug and outer ear canal reveal that the observed maximum is most likely caused by resonances in the test bore, whereas the negative deviations may be linked to the ear canal's residual cavity between the hearing protector and the eardrum. Further research will show to what extend the transfer functions obtained in this research project can be applied to human subjects.

ACKNOWLEDGMENTS

The authors would like to thank the firm Variphone for the fabrication of the custom-made earplugs.

Berger, E. H. (**2005**). "Preferred methods for measuring hearing protector attenuation," *Environmental Noise Control* (Inter Noise, Brazil).

Berger, E. H. (**2007**). "Introducing F-MIRE testing-background and concepts," Technical Report No. E-A-R 06-29/HP, E.A.RCal Laboratory, Indianapolis.

Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (**1998**). "Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part III. The validity of using subject-fit data," J. Acoust. Soc. Am. **103**, 665–672.

Berger, E. H., Franks, J. R., and Lindgren, F. (**1996**). *International Review of Field Studies of Hearing Protection Attenuation* (Thieme Medical Pub, New York), pp. 361–377.

Botteldooren, D. (**1995**). "Finite-difference time-domain simulation of low-frequency room acoustic problems," J. Acoust. Soc. Am. **98**, 3302–3308.

Botteldooren, D. (**1997**). "Vorticity and entropy boundary conditions for acoustical finite-difference time-domain solutions," J. Acoust. Soc. Am. **102**, 170–178.

Brüel & Kjær (**1996**). "Microphone handbook," Technical Report.

Burks, J. A., and Michael, K. L. (**2003**). "A new best practice for hearing conservation," in *Noise-Con*, Cleveland, OH, June 23–25.

European Parliament and Council. (**2003**). "Directive 2003-10-EC on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise)," 2003-10-EC.

Farmer-Fedor, B., and Rabbitt, R. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.

Franks, J. (**2003**). "Comparison of the regulatory noise reduction rating (NRR) and the required ANSI S3.19 test method with real world outcomes and results from testing with the new ANSI S12.6B method," in *Workshop on Hearing Protector Devices* (United States Environmental Protection Agency, Washington, D.C.).

Franks, J., Murphy, W., Harris, D., Johnson, J., and Shaw, P. (**2003**). "Alternative field methods for measuring hearing protector performance," Am. Ind. Hyg. Assoc. J. **64**, 501–509.

Gan, R. Z., Sun, Q., Feng, B., and Wood, M. W. (**2006**). "Acoustic-structural coupled finite element analysis for sound transmission in human ear-pressure distribution," Med. Eng. Phys. **28**, 395–404.

Hager, L. D. (**2006**). *Fit testing ear plugs*, Occupational Health & Safety.

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

Hellstrom, P.-A., and Axelsson, A. (**1993**). "Miniature microphone probe tube measurements in the external auditory canal," J. Acoust. Soc. Am. **93**, 907–919.

Hillström, L., Mossberg, M., and Lundberg, B. (**2000**). "Identification of complex modulus from measured strains on an axially impacted bar using least squares," J. Sound Vib. **230**, 689–707.

Hillström, L., Valdek, U., and Lundberg, B. (**2003**). "Estimation of the state vector and identification of the complex modulus of a beam," J. Sound

Vib. **261**, 653–673.
Hiselius, P. (**2004**). "Method to assess acoustical two-port properties of earplugs," Acta. Acust. Acust. **90**, 137–151.
Hiselius, P. (**2005**). "Attenuation of earplugs—objective predictions compared to subjective reat measurements," Acta. Acust. Acust. **91**, 764–770.
Howe, M. S. (**1998**). *Acoustics of Fluid-Structure Interactions, Cambridge Monographs on Mechanics* (Cambridge University Press, Cambridge).
Hsu, Y.-L., Huang, C.-C., Yo, C.-Y., Chen, C.-J., and Lien, C.-M. (**2004**). "Comfort evaluation of hearing protection," Int. J. Ind. Ergonom. **33**, 543–551.
Hudde, H., and Engel, A. (**1998**). "Measuring and modeling basis properties of the human middle ear and ear canal. Part II: eardrum impedances, transfer functions and model calculations," Acust. Acta Acust. **84**, 1091–1109.
International Standard Organisation (ISO). (**1998**). "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes-Part 2: Transfer-function method," ISO 10534-2.
Jarzynski, J., Balizer, E., Fedderly, J. J., and Lee, G. (**2003**). *Encyclopedia of Polymer Science and Technology* (Wiley, New York), pp. 1–47.
Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2638.
Kringlebotn, M. (**1988**). "Network model of the human middle ear," Scand. Audiol. **17**, 75–85.
The LEE Company (**2006**). "Laws for gas-how to calculate flow resistance for gasses," www.theleeco.com (Last viewed March 27, 2008).
Margolis, R., Paul, S., Saly, G. L., Sachern, P. A., and Keefe, D. (**1999**). "Wideband reflectance tympanometry in normal adults," J. Acoust. Soc. Am. **106**, 265–280.
Nakazawa, M., and Nishikata, A. (**2005**). "Development of sound localization system with tube earphone using human head model with ear canal," IEICE Trans. Fundamentals **E88-A**, 3584–3592.
Neitzel, R., Somers, S., and Seixas, N. (**2006**). "Variability of real-world hearing protector attenuation measurements," Ann. Occup. Hyg. **50**, 679–691.
Nélisse, H., Gaudreau, M.-A., Voix, J., Laville, F., and Boutin, J. (**2007**). "A preliminary study on the measurement of effective hearing protection device attentuation during a work-shift," in *Noise at Work 2007 - Proceedings* (CIDB, INCEEUROPE, Association AINF, Lille), pp. 1295–1322.
Pääkkönen, R., Savolainen, S., Myllyniemi, J., and Lehtomaäki, K. (**2000**). "Ear plug fit and attenuation - an experimental study," Acust. Acta Acust. **86**, 481–484.
Parmentier, G., Dancer, A., Buck, K., Kronenberger, G., and Beck, C. (**2000**). "Artificial head (ATF) for evaluation of hearing protectors," Acust. Acta Acust. **86**, 847–852.
Pascal, J., Bourgeade, A., Lagier, M., and Legros, C. (**1998**). "Linear and nonlinear model of the human middle ear," J. Acoust. Soc. Am. **104**, 1509–1516.
Pierce, A. D. (**1997**). *Encyclopedia of Acoustics*, John Wiley & Sons, New York, 21–37.
Schroeter, J. (**1985**). "The use of acoutsical test fixtures for the measurement of hearing protector attenuation. Part I: Review of previous work and the design of an improved test fixture," J. Acoust. Soc. Am. **79**, 1065–1081.
Schroeter, J., and Poesselt, C. (**1986**). "The use of acoustical test fixtures for the measurement of hearing protector attenuation. Part II: Modeling the external ears, simulating bone conduction, and comparing test fixture and real-ear data," J. Acoust. Soc. Am. **80**, 505–527.
Shaw, E. A. G., and Stinson, M. R. (**1981**). "Network concepts and energy flow in the human middle ear," J. Acoust. Soc. Am. **69**, S44.
Tian, X., and Qing, H. (**2003**). "Finite difference computation of head-related transfer function for human hearing," J. Acoust. Soc. Am. **113**, 2434–2441.
Toivonen, M., Pääkkönen, R., Savolainen, S., and Lethtomäki, K. (**2002**). "Noise attenuation and proper insertion of earplugs into ear canals," Ann. Occup. Hyg. **46**, 527–530.
Tonndorf, J. (**1988**). *Physiology of the Ear* (Raven, New York), pp. 29–39.
Voix, J. (**2006**). "Mise au point d'un bouchon d'oreille 'intelligent' (Development of a 'smart' earplug)," Ph.D. thesis, Ecole de Technologie Supérieure Université du Québec.
Voss, S., and Allen, J. (**1994**). "Meaurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.
Wit, H. P., Damme, K. J., and Van Spoor, C. W. (**1987**). *Fysica voor de fysiotherapeut (Physics for the physiotherapist)* (Bunge, Utrecht).

Reliability of estimating the room volume from a single room impulse response

Martin Kuster[a)]
Laboratory of Acoustic Imaging and Sound Control, Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 23 January 2008; revised 13 May 2008; accepted 13 May 2008)

The methods investigated for the room volume estimation are based on geometrical acoustics, eigenmode, and diffuse field models and no data other than the room impulse response are available. The measurements include several receiver positions in a total of 12 rooms of vastly different sizes and acoustic characteristics. The limitations in identifying the pivotal specular reflections of the geometrical acoustics model in measured room impulse responses are examined both theoretically and experimentally. The eigenmode method uses the theoretical expression for the Schroeder frequency and the difficulty of accurately estimating this frequency from the varying statistics of the room transfer function is highlighted. Reliable results are only obtained with the diffuse field model and a part of the observed variance in the experimental results is explained by theoretical expressions for the standard deviation of the reverberant sound pressure and the reverberation time. The limitations due to source and receiver directivity are discussed and a simple volume estimation method based on an approximate relationship with the reverberation time is also presented.
 [DOI: 10.1121/1.2940585]

PACS number(s): 43.55.Gx, 43.55.Br, 43.55.Mc [NX] Pages: 982–993

I. INTRODUCTION

The general course in room acoustics research is to compare measurements of the room impulse response (RIR) or total sound pressure level with predictions obtained from room acoustic models using geometrical and acoustical room parameters.[1,2] An interesting problem is to reverse the process and observe to what extent and accuracy these parameters can be retrieved from measured data. Depending on the approach followed, this fits into the subjects of inverse methods or parameter extraction. In a room acoustics context, the most important geometry parameters are the room volume and the source-to-receiver distance. In the present paper the focus is on the estimation of the former, but the estimation of the latter is also investigated.

The ease and accuracy with which the room volume can be estimated from a single RIR is relevant to the understanding of room acoustics for the following reasons. The (combination of) parameters extracted from the RIR for the volume estimation are those that do change with a change in room volume. In this context it is interesting to note that a number of perceptual experiments performed by Cabrera and colleagues indicated that auditory room size perception is related to clarity index.[3,4] Further, if it proves to be very difficult to obtain accurate volume estimates, it can be concluded that the exact value of this parameter does not greatly affect the RIR. Apart from the relevance to basic room acoustics research, the estimation of the room volume by acoustic means can have practical applications in cases where, for a number of possible reasons, the room volume cannot be determined by other means.

At least three possible approaches can be identified for the estimation of the room parameters. The first approach requires geometric arrays of receiver positions and was shown previously to provide detailed room information but cannot be used with a single RIR.[5] The second approach is based on the extraction of acoustic parameters from a RIR that are then used inversely with one of the standard room acoustic models. A number of authors have used this approach to find a, not necessarily unique, optimum room parameter set that, when fed into the room acoustic model, results in the desired target values for the acoustic parameters.[6,7] The third approach is based on the extraction of more general signal parameters, of which the acoustic parameters may be a subset, that are then used in conjunction with "blind" methods such as maximum-likelihood or neural networks. The extraction of suitable signal parameters for the parametrization of RIRs has been performed by Hulsebos[8] and van der Vorm,[9] but the found parameters are not applicable to the estimation of geometrical room parameters. Blind methods in room acoustics have been investigated for example by Li and Cox[10] and Ratnam *et al.*[11] In the present paper, the second approach is followed because it can be used with a single RIR and has the potential of using the limited available data more effectively than the third approach. The three room acoustic models employed are based on geometrical acoustics, eigenmode or diffuse field assumptions. The suitability of each model and consequent success of the estimation method is considered separately.

The framework, within which the estimation methods are to be applied, is as follows. No knowledge is to be assumed about either the source or receiver characteristics, their position within the room, or the distance between them. Further, no assumptions are employed about the acoustic characteristics of the room. Several limitations that are

[a)]Electronic mail: kuster_martin@hotmail.com.

TABLE I. List of measured rooms together with the number of receiver positions N_R, the geometric room volume V_{geo}, the broadband reverberation time T_{60}, the absorption area A, and the Schroeder frequency $f_{Schroeder}$.[a]

Name (location)[a]	N_R	V_{geo}[b] (m^3)	T_{60} (s)	A (m^2)	$f_{Schroeder}$ (Hz)	Shape, Remarks
Lavatory (SARC, QUB)	18	5	0.3	2.6	490	Rectangular, few absorption
Office (LG023, SARC, QUB)	15	60	0.6	16	200	Rectangular, corner protrusion
Listening room (LG013, SARC, QUB)	10	131	0.3	70	96	Rectangular, special treatment
Multimedia room (SARC, QUB)	14	150	0.4	60	103	Rectangular, corner protrusion
Lecture hall A (Zaal G, TU Delft)	143	180	0.9	32.2	141	Rectangular, tiered seating
Lecture hall B (School of Music, QUB)	4	550*	1.0	89	85	Rectangular with bay window
McMordie Hall (School of Music, QUB)	9	850*	1.4	98	85	Rectangular plan, roofed ceiling
Harty Room (School of Music, QUB)	18	1 150*	1.4	132	70	Stage, side choirs, roofed ceiling
Sonic Laboratory (SARC, QUB)	30	3 200	0.7	736	30	Rectangular, grid floor at 4 m
Whittla Hall (QUB)	8	8 400*	1.8	751	30	Rectangular plus stage house
Concert hall A (Concertgebouw Amsterdam)	420	19 000	2.6	1180	23	Horseshoe, columns, balconies
Concert hall B (De Doelen, Rotterdam)	512	24 000	2.3	1760	19	Irregular, inner shell

[a]TU=Delt University of Technology, QUB=Queen's University Belfast, and SARC=Sonic Arts Research Centre.
[b]An asterisk indicates that it has been determined from imcomplete dimension data and may be inaccurate.

caused by these conditions or the possibilities that arise from a relaxation thereof are discussed at various stages in the present paper.

II. DESCRIPTION OF ROOMS INVESTIGATED

A summary of all rooms included in the investigation is given in Table I. In Table I, V_{geo} is the value of the room volume obtained from architectural drawings and/or geometrical measurements. In the rooms with a suspended ceiling, the first figure is the measurement up to the acoustic ceiling and the figure in parentheses is the approximate volume up to the fixed ceiling.

In lecture hall A, the room impulse responses have been measured using the maximum-length sequence method with a polyhedral loudspeaker (designed to produce omnidirectional radiation over a wide frequency bandwidth) and a sampling frequency of 14 980 Hz. The measurements in concert halls A and B have been performed with the same polyhedral loudspeaker but using a logarithmic sweep and a sampling frequency of 16 kHz. In the remaining rooms, the measurements were performed using a logarithmic sweep with a Mackie HR 824 loudspeaker (a commercial studio monitor) as a sound source and a sampling frequency of 48 kHz. The measurement microphone for all rooms was the omnidirectional channel of a SoundField MKV microphone system.

In lecture hall A and concert halls A and B, the measurement positions described a line across the entire width of the hall and the offset parameter used in some of the figures is the distance from the central receiver position. In the remaining rooms a varying number of representative receiver positions have been selected. Only one source position was used in all rooms.

III. GEOMETRICAL ACOUSTICS METHOD

Within geometrical acoustics, sound waves are replaced by sound rays and reflected waves are replaced by (specular) reflections. The temporal density of reflections is given by[12]

$$\frac{dN_t}{dt} = 4\pi\frac{c^3t^2}{V}, \tag{1}$$

and therefore depends on time t, the room volume V, and the wave speed c in air. The wave speed does not vary considerably within a practical temperature range and can be considered to be known. Equation (1) can thus be rearranged to yield the volume from the temporal reflection density.

In a modeled RIR $h(t)$ the reflections can be identified because it is assumed here that they are represented by scaled Kronecker delta functions. The binary signal $h_{refl}(t)$ is then constructed from $h(t)$ in the following manner:

$$h_{refl}(t) = \begin{cases} 0, & \forall|h(t)| = 0 \\ 1, & \forall|h(t)| \neq 0. \end{cases} \tag{2}$$

As illustrated in Fig. 1(b) for the example RIR in Fig. 1(a), $h_{refl}(t)$ is essentially a pulse width modulated signal with the modulation density equal to the reflection density. An esti-

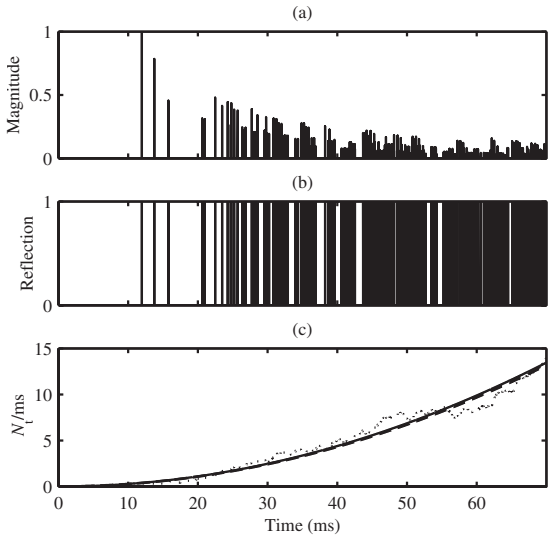


FIG. 1. (a) Modeled RIR $h(t)$, (b) identified reflections in $h_{refl}(t)$, and (c) estimated (dotted curve) and theoretical (solid curve) reflection density dN_t/dt. The almost coincident dashed curve is the least-squares fit to the dotted curve.

mate of the latter can be obtained after convolving $h_{\mathrm{refl}}(t)$ with a low pass moving average filter $f(t)$,

$$dN_t/dt \approx h_{\mathrm{refl}}(t) * f(t), \tag{3}$$

with the $(*)$ the convolution operator and the filter $f(t)$ of length T given by

$$f(t) = \begin{cases} 1/T, & \forall -T/2 \leqslant t \leqslant T/2 \\ 0, & \forall -T/2 > t > T/2. \end{cases} \tag{4}$$

The dotted curve in Fig. 1(c) shows the estimated reflection density, which clearly exhibits deviations from the theoretical behavior indicated by the solid curve. Because it is known that the density increases with the square of time, a least-squares fit can be applied to the estimated density and the result is the dashed curve in Fig. 1. For this particular example, there is virtually no difference between this curve and the theoretical reflection density given by the solid curve.

In order to assess the performance of this proposed volume estimation method, RIRs have been modeled with a mirror image source model (discrete time, frequency-independent reflection coefficient[13]) in rectangular rooms ranging in volume V from 10 to 10 000 m^3. The room dimensions (L_x, L_y, L_z) are given by

$$L_x = \varphi_x \left(\frac{V}{\varphi_x \varphi_y} \right)^{1/3}, \quad L_y = \varphi_y \left(\frac{V}{\varphi_x \varphi_y} \right)^{1/3}, \tag{5a}$$

$$L_z = \left(\frac{V}{\varphi_x \varphi_y} \right)^{1/3}, \tag{5b}$$

with φ_x a random variable with uniform distribution between 1.2 and 2 and φ_y a random variable with uniform distribution between 0.5 and 0.83. This procedure ensures that the aspect ratio of the room dimensions varies between 1.2:1:0.83 and 2:1:0.5. A similar procedure has been used for the positioning of the source and receiver within the room. The sampling frequency was 192 kHz and the reflections coefficient of all six walls was set to 0.6.

Due to the large range in room volumes, estimating the reflection density within a fixed time interval is prone to errors because small rooms have a very large density at the upper time limit and large rooms have a very small density at the lower time limit. The reflection density was thus estimated over a varying time interval defined by the arrival times of the first 500 reflections. The length T of the moving average filter was set to 20 ms. Informal experiments have shown that a variation by a factor of 2 on either side is acceptable.

Using these parameters, the room volume determined from the reflection density has been estimated for a total of 2000 modeled rooms. The mean error between true and estimated room volume was found to be 3.8% with a standard deviation of 5.7%. This numerical result shows that the room volume can be estimated fairly accurately under idealized conditions. The success of the method depends crucially on the reliable estimation of all individual reflections; this issue is now investigated further both theoretically and experimentally in the following two sections.

A. Resolution limit in the time domain

One necessary but not sufficient condition for the reflections to be represented by Kronnecker deltas is that the source impulse has infinite frequency bandwidth. In practice, this condition can never be met. Instead, a source impulse is now considered whose frequency response is of uniform magnitude and zero phase up to a maximum frequency $\omega_{\max}$. Its transfer function $H_{\mathrm{Sc}}(\omega)$ can thus be written as

$$H_{\mathrm{Sc}}(\omega) = \begin{cases} 1 & \text{for} |\omega| \leqslant \omega_{\max} \\ 0 & \text{for} |\omega| > \omega_{\max}. \end{cases} \tag{6}$$

From standard Fourier theory, its impulse response $h_{\mathrm{Sc}}(t)$ follows as

$$h_{\mathrm{Sc}}(t) = \frac{1}{\pi} \frac{\sin \omega_{\max} t}{t}. \tag{7}$$

The bandlimited RIR is then obtained by convolving $h(t)$, containing the scaled Kronnecker deltas, with $h_{\mathrm{Sc}}(t)$.

A condition is now required that specifies when two reflections arriving successively in time are separable and thus identifiable in the room impulse response. For this purpose, the Rayleigh resolution criterion is adopted.[14] It states that two impulses are barely resolved if the maximum of the first is located at the first zero of the second impulse. With $h_{\mathrm{Sc}}(t)$, this occurs when $\omega_{\max} \Delta t = \pi$ or

$$\Delta t = 1/2 f_{\max}, \tag{8}$$

where $f_{\max} = \omega_{\max}/2\pi$. An additional requirement is that the two impulses are $\pi/2$ out of phase. In the context of room acoustics, the phase differences are mainly caused by the imaginary part of the reflection coefficient of the walls and the directionally varying impulse response of source and receiver. Depending on these factors, the resolution criterion is either an over- or underestimate.

The inverse of the reflection density in Eq. (1) is the (average) time interval Δt between the arrival of successive reflections. Equating it with Δt from Eq. (8) results in a maximum time t_{up} up to which individual reflections are distinguishable,

$$t_{\mathrm{up}} = \sqrt{\frac{V f_{\max}}{2\pi c^3}}. \tag{9}$$

This result shows that the larger the volume V and maximum frequency $f_{\max}$, the larger the value t_{up} can assume.

Unfortunately, Eq. (1) and therefore also Eq. (9) are not directly applicable to RIRs measured in real rooms because the expressions have been derived for specular reflections and neglect the effects of scattering from rough surfaces and diffraction from finite-size surfaces. In order to circumvent this, $f_{\max}$ must be decreased to a, room dependent, value whose corresponding acoustic wavelength is much larger than the surface roughness and the size of the reflecting surfaces.

As practical examples, lecture hall A and concert hall A are considered and $f_{\max}$ is determined from a visual inspection of the room itself or photographs thereof. For lecture hall A, it is assumed that the smallest acoustic wavelength should be 30 cm and thus $f_{\max} \approx 1100$ Hz, which implies fur-

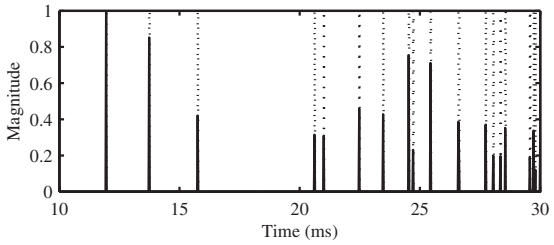


FIG. 2. Magnitude of modeled RIR $h(t)$ at the central receiver position in lecture hall A. The vertical dotted lines represent the signal $h_{\text{peaks}}(t)$.

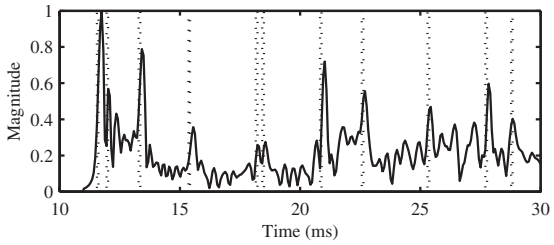


FIG. 3. Magnitude of measured RIR $h(t)$ at the central receiver position in lecture hall A. The vertical dotted lines represent the signal $h_{\text{peaks}}(t)$.

ther that $t_{\text{up}}=28$ ms and approximately 20 reflections should be distinguishable. For concert hall A, it is assumed that the smallest acoustic wavelength should be 2.0 m and thus $f_{\max}\approx 175$ Hz, which then implies that $t_{\text{up}}=110$ ms and approximately nine reflections should be distinguishable.

Note that the expression for t_{up} has been derived under the assumption that the real part of the reflection coefficient of all reflecting surfaces is approximately equal and that both source and receiver have omnidirectional directivity.

B. Identification of reflections in measured RIRs

In Eq. (2), the signal $h_{\text{refl}}(t)$ was formed by identifying the reflections as the only nonzero samples in the RIR $h(t)$. With measured RIRs this approach cannot be followed because the RIR magnitude is almost always nonzero. Instead, the extraction of the peaks in measured RIRs is now performed and it will then be considered whether the peaks correspond to the desired specular reflections. One way of extracting the peaks in a RIR is through adaptive thresholding known from image processing, see, e.g., Gonzales.[15] The rationale is that the magnitude of a peak is a factor ϵ above the magnitude average of a number of neighboring samples.

Using the mean as the method of averaging, the local magnitude mean at time t is given by

$$\mu_{\text{local}}(t)=\frac{1}{T_{\mu_{\text{local}}}}\int_{t-T_{\mu_{\text{local}}}/2}^{t+T_{\mu_{\text{local}}}/2}|h(\tau)|d\tau, \tag{10}$$

where $T_{\mu_{\text{local}}}$ is the averaging time. The binary signal containing the extracted peaks then follows as

$$h_{\text{peaks}}(t)=\begin{cases}0, & \forall h(t)<\epsilon\mu_{\text{local}}(t)\\ 1, & \forall h(t)\geqslant\epsilon\mu_{\text{local}}(t),\end{cases} \tag{11}$$

with ϵ the thresholding parameter. In applying this method to modeled and measured RIRs, the average was taken over $T_{\mu_{\text{local}}}=2$ ms and the threshold was set to $\epsilon=2$. Both values were determined experimentally and, as will be shown in the following, may not be optimal for all RIRs.

As a first step, the adaptive thresholding is applied to the RIR at the central receiver position in lecture hall A that was modeled with the mirror image source model. Together with the magnitude of the RIR, the result after thresholding is shown by the vertical dotted lines in Fig. 2 and it can be seen that all the present peaks corresponding to specular reflections are correctly identified. In this case, $h_{\text{peaks}}(t)=h_{\text{refl}}(t)$.

In Fig. 3, the result is shown if the same procedure is applied to the magnitude of the RIR measured at the same position in lecture hall A. Compared to Fig. 2, the number of extracted peaks is less. This is a consequence of the higher complexity of a measured RIR caused by the detailed impulse response of source and receiver, complex-valued, frequency-dependent reflection coefficients, the presence of nonspecular reflections, and other factors. On the other hand, the two identified peaks between 18 and 20 ms are not present in the modeled RIR. They are most likely caused by reflections from objects not modeled with the mirror image source model.

For a more comprehensive performance assessment, the peak extraction procedure has been performed on the RIRs measured across the width of lecture hall A and compared to the modeled RIRs at the same receiver positions. The two sets of RIRs are shown in Figs. 4(a) and 4(b), respectively. The result after applying adaptive thresholding to the magnitude of the measured RIRs is shown in Fig. 4(c). Compared to Fig. 4(a), it can be observed that the main features are reproduced. The deviations are again caused by differences between measured and modeled RIRs and in particular because some of the reflections in the measured RIR are weaker and more diffuse (mainly from the right sidewall)

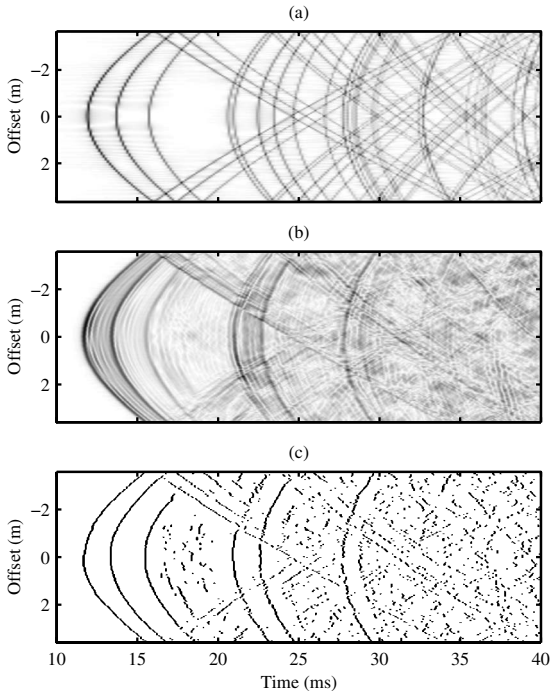


FIG. 4. RIR across the width of lecture hall A, (a) from model, (b) magnitude from measurements, and (c) result after applying adaptive thresholding to (b).

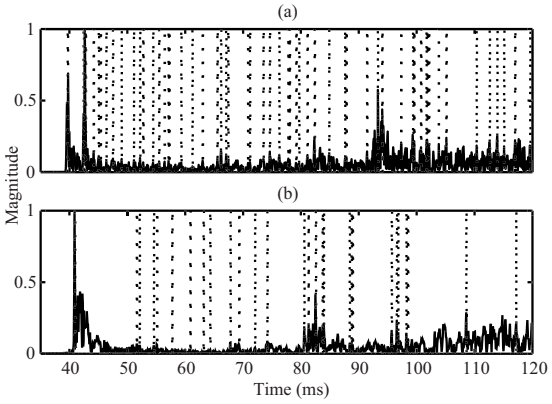


FIG. 5. Measured RIR in (a) concert hall A and (b) concert hall B with vertical dotted lines representing the signal $h_{\text{peaks}}(t)$.

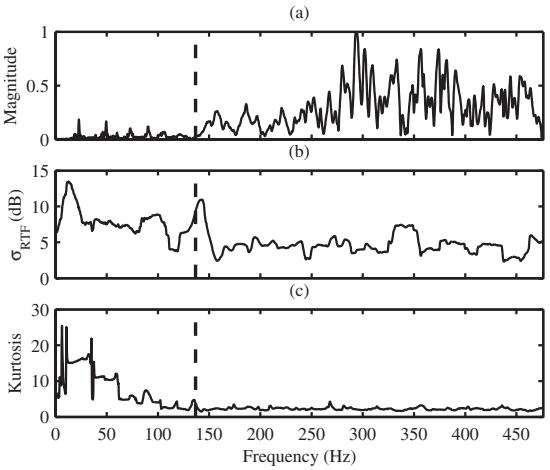


FIG. 6. (a) RTF magnitude in lecture hall A, (b) standard deviation of the logarithmic magnitude, and (c) kurtosis of the magnitude. The vertical dashed line indicates the theoretical Schroeder frequency.

and the measured RIR contains reflections from objects that are not present in the modeled RIR (mainly seen between 17 and 22 ms). Depending on the time interval in question, it seems fair to conclude that approximately 50% of the extracted peaks correlate with the specular reflections in Fig. 4(a). It has been contemplated in Sec. III A that for this room 20 reflections should be identifiable for a time interval up to 28 ms and the agreement is indeed best in this time interval. But the number of extracted peaks is less than 10.

Another issue is whether the adaptive thresholding parameters $T_{\mu_{\text{local}}}$ and ϵ found suitable for the RIRs in lecture hall A will yield usable results for other rooms as well. For this purpose, the same procedure has been applied to a measured RIR in both concert halls A and B and the results can be seen in Fig. 5. In both graphs the number of identified peaks is very large and most of them do not seem to correspond to specular reflections. From geometrical considerations, for concert hall A two sidewall reflections should arrive at 95 and 100 ms and for concert hall B one strong reflection should arrive at 85 ms and approximately another five reflections between 100 and 120 ms. The number of extracted peaks is therefore far too high. This observation is also supported by the theoretical result from Sec. III A that predicted for concert hall A nine identifiable reflections in the time interval up to 110 ms.

From these experimental results it is now concluded that systematically identifying specular reflections from measured RIRs is not a feasible approach in most rooms encountered in practice. Apart from the direct sound and depending on the acoustic characteristics of the room, at most one to five reflections can be extracted with confidence. As the room volume estimation through the reflection density explicitly relies on identifying the reflections, it is further concluded that this method is not a viable approach. It must also be noted that the theoretical expression for the reflection density is inaccurate for values of the time variable where the first few reflections arrive.[12]

IV. EIGENMODE METHOD

Whereas the approach for the room volume estimation presented in the previous section was based on the RIR in the time domain, the current section focuses on the room transfer function (RTF) in the frequency domain. Analogous to the reflection density, a possible approach for the room volume estimation is through the modal density. This requires the identification of individual eigenmodes and, since only the density of oblique modes is proportional to the room volume,[16] will need to be performed in a frequency region where rooms with at least a moderate amount of absorption do normally feature significant frequency overlap between neighboring eigenmodes (i.e., above the Schroeder frequency). For this reason, this approach will not be followed further in the current paper. Instead, it is noted that the Schroeder frequency is defined as[17,18]

$$f_{\text{Schroeder}} \approx 2000\sqrt{\frac{T_{60}}{V}} \tag{12}$$

and depends only on the room volume and the reverberation time T_{60}. The reverberation time can be estimated reliably and if it proves possible to estimate the Schroeder frequency experimentally from a RTF, the theoretical expression for the Schroeder frequency can be rearranged to yield an estimate of the room volume.

A. RTF statistics around the Schroeder frequency

Above the Schroeder frequency, the RTF magnitude statistics has a kurtosis of three (theoretical value for a Gaussian distribution) and a standard deviation of 5.57 dB.[19,17] It is anticipated that these two statistical parameters assume different values in the region below the Schroeder frequency and in order to verify this, they have been estimated on the RTFs in a sliding rectangular frequency window of 24 Hz width for a frequency range from 0 to 500 Hz. The measured RIRs were either zero-padded or truncated before transformation to the frequency domain such that the frequency resolution was always 0.3 Hz and therefore the statistics are estimated over 80 samples. The direct sound component was always included.

In Fig. 6, the results for lecture hall A show that both the standard deviation and the kurtosis fluctuate around the known statistical values above the Schroeder frequency,

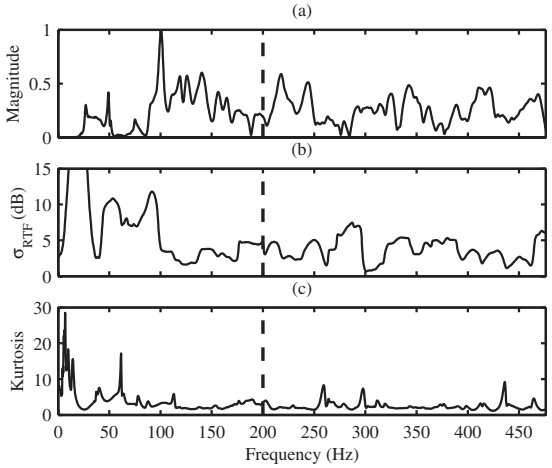


FIG. 7. (a) RTF magnitude in the office, (b) standard deviation of the logarithmic magnitude, and (c) kurtosis of the magnitude. The vertical dashed line indicates the theoretical Schroeder frequency.

whereas below it their values are significantly different. It can also be seen in Fig. 6(a) that the overall RTF magnitude is reduced below the Schroeder frequency, but this phenomenon was found to be caused by a reduced sound power output from the measurement loudspeaker. In Fig. 7, the results for the office (LG023 in Table I) show that the kurtosis fluctuates around the theoretical statistical value for the whole frequency bandwidth except at the very low end where the results are biased because less than the 80 frequency samples are available for the estimation of the statistics. The standard deviation exhibits a similar behavior except that it fluctuates around the theoretical value starting from 100 Hz. It is worth mentioning that analogous graphs from RTFs measured in other rooms showed that the estimated statistics often changes at a lower frequency than the theoretical Schroeder frequency.

B. Room volume from Schroeder frequency

The success of the room volume estimation method based on the Schroeder frequency is dependent on the accuracy with which the Schroeder frequency can be estimated from the RTF magnitude statistics and this poses the following challenges. The unbiased estimation of the magnitude statistics requires a width of the frequency window that incorporates more than a single magnitude peak or dip and consequently the width must be at least 20 Hz. As can be seen from Table I, the value of the Schroeder frequency in the larger rooms is of the same order and leaves no margin to estimate the magnitude statistics below it and many loudspeakers do also not produce sufficient sound power at these low frequencies. Further problems are that the magnitude statistics below the Schroeder frequency are not independent of source and receiver position and that a strong direct sound component has a bias effect on the magnitude statistics and may thus need to be excluded.

To ascertain whether the method can provide any affirmative results for the estimated room volume, its performance has been tested on all of the RTFs in nearly all of the available rooms. The experimentally observed Schroeder frequency was taken as the highest frequency sample where either the kurtosis was above 7 or the standard deviation was above 7 dB. These two values are slightly higher than their respective asymptotic values and were determined empirically by looking at graphs analogous to those shown in Figs. 6 and 7 from various RIRs. The numerical results in terms of mean and standard deviation between the receiver positions in each room are given in Table II and Fig. 8 shows a plot of the estimated versus the geometrical room volume.

TABLE II. Mean μ and relative standard deviation σ/μ of estimated volume V_{Sch} together with room volume V_{geo} from geometrical measurements. The numbers in parentheses following the name of the room indicate the fraction of receiver positions where a valid estimate was obtained.

Name	V_{geo} (m^3)	$\mu_{V_{\mathrm{Sch}}}$ (m^3)	$\sigma/\mu_{V_{\mathrm{Sch}}}$
Office (5/15)	60(80)	570	0.92
Listening room (0/10)	131	⋯	⋯
Lecture room A (61/143)	180(220)	533	0.96
Lecture room B (3/4)	550	15 000	0.11
McMordie Hall (5/9)	850	12 000	1.20
Harty Room (6/18)	1 150	7 000	1.78
Sonic Laboratory (12/30)	3 200	7 400	0.70
Whittla Hall (4/8)	8 400	38 000	0.96
Concert hall A (200/420)	19 000	20 000	0.94
Concert hall B (279/512)	24 000	906	0.92

From Table II, the method yields valid results at usually less than half of the receiver positions (valid means that the estimated Schroeder frequency is neither zero nor the maximum of the frequency range considered). Moreover, some of the estimates are either too large or too small, by almost two orders of magnitudes. Figure 8 shows that the correlation between estimated and true room volume is poor. Taking into account that it can already be guessed that the volume of the rooms encountered in practice is in the range between 5 and 40 000 m^3, it is concluded that this method yields neither consistent nor useful results.

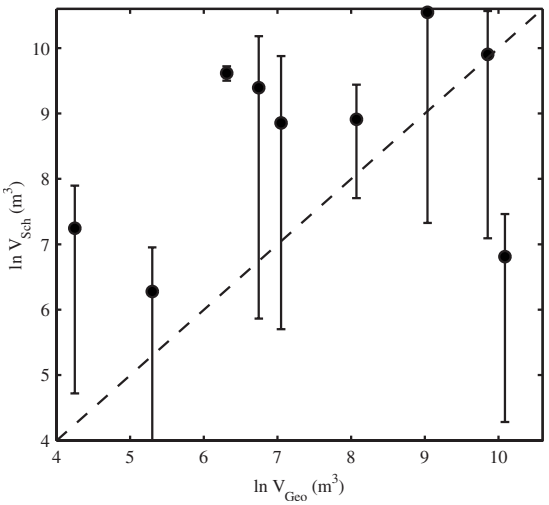


FIG. 8. The logarithm of the mean of V_{Sch} vs V_{Geo} with the error bars corresponding to the standard deviation between the receiver positions in each room. The dashed line indicates $V_{\mathrm{Sch}}=V_{\mathrm{Geo}}$.

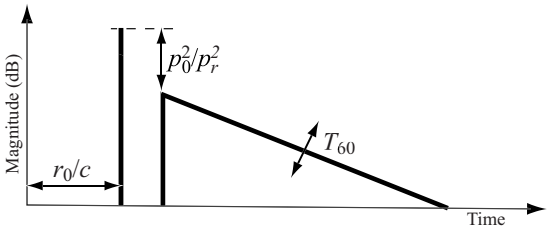


FIG. 9. Diagrammatic representation of the logarithmic magnitude in a RIR in terms of the arrival time of the direct sound r_0/c, the direct-to-reverberant ratio p_0^2/p_r^2, and the reverberation time T_{60}.

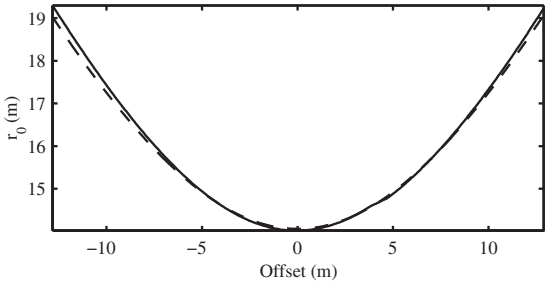


FIG. 10. Source-to-receiver distance r_0 across the width of concert hall B measured geometrically (---) and estimated acoustically (—).

V. DIFFUSE FIELD METHOD

From an energetic perspective, the logarithmic magnitude of a RIR can be described by the arrival time of the direct sound r_0/c, the reverberation time T_{60}, and the direct-to-reverberant ratio p_0^2/p_r^2 as illustrated in Fig. 9. If the source-to-receiver distance r_0 and the acoustic properties of the walls are kept constant when moving from a small room to a larger room, it seems logical to expect that the ratio p_0^2/p_r^2 would increase due to the decrease in reverberant energy density per unit volume. This observation is now formalized mathematically and forms the essence of the room volume estimation method based on diffuse field acoustics.

A. Theoretical basis

In a diffuse field, the mean square pressure of the reverberant sound field is given by[20]

$$\overline{p_r^2} = \rho_0 c W\left(\frac{T_{60}c}{6\ln(10)V}\right), \tag{13}$$

where $\rho_0 c$ is the specific acoustic impedance of air and W the sound power. The mean square reverberant pressure can be calculated from a RIR by summing all squared amplitudes therein but leaving out the direct sound component. The room volume can therefore be obtained from Eq. (13) but requires knowledge of W. This quantity is usually unknown. To circumvent this problem, Eq. (13) is accompanied by the equation for the mean square direct sound pressure given by[20]

$$\overline{p_0^2}(r_0) = \rho_0 c W\left(\frac{1}{4\pi r_0^2}\right), \tag{14}$$

where r_0 is the source-to-receiver distance. Again, the mean square pressure of the direct sound can be obtained from the squared amplitudes in the RIR.

When combining Eq. (14) with Eq. (13), the following expression for the room volume results:

$$V_{\text{classic}} = \frac{\overline{p_0^2}(r_0)}{\overline{p_r^2}}\frac{4\pi r_0^2 c T_{60}}{6\ln(10)}. \tag{15}$$

Alternatively, the reverberant pressure according to revised diffuse field theory is given by[1,2]

$$\overline{p_r^2} = \rho_0 c W\left(\frac{T_{60}c}{6\ln(10)V}\right)[e^{-r_0/c6\ln(10)}], \tag{16}$$

which leads to a slightly different alternative equation for the room volume given by

$$V_{\text{revised}} = \frac{\overline{p_0^2}(r_0)}{\overline{p_r^2}}\frac{4\pi r_0^2 c T_{60}}{6\ln(10)}[e^{-r_0/c6\ln(10)}]. \tag{17}$$

Moreover, in order to take the directivity of source and receiver into account, Eq. (14) has to be modified as follows:

$$\overline{p_0^2}(r_0) = Q_{\text{src}}(\mathbf{r}_0)Q_{\text{rec}}(-\mathbf{r}_0)\rho_0 c W\left(\frac{1}{4\pi r_0^2}\right), \tag{18}$$

with $Q_{\text{src/rec}}(\pm\mathbf{r}_0)$ the directivity factor of the source/receiver in the direction of the receiver/source, respectively. For the reverberant sound pressure, the directivity factors do not need to be included because they are by definition unity. Equation (18) does of course also alter Eqs. (15) and (17).

B. Calculation of the required parameters

The calculation of the room volume according to Eq. (15) or Eq. (17) requires parameter values for r_0, p_0^2, p_r^2, T_{60} and also in principle $Q_{\text{src}}(\mathbf{r}_0)$ and $Q_{\text{rec}}(-\mathbf{r}_0)$. The robust automatic estimation of each parameter from a RIR is now briefly discussed. Numerical results for each parameter are illustrated for RIRs measured across the width of concert hall B.

In a correctly measured RIR, the source-to-receiver distance r_0 can be estimated from the initial delay $\tau_0 = r_0/c$ of the direct sound arrival time. In the estimation procedure, τ_0 was assigned the value of the first time sample whose magnitude was less than 22 dB below the maximum magnitude in the entire RIR. This measure is necessary because the direct sound does not always correspond to the largest magnitude in the RIR. The very good agreement between geometrically measured and estimated r_0 is shown in Fig. 10. A general offset between the two curves is evident but this may also have been caused by incorrect geometrical measurements of r_0. More important, the estimated r_0 is consistent and shows no outliers.

For the estimation of the direct sound pressure p_0^2, the squared amplitudes in a time window, extending from −1 to 1.5 ms relative to the identified direct sound arrival time, are summed. For a better temporal resolution, the data have been resampled to 64 kHz. The estimated magnitude

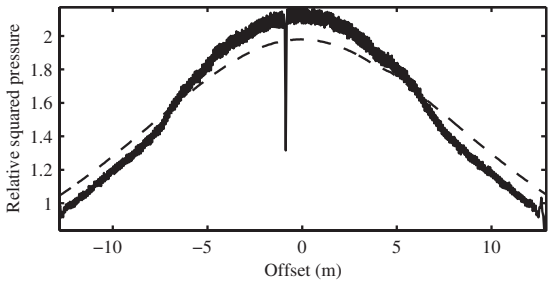


FIG. 11. Magnitude of the squared direct sound pressure p_0^2 across the width of concert hall B from the theoretical inverse square law (---) and estimated acoustically (—). The outlier slightly to the left of the zero offset is due to a measurement anomaly.

and the theoretical behavior according to the inverse square law are shown in Fig. 11. Because of an arbitrary scaling value, the two curves have been rescaled for equal mean. The agreement between the two curves is quite good even though there is a systematic deviation at the central half of the receiver positions.

For the calculation of the reverberant pressure and the reverberation time, the method by Lundeby *et al.*[21] has been used to find the crossover point between the sound decay and the stationary noise floor and T_{60} is then obtained from a straight line fit to the logarithm of the energy decay curve obtained from Schroeder backwards integration.[22] Figure 12 shows the estimated value of T_{60} across the width of concert hall B, which demonstrates that the reverberation time does fluctuate with receiver position. The reverberant pressure has been obtained by summing the squared amplitudes in the RIRstarting from the end of the time window used for the direct sound pressure and stopping at the crossover point found by Lundeby's method. Figure 13 shows the estimated value of p_r^2 across the width of concert hall B and illustrates that this parameter does also fluctuate with receiver position.

Finally, no procedure has been found to estimate the directivity factor of the source and receiver from the given input data and it has thus been assumed that $Q_{src}(\mathbf{r}_0) = Q_{rec}(-\mathbf{r}_0) = 1$. It needs to be mentioned that it is possible to obtain the directivity data by other means such as from the manufacturer's datasheet or from measurements in an anechoic chamber. Because the directivity data are required for a arbitrary three-dimensional direction and for a wide frequency range, the present author has decided to not pursue this avenue further because it would severely limit the method's applicability to laboratory experiments.

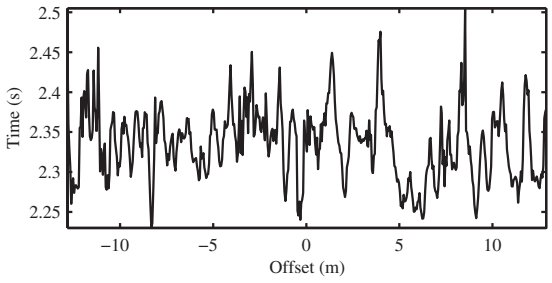


FIG. 12. Estimated reverberation time T_{60} across the width of concert hall B. The relative standard deviation between receiver positions is 0.02.

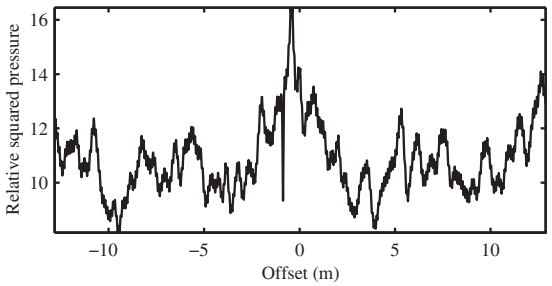


FIG. 13. Estimated magnitude of the squared reverberant sound pressure p_r^2 across the width of concert hall B. The relative standard deviation between receiver positions is 0.12.

C. Results

The performance of the room volume estimation method has been evaluated with all RIRs measured in the twelve rooms listed in Table I. Initial results obtained in the different rooms showed that with the polyhedral loudspeaker the volume estimates were of the correct order, whereas with the studio monitor all estimates were by an approximate factor ten too large. This fact was attributed to the difference in directivity between the two sound sources. For the studio monitor, the directivity factor in the direction of the receiver significantly exceeds unity at higher frequencies. This results in an increased direct sound pressure and, when inserting into Eq. (15) or Eq. (17), also in an overestimated room volume. The problem has been circumvented by restricting the frequency bandwidth to frequencies where the source radiation is approximately omnidirectional. Experiments with several loudspeakers have shown that an upper frequency limit of 700 Hz is appropriate. At the same time, a lower frequency limit of 200 Hz, higher than the Schroeder frequency in most rooms, was also introduced. It will become evident in Sec. V E that the bandwidth is detrimental to the accuracy of the room volume estimate.

After incorporating these measures, Table III shows the numerical results for volume estimates $V_{classic}$ and $V_{revised}$ from classical and revised diffuse field theory in terms of mean and standard deviation between all N_R receiver positions in a room. Quoting the values for mean and standard deviation is based on the assumption that the data follow a normal distribution. Performing the Kolmogorov–Smirnov test[23] on the data from lecture hall A and concert hall B proved that this assumption is justified. Naturally, the result obtained with a low number of receiver positions are statistically less representative. For concert hall B, Fig. 14 illustrates the variation of $V_{revised}$ and $V_{classic}$ across the width of the room.

Excluding the results for concert hall A, the office, the Sonic Laboratory and the lavatory, $\mu_{V_{revised}}$ is generally within ±50% of V_{geo}. This good correlation is perhaps better visualized by plotting $V_{revised}$ against V_{geo} as shown in Fig. 15, which is also to be compared with Fig. 8 for the room volume estimation method based on the Schroeder frequency. Even though $\mu_{V_{revised}}$ is generally closer to V_{geo} than $\mu_{V_{classic}}$ is, in most rooms revised theory produces only marginally more consistent results than classical theory as evident by the comparable standard deviation. Without the four

TABLE III. Mean μ and relative standard deviation σ/μ of estimated volumes V_{classic} and V_{revised} together with room volume V_{geo} from geometrical measurements.

Name	V_{geo} (m^3)	$\mu_{V_{\text{classic}}}$ (m^3)	$\sigma/\mu_{V_{\text{classic}}}$	$\mu_{V_{\text{revised}}}$ (m^3)	$\sigma/\mu_{V_{\text{revised}}}$
Lavatory	5(7)	14	0.28	11	0.27
Office	60(80)	231	0.28	161	0.27
Listening room	131	206	0.44	130	0.39
MMedia room	150	303	0.33	213	0.36
Lecture hall A	180(220)	297	0.20	248	0.21
Lecture hall B	550	714	0.22	539	0.19
McMordie Hall	850	1 190	0.29	1 020	0.31
Harty Room	1 150	1 480	0.31	1 130	0.33
Sonic Laboratory	3 200	3 340	0.52	2 020	0.44
Whittla Hall	8 400	14 500	0.24	9 450	0.19
Concert hall A	19 000	5 460	0.25	4 540	0.23
Concert hall B	24 000	26 900	0.14	20 400	0.16

exceptions already mentioned, it can be concluded that the room volume estimation method based on diffuse field acoustics delivers consistent results for rooms ranging in size from 100 m^3 to 20 000 m^3, i.e., over nearly three orders of magnitude, and having substantially different acoustic characteristics.

D. Exception to the general result trend

The results presented in Table III have shown that concert hall A, the office, the Sonic Laboratory, and the lavatory are exceptions. In all of these rooms $\mu_{V_{\text{revised}}}$ differs by more than 50% from V_{geo}. In the office and the lavatory, the Schroeder frequency is equal to or larger than the lower frequency limit. Experiments revealed that in these two rooms $\mu_{V_{\text{revised}}}$ moves closer to V_{geo} when the frequency limit is increased. But it is shown further below that decreasing the frequency bandwidth has the detrimental effect of increasing the variance in the volume estimates.

In the Sonic Laboratory, $\mu_{V_{\text{revised}}}$ is slightly below −50% from V_{geo} and $\sigma/\mu_{V_{\text{revised}}}$ is the highest value of all rooms investigated. A particular design feature of this room is that the performers and the audience are located on a metal grid floor below which there is an undercroft of substantial volume. This raises the question whether the room should be treated as a single volume or two coupled volumes. Due to this issue, the homogeneity of the sound field is questionable.

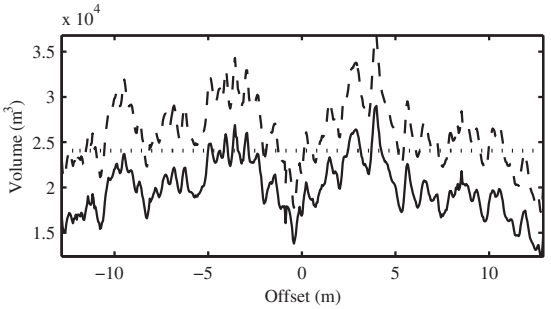


FIG. 14. Estimated room volumes V_{revised} (—) and V_{classic} (---) across the width of concert hall B. The true room volume is indicated by the dotted horizontal line.

In concert hall A, $\mu_{V_{\text{revised}}}$ is a factor of 4 too small and this result is consistent as shown by the low value for $\sigma/\mu_{V_{\text{revised}}}$. The estimates for r_0 and the T_{60} have been checked and found to yield plausible values, which leaves the squared pressure of the direct and reverberant sound. Cross referencing the ratio between the two with the same quantity in concert hall B of similar volume and complexity, it was found that the ratio is far too small for the same source-to-receiver distance. The most likely cause seems to be an uncharacteristically low magnitude of the direct sound but a further investigation into this anomaly was not possible because the data were not measured by the author and no further knowledge of the measurement setup was available. By comparing with the floor reflection, it was estimated that the squared direct sound pressure should theoretically be larger by an approximate factor three. This would increase the room volume estimate by the same factor and consequently bring it close to the value for V_{geo}.

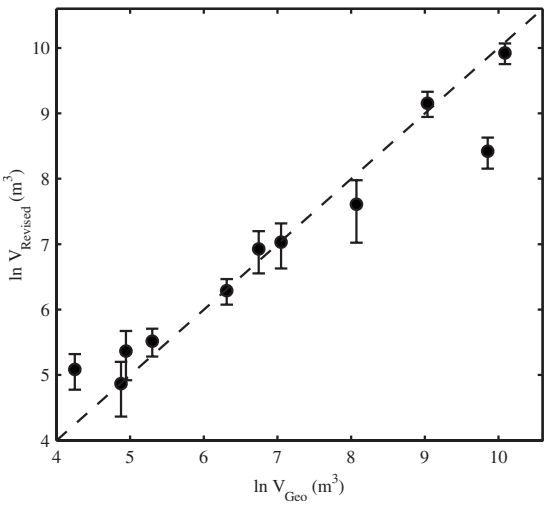


FIG. 15. The logarithm of the mean of V_{revised} vs V_{geo} with the error bars corresponding to the standard deviation between the receiver locations in each room. The dashed line indicates $V_{\text{revised}}=V_{\text{geo}}$.

E. Variance in the estimated parameters

It is now attempted to explain the variance of the room volume estimates in Table III in terms of both the theoretical and experimentally observed variance of the parameters involved in the calculation.

1. Variance in reverberant sound pressure and reverberation time

As mentioned in Sec. IV A, the standard deviation of the sound pressure for a single frequency is 5.57 dB. This value is decreased if the standard deviation is measured in a frequency band. Both Schroeder[24] and Lubman[25] considered this case and arrived at the same approximate equation for the standard deviation $\sigma_{\overline{p_r^2}}$ of the reverberant sound pressure level

$$\sigma_{\overline{p_r^2}} \approx \frac{5.57}{\sqrt{1+\frac{3.3BT_{60}}{13.8}}}\ (\mathrm{dB}), \tag{19}$$

where B is the equivalent bandwidth.[24] Equation (19) exhibits the correct asymptotic behavior in as much that for $B \to 0$ the single frequency value of 5.57 dB is obtained and for $B \to \infty$ the standard deviation tends to zero. Chiles and Barron[2] found the scatter of the sound pressure level in a scale and computer model to be larger than theoretically predicted by Eq. (19) but the agreement improves if only the late reverberant sound is considered.

An approximate expression for the relative standard deviation of the reverberation time can be obtained from formulas derived in Refs. 26 and 27 and is given by[28]

$$\frac{\sigma_{T_{30}}}{T_{30}} \approx \frac{0.59}{\sqrt{BT_{30}}}. \tag{20}$$

Here, T_{30} instead of T_{60} is used because the decay was measured over 30 dB and B is now the statistical bandwidth of the RIR or the bandpass filter. [29]

With the bandpass filter used in the current paper to attenuate the spectrum at frequencies below the Schroeder frequency and above the frequency where the source radiation is no longer omnidirectional, the equivalent or statistical bandwidth $B \approx 540$ Hz. Converting the dB value from Eq. (19) into linear units, the predicted relative standard deviation of the reverberant sound pressure for the rooms listed in Table III is in the range $\sigma_{\overline{p_r^2}}/\overline{p_r^2} \approx 0.06-0.21$ (depending on the reverberation time of the room in question). Similarly, Eq. (20) predicts that $\sigma_{T_D}/T_D \approx 0.02$ to 0.05. For concert hall B, the relevant experimental values, quoted in the captions of Fig. 13 and Fig. 12, are 0.12 and 0.02, respectively. Both of these values are in the respective range predicted by theory.

The values for the standard deviation of the reverberation time and reverberant sound pressure do account for part of the experimentally observed standard deviation in the estimated room volume and they do explain why the rooms with larger reverberation time mostly exhibit a smaller deviation in the room volume estimates. The combined theoretical standard deviation due to the T_{60} and $\overline{p_r^2}$ is given by[30]

$$\frac{\sigma_{V_{\mathrm{theory}}}}{V} = \sqrt{\frac{\sigma^2_{\overline{p_r^2}}}{(\overline{p_r^2})^2} + \frac{\sigma^2_{T_{60}}}{(\overline{T}_{60})^2} - 2\frac{\sigma_{\overline{p_r^2},T_{60}}}{\overline{p_r^2}\overline{T}_{60}}}, \tag{21}$$

where $\sigma_{\overline{p_r^2},T_{60}}$ is the covariance between the parameters $\overline{p_r^2}$ and T_{60} that will be nonzero because the variation in both parameters stems from the same physical wave phenomena. It has been found that the experimentally observed $\mu_{V_{\mathrm{revised}}}$ for any room listed in Table III is always underpredicted by the value given by Eq. (21). On the bandwidth issue, the volume estimation for the RIRs from concert hall B has been repeated with an upper frequency limit of 10 kHz and the result was a reduction of $\sigma/\mu_{V_{\mathrm{revised}}}$ to 0.10.

2. Variance in source-to-receiver distance and direct sound pressure

It was already concluded from Fig. 10 that the error in estimating the source-to-receiver distance from the initial time delay r_0/c is fairly small. The variation in the estimated squared direct sound pressure can be obtained by measuring the relative standard error between the two curves in Fig. 11; its is 0.06 for concert hall B. This error would need to be incorporated into Eq. (21) by an extra term under the square root. It seems reasonable to assume that this term is independent of the other terms due to the reverberation time and the reverberant sound pressure.

VI. CORRELATION BETWEEN ROOM VOLUME AND T_{60}

Apart from the directivity issue, the main drawback of the room volume estimation method presented in Sec. V is that it relies on correctly measured RIRs where the initial time delay can be used to calculate the source-to-receiver distance. In this section, it is investigated whether an approximate relationship between the room volume and the reverberation time can be used for a simplified estimation of the room volume.

Suppose it can be assumed that the surface area S of a room is related to the volume by $S=\beta V^{2/3}$. The minimum value of $\beta=6$ results for a cube and for a room with an aspect ratio of 2:1:1, $\beta=6.3$. In the following it is assumed that $\beta=6.4$ is an average representative value for the rooms encountered in practice. The Sabine equation for the reverberation time then reads

$$T_{60} \approx \frac{0.161V}{6.4\bar{\alpha}V^{2/3}} = \frac{0.161V^{1/3}}{6.4\bar{\alpha}}. \tag{22}$$

In Fig. 16, the average T_{60} per room in the 500 Hz octave band is plotted as a function of the natural logarithm of the room volume for all the rooms measured. The solid curve is the linear least-square fit of Eq. (22) for all the rooms indicated by the filled circles. It results in an average absorption value of $\bar{\alpha}=0.26$. The rooms excluded thus either have an uncharacteristically low or high value for $\bar{\alpha}$.

Alternatively, the dashed line in Fig. 16 is the linear least-square fit of a straight line through the plotted data. Its approximate equation is

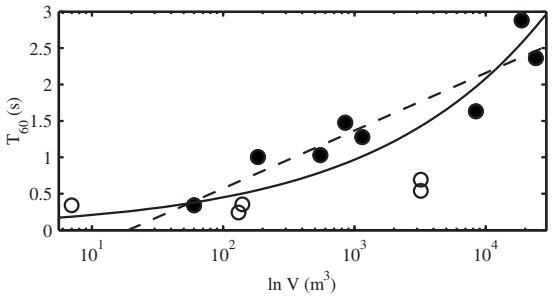


FIG. 16. T_{60} in the 500 Hz octave band vs the logarithm of the room volume V. The solid curve is the least-squares fit of Eq. (??) through all the closed circles. The open circles indicate the volume in rooms with particular acoustics. These are in order of increasing volume: Lavatory, office, the listening room, and the Sonic Laboratory. The dashed line is the linear least-squares fit $T_{60} \propto \ln V$.

$$T_{60} \approx 0.34 \ln(V) - 1. \tag{23}$$

Table IV shows the results after rearranging Eq. (23) to yield the volume from the reverberation time in the 500 Hz octave band (the 1 kHz octave band can similarly be used) for all rooms investigated. As expected, the room volume results show fairly large errors for those rooms which have not been included in the least-squares fit. Except for the Sonic Laboratory and the lavatory, the results are, however, mostly of the correct order of magnitude. A brief comparison between using Eq. (22) or Eq. (23) with $\bar{\alpha}=0.26$ for the room volume revealed that Eq. (23) yielded more accurate results in the Harty Room, McMordie Hall, and lecture room A whereas the results were slightly more inaccurate in concert hall A, the Sonic Laboratory and the Whittla Hall.

A. Source-to-receiver distance estimation

Due to the promising room volume results in Table IV, it seems logical to reconsider Eq. (15) and investigate to what extent it can be used to estimate the source-to-receiver distance. Inserting Eq. (23) into Eq. (15) and rearranging for r_0 yields

$$r_0' \approx \sqrt{\frac{\overline{p_r^2}}{\overline{p_0^2}(r_0)} \frac{6 \ln(10)}{4\pi c T_{60}} e^{(T_{60}-1)/0.34}}. \tag{24}$$

The last two columns of Table IV show the mean and standard deviation of the relative difference between r_0' and the true source-to-receiver distance r_0 in each room. In just under half of the rooms this procedure results in a mean error of approximately 20%. Larger errors of 70% result in rooms where the value for the volume was already inaccurate. The result for concert hall A is of course an exception because of issues already discussed in Sec. V D. Note that one reason for the small relative errors is the square root operation in Eq. (24).

VII. CONCLUSION

The estimation of the room volume from a given room impulse response has been investigated. The measurement data were obtained from several receiver positions in a total of 12 rooms of varied size and acoustic characteristics. The investigated methods were based on geometrical acoustics, eigenmode, and diffuse field models. It was found that the estimation method based on the temporal reflection density fails because of the difficulty in identifying reflections. The estimation method based on the Schroeder frequency was found to deliver inconsistent results because of the difficulty in experimentally determining the Schroeder frequency.

The estimation method based on diffuse field acoustics was found to deliver consistent results for room ranging in volume over almost three orders of magnitude. With the exceptions of the results in four rooms, the average of the estimated room volume between receiver positions was within $\pm 50\%$ of the true room volume with a standard deviation of approximately 0.25. The deviant results in the four excepted rooms were explained by a measurement error, a particularly high Schroeder frequency and a particular room design feature. A part of the experimentally observed standard deviation was explained by theoretical expressions for the standard deviation of the reverberant sound pressure and the reverberation time.

TABLE IV. Mean μ and relative standard deviation σ/μ of estimated volume from inverting Eq. (23) in the 500 Hz octave band. The last two columns express the error between the true r_0 and estimated source-to-receiver distance r_0', where $\epsilon_{r_0}=|r_0-r_0'|/r_0$.

Name	V_{geo} (m^3)	μ_V (m^3)	σ/μ_V	$\mu_{\epsilon_{r_0}}$	$\sigma_{\epsilon_{r_0}}$
Lavatory	5(7)	51	0.10	0.78	0.31
Office	60(80)	50	0.04	0.41	0.08
Listening room	131	38	0.04	0.46	0.16
MMedia room	150	53	0.06	0.40	0.10
Lecture room A	180(220)	287	0.07	0.12	0.11
Lecture room B	550	378	0.13	0.14	0.19
McMordie Hall	850	1 370	0.08	0.22	0.21
Harty Room	1 150	784	0.16	0.13	0.19
Sonic Laboratory	3 200	120	0.23	0.71	0.12
Whittla Hall	8 400	2 180	0.10	0.50	0.08
Concert hall A	19 000	70 000	0.43	10.8	6.3
Concert hall B	24 000	22 560	0.09	0.13	0.08

Two drawbacks of this method are that it (i) relies on the presence of the initial time delay in the room impulse response for the estimation of the source-to-receiver distance and (ii) essentially assumes omnidirectional source and receiver. It was found that estimating the room volume from an approximate relationship with the reverberation time does not suffer from these drawbacks but does only provide accurate results for rooms whose absorption is neither uncharacteristically low nor high.

[1]M. Barron and L. J. Lee, "Energy relations in concert auditoriums. I," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[2]S. Chiles and M. Barron, "Sound level distribution and scatter in proportionate spaces," J. Acoust. Soc. Am. **116**, 1585–1595 (2004).
[3]D. Cabrera, D. Jeong, H. J. Kwak, and J.-Y. Kim, "Auditory room size perception for modeled and measured rooms," in Proceedings of the 2005 Congres and Exposition on Noise Control Engineering, Rio, 2005.
[4]C. B. Pop and D. Cabrera, "Auditory room size perception for real rooms," in Proceedings of ACOUSTICS 2005, Australian Acoustics Society, Busselton, 2005.
[5]M. Kuster, D. de Vries, E. M. Hulsebos, and A. Gisolf, "Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response," J. Acoust. Soc. Am. **116**, 2126–2137 (2004).
[6]M. Monks, B. M. Oh, and J. Dorsey, "Audioptimization: Goal-based acoustic design," IEEE Comput. Graphics Appl. **20**, 76–91 (2000).
[7]A. T. Fürjes, F. Augusztinovicz, and E. Arató-Borsi, "A new method for the objective qualification of rooms," Acta Acust. **86**, 911–918 (2000).
[8]E. M. Hulsebos, "Auralization using wave field synthesis," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 2004.
[9]J. van der Vorm, "Transform coding of audio impulse responses," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2003.
[10]F. Li and T. J. Cox, "Speech transmission index from running speech: A neural network approach," J. Acoust. Soc. Am. **113**, 1999–2008 (2003).
[11]R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," J. Acoust. Soc. Am. **114**, 2877–2892 (2003).
[12]H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 2000), Chap. 4.
[13]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).
[14]E. Hecht, *Optics*, 2nd ed. (Addision-Wesley, Wokingham, 1987), Chap. 10.
[15]R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 1st ed. (Addison-Wesley, Reading, Mass., 1992), Chap. 7.
[16]H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 2000), Chap. 3.
[17]M. Schroeder, "Statistical parameters of the frequency response curves of large rooms," J. Audio Eng. Soc. **35**, 299–306 (1987).
[18]M. Schroeder, "Normal frequency and excitation statistics in rooms: Model experiments with electric waves," J. Audio Eng. Soc. **35**, 307–316 (1987).
[19]Valid for the case where the receiver is sufficiently far away from the source in terms of direct-to-reverberant energy ratio.
[20]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), Chap. 13.
[21]A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of measurements in room acoustics," Acustica **81**, 344–355 (1995).
[22]M. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).
[23]A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the theory of statistics*, McGraw–Hill Series in Probability and Statistics, 3rd ed. (McGraw–Hill, New York, 1974).
[24]M. R. Schroeder, "Effect of frequency and space averaging on the transmission responses of multimode media," J. Acoust. Soc. Am. **46**, 277–283 (1969).
[25]D. Lubman, "Fluctuations of sound with position in a reverberant room," J. Acoust. Soc. Am. **44**, 1491–1502 (1968).
[26]J. L. Davy, I. P. Dunn, and P. Dubout, "The variance of decay rates in reverberation rooms," Acustica **43**, 12–25 (1979).
[27]J. L. Davy, "The variance of impulse decays," Acustica **44**, 51–56 (1980).
[28]M. Barron, "Thoughts on the room acoustic enigma: The state of diffusion," *Proceedings of the Institute of Acoustics*, Oxford, 2005, Vol. 27.
[29]J. L. Davy and I. P. Dunn, "The statistical bandwidth of Butterworth filters," J. Sound Vib. **115**, 539–549 (1987).
[30]H. Ku, "Notes on the use of propagation of error formulas," J. Res. Natl. Bur. Stand., Sect. C **70**, 263–273 (1966).

Noise reduction combining time-frequency ε-filter and M-transform

Tomomi Abe[a)]
Major in Pure and Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Mitsuharu Matsumoto[b)] and Shuji Hashimoto[c)]
Department of Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

(Received 4 December 2007; revised 13 May 2008; accepted 20 May 2008)

This paper introduces noise reduction combining time-frequency ε-filter (TF ε-filter) and time-frequency M-transform (TF M-transform). Musical noise is an offensive noise generated due to noise reduction in the time-frequency domain such as spectral subtraction and TF ε-filter. It has a deleterious effect on speech recognition. To solve the problem, M-transform is introduced. M-transform is a linear transform based on M-sequence. The method combining the time-domain ε-filter (TD ε-filter) and time-domain M-transform (TD M-transform) can reduce not only white noise but also impulse noise. Musical noise is isolated in the time-frequency domain, which is similar to impulse noise in the time domain. On these prospects, this paper aims to reduce musical noise by improving M-transform for the time-frequency domain. Noise reduction by using TD M-transform and the TD ε-filter is first explained to clarify its features. Then, an improved method applying M-transform to the time-frequency domain, namely TF M-transform, is described. Noise reduction combining the TF ε-filter and TF M-transform is also proposed. The proposed method can reduce not only high-level nonstationary noise but also musical noise. Experimental results are also given to demonstrate the performance of the proposed method.
 [DOI: 10.1121/1.2940584]

PACS number(s): 43.60.Ac, 43.60.Wy, 43.60.Hj [EJS] Pages: 994–1005

I. INTRODUCTION

Noise reduction plays an important role in acoustical signal processing. There are two types of approaches for noise reduction: one is to use multiple inputs and the other is to use a single input. A typical approach using multiple inputs is microphone array.[1–3] There are many studies based on microphone array such as the delay-sum type microphone array[4] and the adaptive microphone array.[5,6] However, they require large-size systems because they use multiple inputs. They also need a precise estimation of the position and directivity of the microphones. There are also blind source separation techniques using multiple inputs such as independent component analysis (ICA)[7,8] and sparseness approach.[9–11] ICA can separate mixed sounds by maximizing statistical independence among them. However, calculation cost is high because higher order statistics are used. In sparseness approaches, we assume that the sources rarely overlap in the frequency domain. Under this assumption, it is possible to extract each signal using time-frequency binary masks. However, due to this assumption, sparseness approaches cannot handle sounds that overlap in the frequency domain. Both ICA and sparseness approaches also need multiple signals as in microphone arrays to separate sounds.

On the other hand, if we can effectively reduce the noise of a single input, there will be a wide range of applications. It will also be easy to reduce the system size because only one signal is required. Some authors have reported model-based approaches to reduce noise.[12] In these approaches, we can extract the specific sound by determining the sound model in advance. However, it is not possible to apply the method to the signals with unknown noise. There are other approaches that use a comb filter.[13] In these approaches, the pitch of the speech signal is estimated, and noise is reduced with a comb filter. However, estimation error results in the degradation of noise reduction performance. Although noise reduction by using sophisticated filters such as the Kalman filter and the iterative Wiener filter[14–17] has also been reported, the calculation cost to estimate signal model is high. In image processing, a median filter is often used for noise reduction.[18,19] However, when we use median filters for speech signal processing, the attack and decay characteristics of speech signals are also smoothed because speech signals change frequently, while the change occurs only near the edge in the image.

To solve the problem, the authors proposed a nonlinear filter called the time-frequency ε-filter (TF ε-filter).[20] An ε-filter is a nonlinear filter in the time domain, which can reduce noise while preserving speech signals.[21,22] In spite of its simple design, the ε-filter has some desirable features for noise reduction and is executed in the time-domain. It does not need noise estimation in advance. It can reduce not only

[a)]Electronic mail: tomomi@shalab.phys.waseda.ac.jp
[b)]Electronic mail: matsu@shalab.phys.waseda.ac.jp
[c)]Electronic mail: shuji@waseda.jp

stationary noise, but also nonstationary noise. It is easy to design and calculation costs are small. A TF ε-filter is an improved version of an ε-filter. Using a TF ε-filter, we can reduce not only relatively small-amplitude noise but also relatively large-amplitude noise. It can also reduce not only stationary noise but also nonstationary noise. It is not necessary to estimate the noise information in advance as is the case with the conventional ε-filter labeled TD ε-filter. However, since the TF ε-filter is a frequency domain filter, some musical noise may occur. It is well known that musical noise also becomes a problem in some other methods in frequency domain such as spectral subtraction (SS).[23–25] Our aim is to reduce musical noise effectively while preserving the performance of noise reduction. Although there are some approaches for reducing musical noise by utilizing the properties of the human auditory system,[26] it is difficult to determine the parameters in the approaches. On the other hand, in other approaches, the spectrogram of the signal is represented as an image, and a method of image processing such as median filters is applied to it in order to reduce musical noise.[27,28] In these approaches, however, it is necessary to estimate the position of musical noise. Their calculation cost is also high.

To solve the problem, we refer to the M-transform.[29] M-transform is a linear transformation technique using M-sequence. By combining M-transform in time domain, namely TD M-transform and TD ε-filter, we can reduce not only white noise but also impulse noise.[30] Musical noise is isolated in the time-frequency domain and is similar to impulse noise. Therefore, we tried to reduce the musical noise by improving M-transform for the time-frequency domain, namely a time-frequency M-transform (TF M-transform) whose procedure is the same as TD M-transform but applied to different data. Using TF M-transform, we can reduce musical noise effectively without estimating the position of musical noise. It is simple and can reduce not only the musical noise but also white noise.

While applying TF M-transform to the conventional methods, we also propose noise reduction combining the TF ε-filter and TF M-transform. In this way, we can reduce various types of noise such as low-level stationary noise, high-level nonstationary noise, and musical noise. The method is also simple and an improved version of the use of the TF ε-filter. In Sec. II, we describe the algorithm of the conventional M-transform. In Sec. III, we describe the algorithm of the TF M-transform. We then propose a method for noise reduction combining the TF ε-filter and TF M-transform. In Sec. V, we show the results of our experiment, followed by conclusions in Sec. VI.

II. NOISE REDUCTION UTILIZING M TRANSFORM

First, we explain M-transform and the noise reduction method utilizing M-transform. Let us consider GF(2) as the Galois field for generation of M-sequence, which is required for M-transform. Let $f(x)$ be the nth degree primitive polynomial defined on Galois field GF(2).[29] Let us also consider the M-sequence signal $\{a_i\}$ whose cycle is $N(=2^n-1)$ generated from $f(x)$. By using a_i, the ith element of $\{a_i\}$ (=0 or 1), m_i is defined as

$$m_i = (-1)^{a_i}. \tag{1}$$

Under the above-presented definition, the autocorrelation ϕ_{mm} of m_i is represented as

$$\phi_{mm} = \frac{1}{N}\sum_{i=0}^{N-1} m_{i-k}m_i = \begin{cases} 1 & (k = 0,N,2N, \ldots) \\ -\dfrac{1}{N} & (\text{otherwise}), \end{cases} \tag{2}$$

where $N=2^n-1$ represents the cycle of the nth degree M-sequence. When we select sufficiently large N as shown in Eq. (2), $\{m_i\}$ and $\{m_j\}$ $(i \neq j)$ have the pseudo-orthogonality, which shows the similarity of M-transform and orthogonal function expansion. Let us consider the matrix $\mathbf{M}_i$ whose size is $N \times N$ described as

$$\mathbf{M}_i = \begin{bmatrix} m_i & m_{i-1} & \cdots & m_{i-N+1} \\ m_{i+1} & m_i & \cdots & m_{i-N+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{i+N-1} & m_{i+N-2} & \cdots & m_i \end{bmatrix}, \tag{3}$$

where i indicates the starting point of the M-sequence and can be selected arbitrary and fixed. We also consider $x(t)$ as the input signal at time t. We then sample $x(t)$ at sampling interval Δt and define $x(k\Delta t)$ as the sampled signal. To simplify the representation, we abbreviate $x(k\Delta t)$ to $x(k)$. k corresponds to the discrete time. Let us define the vector $\mathbf{x}_k$ whose length is N, a cycle of the M-sequence, beginning from kth signal as

$$\mathbf{x}_k = [x(k), x(k+1), \ldots, x(k+N-1)]^T, \tag{4}$$

where $\mathbf{x}^T$ represents the transposed vector of $\mathbf{x}$. The vector $\mathbf{a}_k$ is defined as

$$\mathbf{x}_k = \mathbf{M}_i\mathbf{a}_k, \tag{5}$$

$$\mathbf{a}_k = [\alpha_0, \alpha_1, \ldots, \alpha_{N-1}]^T = (\mathbf{M}_i^T\mathbf{M}_i)^{-1}\mathbf{M}_i^T\mathbf{x}_k. \tag{6}$$

$\mathbf{a}_k$ is the M-transform of $\mathbf{x}_k$. Next, we explain noise reduction using the M-transform. Let us define the impulse noise vector $\mathbf{d}$ as

$$\mathbf{d} = [0,0, \ldots, d_s, 0, \ldots, 0]^T, \tag{7}$$

where d_s and s represent the amplitude of the impulse noise and the location of the impulse noise, respectively. When we consider the M-transform of $\mathbf{d}$, the rth element α_r is described as

$$\alpha_r = \frac{1}{N+1}(m_{r+s} - 1)d_s. \tag{8}$$

The impulse noise can be transformed to the M-sequence signal whose amplitude is small by the M-transform as shown in Eq. (8). On the other hand, the cross correlation of m_i and $x(k)$ is described as

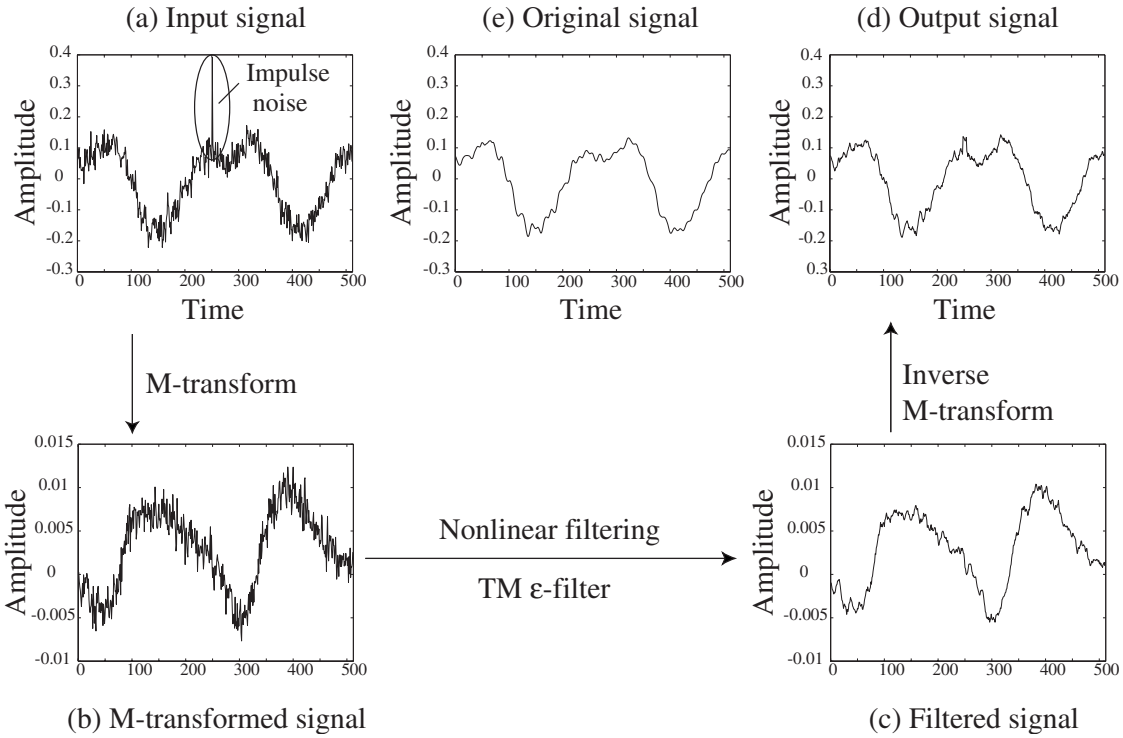


FIG. 1. Process of noise reduction using M-transform.

$$\phi_{mx}(u) = \frac{1}{N}\sum_{l=0}^{N-1} m_{l-u}x(l) = \frac{1}{N}\sum_{l=0}^{N-1} m_{l-u}\sum_{v=0}^{N-1} \alpha_v m_{l-v}$$
$$= \frac{1}{N}\sum_{v=0}^{N-1} \alpha_v \sum_{l=0}^{N-1} m_{l-u}m_{l-v} = \sum_{v=0}^{N-1} \alpha_v \phi_{mm}(v-u)$$
$$= \frac{N+1}{N}\alpha_u - \frac{1}{N}\sum_{v=0}^{N-1} \alpha_v. \tag{9}$$

When the average of $x(k)$ is zero, α_u determines $\phi_{mx}(u)$. Hence white noise is also transformed into small amplitude noise by the M-transform because white noise is noncorrelated with the M-sequence signal. Therefore, we can reduce the transformed impulse noise and white noise by employing a nonlinear filter which can reduce the small amplitude noise such as the ε-filter. Let us consider α_v as the input signal. The output signal β_v of ε-filter is described as

$$\beta_v = \alpha_v + \sum_{l=-R}^{R} a(l)F(\alpha_{v+l} - \alpha_l), \tag{10}$$

where $a(l)$ represents the filter coefficient. $a(l)$ is usually constrained as follows:

$$\sum_{l=-R}^{R} a(l) = 1. \tag{11}$$

The window size of the ε-filter is $2R+1$. $F(x)$ is the nonlinear function described as

$$|F(x)| \leqslant \varepsilon_0 : -\infty \leqslant x \leqslant \infty, \tag{12}$$

where ε_0 is a constant. The ε-filter can reduce small amplitude noise while preserving the speech signal. Figure 1 shows the process of the noise reduction utilizing M-transform. Noise reduction using M-transform is effective as shown in Fig. 1. In this paper, we label the ε-filter in the M domain "TM ε-filter."

III. MUSICAL NOISE REDUCTION UTILIZING TF M TRANSFORM

Figure 2 shows the spectrogram of the signal processed by the TF ε-filter. In Fig. 2, black pixels represent the points where the power is large, while white pixels represent the points where the power is small. As shown in Fig. 2, musical noise, which is similar to impulse noise, is isolated in the time-frequency domain.

Then, we apply the noise reduction combining the M-transform and ε-filter to the time-frequency domain. Let us consider $x(k)$ as the input signal as well as Sec. II. We transform $x(k)$ to $X(\kappa,\omega)$ by short-term Fourier transform (STFT) as follows:

$$X(\kappa,\omega) = \sum_{l=-\infty}^{\infty} x(\kappa+l)W(l)e^{-j\omega l}, \tag{13}$$

where κ and ω represent the time frame and the angular frequency, respectively. $W(l)$ represents the window function. j represents the imaginary unit. $\mathbf{X}_\omega(\kappa)$ represents the signal vector at time frame κ in angular frequency ω. $\mathbf{X}_\omega(\kappa)$ is described as

$$\mathbf{X}_\omega(\kappa) = [X(\kappa,\omega), X(\kappa+1,\omega), \ldots, X(\kappa+N-1,\omega)]^T. \tag{14}$$

Let us define the operator $\otimes$ as follows:

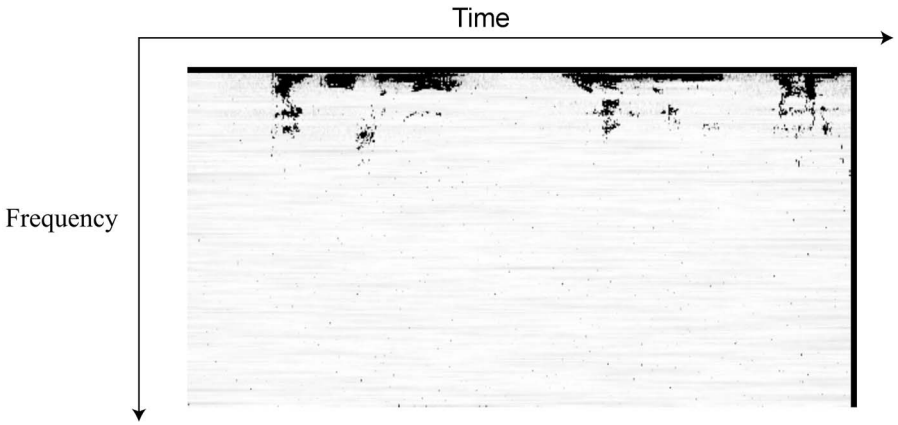


FIG. 2. Spectrogram of signal processed by the TF ε-filter.

$$\mathbf{G} \otimes \mathbf{H} = [G_1H_1, G_2H_2, \ldots, G_LH_L]^T, \tag{15}$$

where $\mathbf{G}$ and $\mathbf{H}$ represent the arbitrary vector whose size is $L \times 1$. $\mathbf{G}$ and $\mathbf{H}$ are described as

$$\mathbf{G} = [G_1, G_2, \ldots, G_L]^T, \tag{16}$$

$$\mathbf{H} = [H_1, H_2, \ldots, H_L]^T. \tag{17}$$

L is an arbitrary counting number. $\mathbf{X}_\omega(\kappa)$ can be described as

$$\mathbf{X}_\omega(\kappa) = \mathbf{P}_\omega(\kappa) \otimes e^{i\boldsymbol{\Theta}_\omega(\kappa)}, \tag{18}$$

where $\mathbf{P}_\omega(\kappa)$ and $\boldsymbol{\Theta}_\omega(\kappa)$ represent the power and the phase of $\mathbf{X}_\omega(\kappa)$, respectively. $\mathbf{P}_\omega(\kappa)$ and $\boldsymbol{\Theta}_\omega(\kappa)$ are described as

$$\mathbf{P}_\omega(\kappa) = [P_\omega(\kappa), P_\omega(\kappa+1), \ldots, P_\omega(\kappa+N-1)]^T, \tag{19}$$

$$\boldsymbol{\Theta}_\omega(\kappa) = [\Theta_\omega(\kappa), \Theta_\omega(\kappa+1), \ldots, \Theta_\omega(\kappa+N-1)]^T. \tag{20}$$

Next, we use the M-transform to $\mathbf{P}_\omega(\kappa)$ for time direction on every ω. The M-transform in the time-frequency domain, namely TF M-transform, is described as

$$\mathbf{A}_{\omega,\kappa} = (\mathbf{M}_i^T\mathbf{M}_i)^{-1}\mathbf{M}_i^T\mathbf{P}_\omega(\kappa)^T, \tag{21}$$

where $\mathbf{A}_{\omega,\kappa}$ represents the M-transformed signal vector of $\mathbf{P}_\omega(\kappa)$ described as follows:

$$\mathbf{A}_{\omega,\kappa} = [A_{\omega,\kappa}(0), A_{\omega,\kappa}(1), \ldots, A_{\omega,\kappa}(N-1)]^T. \tag{22}$$

$\mathbf{M}_i$ represents the matrix defined in Eq. (3). The musical noise is expected to be transformed to the small amplitude noise by this procedure. Therefore, we can reduce the M-transformed musical noise utilizing the ε-filter. We apply the ε-filter to $A_{\omega,\kappa}(q)$ as follows:

$$V_{\omega,\kappa}(q) = \sum_{l=-K}^{K} a(l)A'_{\omega,\kappa}(q+l), \tag{23}$$

where

$$A'_{\omega,\kappa}(q+l) = \begin{cases} A_{\omega,\kappa}(q) & (|A_{\omega,\kappa}(q+l) - A_{\omega,\kappa}(q)| > \varepsilon_M) \\ A_{\omega,\kappa}(q+l) & (|A_{\omega,\kappa}(q+l) - A_{\omega,\kappa}(q)| \leqslant \varepsilon_M), \end{cases} \tag{24}$$

and ε_M is constant. On the other hand, $a(l)$ is a filter coefficient. The window size of the ε-filter is $2K+1$. Let us define $\mathbf{V}_{\omega,\kappa}$ as

$$\mathbf{V}_{\omega,\kappa} = [V_{\omega,\kappa}(0), V_{\omega,\kappa}(1), \ldots, V_{\omega,\kappa}(N-1)]^T. \tag{25}$$

Then, the inverse M-transform is applied to $\mathbf{V}_{\omega,\kappa}$ as follows:

$$\mathbf{U}_\omega(\kappa) = \mathbf{M}_i\mathbf{V}_{\omega,\kappa}. \tag{26}$$

We then append the phase $\boldsymbol{\Theta}_\omega(\kappa)$ to $\mathbf{U}_\omega(\kappa)$,

$$\mathbf{Y}_\omega(\kappa) = \mathbf{U}_\omega(\kappa) \otimes e^{i\boldsymbol{\Theta}_\omega(\kappa)}. \tag{27}$$

Finally, we transform $Y(\kappa,\omega)$, the element of $\mathbf{Y}_\omega(\kappa)$, to $y(k)$ by inverse short-term Fourier transformation (ISTFT). $y(k)$ is obtained as

$$y(k) = \sum_{n=-\infty}^{\infty} Y(\kappa,\omega)e^{j\omega\kappa}. \tag{28}$$

Note that we repeatedly calculate Eq. (28) under the same window size and hop size as STFT. In this paper, we label the ε-filter in this method the "FM ε-filter."

IV. NOISE REDUCTION COMBINING TF ε FILTER AND TF M TRANSFORM

In this section, we propose noise reduction combining the TF ε-filter and TF M-transform. Figure 3 illustrates the proposed method with a block diagram. Let us consider $x(k)$, and $X(\kappa,\omega)$ transformed from $x(k)$ by STFT as in Sec. III,

$$X(\kappa,\omega) = \sum_{l=-\infty}^{\infty} x(\kappa+l)W(l)e^{-j\omega l}, \tag{29}$$

where $W(l)$ represents the window function. Next we use the TF ε-filter, that is an ε-filter used in the time-frequency domain, as shown by (2) in Fig. 3. In this procedure, $T(\kappa,\omega)$ is obtained as follows:

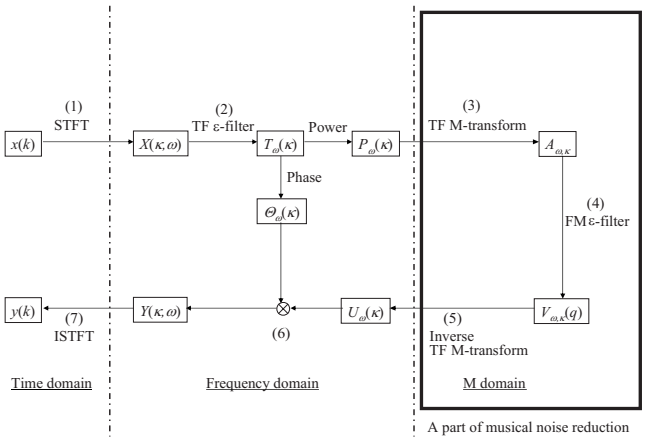


FIG. 3. Block diagram of the proposed method.

$$T(\kappa,\omega)=\sum_{l=-J}^{J} a(l)X'(\kappa+l,\omega), \tag{30}$$

where

$$X'(\kappa+l,\omega)=\begin{cases} X(\kappa,\omega) & (||X(\kappa,\omega)|-|X(\kappa+l,\omega)||>\varepsilon_F) \\ X(\kappa+l,\omega) & (||X(\kappa,\omega)|-|X(\kappa+l,\omega)||\leqslant\varepsilon_F), \end{cases} \tag{31}$$

and ε_F is a constant. $a(l)$ is a filter coefficient. The window size of the TF ε-filter is $2J+1$. It should be noted that $X(\kappa,\omega)$ is not a real number but a complex number. For a better understanding, we briefly explain the concept of the TF ε-filter. When we consider the distribution of speech signal and noise in the time-frequency domain, the following assumptions are usually satisfied.

Assumption 1. Speech signal has greater variation in power than noise signal in the time-frequency domain.

Assumption 2. Noise signal is distributed more uniformly and with less variation in the time-frequency domain.

Figure 4 illustrates the difference in performance when we apply the TF ε-filter to speech signal and noise. The horizontal axis and the vertical axis represent the real axis and imaginary axis, respectively. In Fig. 4, an asterisk and cross represent the remarkable point and the other signal points, respectively, in the same window as the remarkable point. Point A in Fig. 4(a) and point B in Fig. 4(b) represent the complex amplitude of the remarkable point. A' and B' represent the complex amplitude of the outputs when we apply the TF ε-filter to point A and B, respectively. Using the TF ε-filter, we first replace the complex amplitude of the signal outside of the shadow area by that of A. We then add the complex spectra of all the points in the same window. Due to the nature of complex spectra, if we have many signals that have similar amplitude but different phase, the real part and imaginary part cancel each other. Note that the noise is reduced not only when the noise amplitude is small but also when it is large because of this procedure. Figure 4(a) represents the basic concept in case of speech signals where

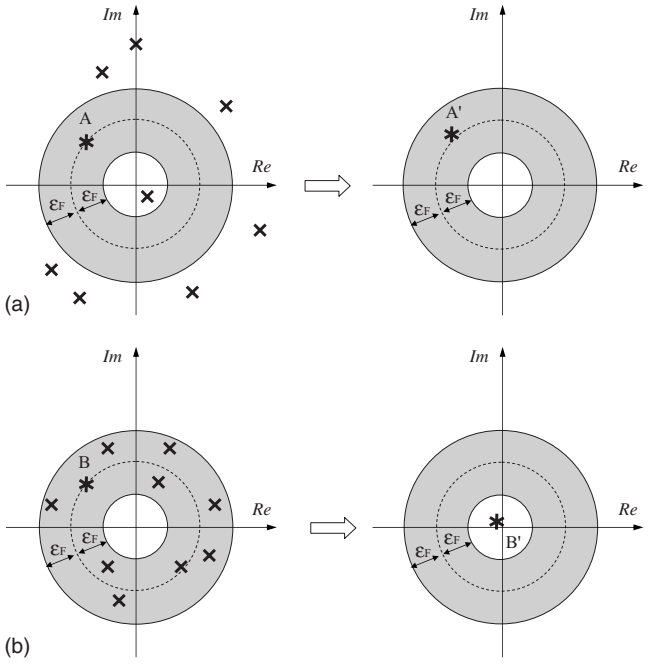


FIG. 4. Difference in performance when the TF ε-filter is applied to speech signal and noise: (a) Speech signal and (b) noise signal.

the power can vary frequently. When we consider a signal whose power varies frequently, the difference between the absolute value of A and that of the other signals is large as shown in Fig. 4(a). For this reason, few or none of the speech signal points in the same window as A fall within the shaded area containing A, and thus A is largely unchanged by the ε-filter. As a result, when we handle speech signal, the complex amplitude of the remarkable point is intact. Figure 4(b) shows the basic concept of noise signals where the power is more uniform and varies less frequently than speech signals. When we consider the noise signal, the difference between the absolute value of B and that of the other signals is relatively small compared with speech signal. Hence, there will be several points in the shaded area containing B, and the summation of these points with B by the filter results in a reduced value of B. In other words, when we handle noise, the complex amplitude of the remarkable point becomes small by using the TF ε-filter. Therefore, we can reduce noise, while preserving the signal by setting ε_F appropriately. Following the above-presented process, we can reduce the relatively large amplitude noise by the TF ε-filter. Although the performance of the proposed method may become less effective in discriminating between unvoiced speech and noise than in discriminating between voiced speech and noise, the difference is aurally inconspicuous.

However, it generates some musical noise because it is the process in the time-frequency domain. Therefore, we apply the TF M-transform to the output of the TF ε-filter. Let us consider $\mathbf{T}_\omega(\kappa)$ described as

$$\mathbf{T}_\omega(\kappa)=[T(\kappa,\omega),T(\kappa+1,\omega),\ldots,T(\kappa+N-1,\omega)]^T. \tag{32}$$

$\mathbf{T}_\omega(\kappa)$ can be rewritten as

$$\mathbf{T}_{\omega}(\kappa) = \mathbf{P}_{\omega}(\kappa) \otimes e^{i\boldsymbol{\Theta}_{\omega}(\kappa)}, \tag{33}$$

where $\mathbf{P}_{\omega}(\kappa)$ and $\boldsymbol{\Theta}_{\omega}(\kappa)$ represent the power and the phase of $\mathbf{T}_{\omega}(\kappa)$, respectively. $\mathbf{P}_{\omega}(\kappa)$ and $\boldsymbol{\Theta}_{\omega}(\kappa)$ are described as follows:

$$\mathbf{P}_{\omega}(\kappa) = [P_{\omega}(\kappa), P_{\omega}(\kappa+1), \ldots, P_{\omega}(\kappa+N-1)]^T, \tag{34}$$

$$\boldsymbol{\Theta}_{\omega}(\kappa) = [\Theta_{\omega}(\kappa), \Theta_{\omega}(\kappa+1), \ldots, \Theta_{\omega}(\kappa+N-1)]^T. \tag{35}$$

Next, we use the M-transform to $\mathbf{P}_{\omega}(\kappa)$ for time direction on every ω as shown by (3) in Fig. 3. The M-transform in the time-frequency domain is described as

$$\mathbf{A}_{\omega,\kappa} = (\mathbf{M}_i^T \mathbf{M}_i)^{-1} \mathbf{M}_i^T \mathbf{P}_{\omega}(\kappa), \tag{36}$$

where $\mathbf{A}_{\omega,\kappa}$ represents the M-transformed signal of $\mathbf{P}_{\omega}(\kappa)$ and is described as

$$\mathbf{A}_{\omega,\kappa} = [A_{\omega,\kappa}(0), A_{\omega,\kappa}(1), \ldots, A_{\omega,\kappa}(N-1)]^T. \tag{37}$$

$\mathbf{M}_i$ represents the matrix defined in Eq. (3). The musical noise is transformed to the small amplitude noise by this procedure. Then, we reduce the small amplitude noise in the M-transformed coordinate utilizing the FM ε-filter. Then, as shown by (4) in Fig. 3, we apply the FM ε-filter to $A_{\omega,\kappa}$ as follows:

$$V_{\omega,\kappa}(q) = \sum_{l=-K}^{K} a(l) A'_{\kappa,\omega}(q+l), \tag{38}$$

where

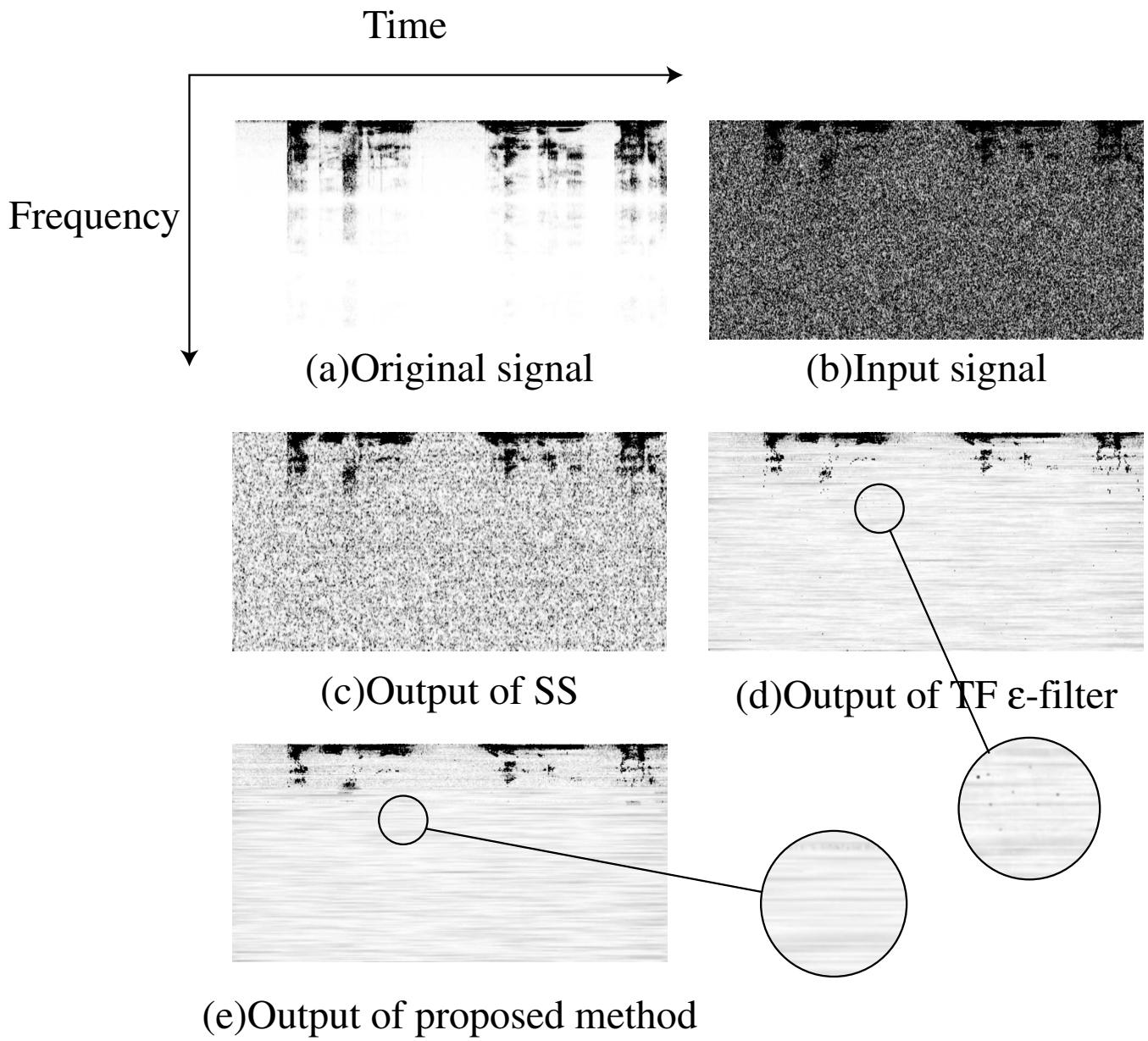


FIG. 5. Experimental results when a signal with stationary noise is used.

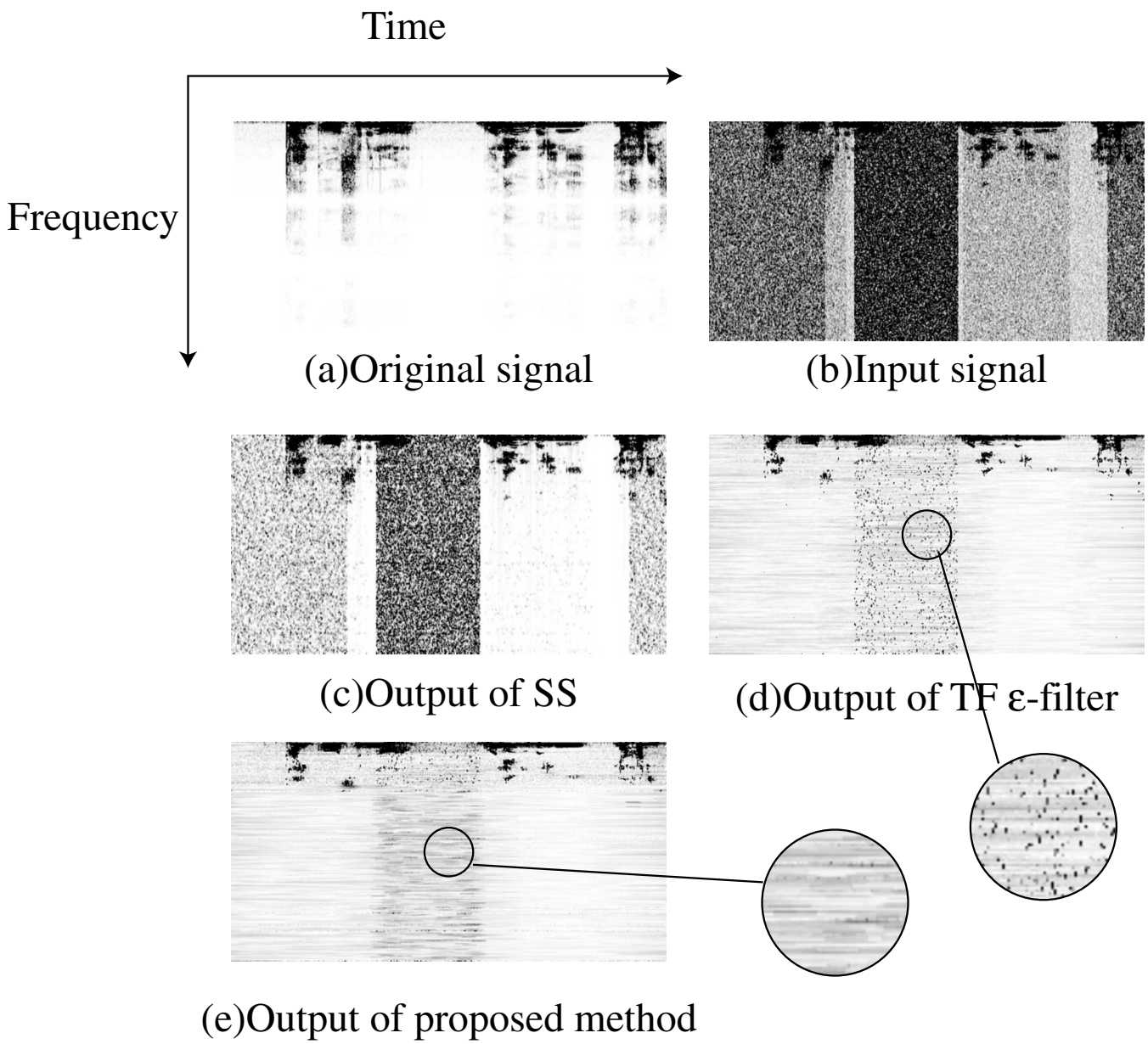


FIG. 6. Experimental results when a signal with nonstationary noise is used.

$$A'_{\omega,\kappa}(q+l) = \begin{cases} A_{\omega,\kappa}(q) & (|A_{\omega,\kappa}(q+l) - A_{\omega,\kappa}(q)| > \varepsilon_M) \\ A_{\omega,\kappa}(q+l) & (|A_{\omega,\kappa}(q+l) - A_{\omega,\kappa}(q)| \leqslant \varepsilon_M), \end{cases} \tag{39}$$

and $a(l)$ is a constant. Let us define $\mathbf{V}_{\omega,\kappa}$ as

$$\mathbf{V}_{\omega,\kappa} = [V_{\omega,\kappa}(0), V_{\omega,\kappa}(1), \ldots, V_{\omega,\kappa}(N-1)]^T. \tag{40}$$

Next, as shown by (5) in Fig. 3, we use inverse M-transform to $\mathbf{V}_{\omega,\kappa}$ as follows:

$$\mathbf{U}_\omega(\kappa) = \mathbf{M}_i \mathbf{V}_{\omega,\kappa}. \tag{41}$$

Then, as shown by (6) in Fig. 3, we append the phase $\mathbf{\Theta}_\omega(\kappa)$ to $\mathbf{U}_\omega(\kappa)$ as follows:

$$\mathbf{Y}_\omega(\kappa) = \mathbf{U}_\omega(\kappa) \otimes e^{i\mathbf{\Theta}_{\omega(\kappa)}}. \tag{42}$$

Finally, we transform $\mathbf{Y}(\omega,\kappa)$ to $y(k)$ by ISTFT as shown by (7) in Fig. 3.

It is considered that musical noise reduction based on TF M-transform requires fewer calculations than conventional methods because it requires only switching and linear operation.

V. EXPERIMENT

A. Experimental conditions

We conducted experiments using speech signal with a noise signal. For the sound source, we used "Japanese Newspaper Article Sentences" edited by the Acoustical Society of Japan. The signal and noise were mixed in the computer. Table I shows the values of common parameters used for all the experiments. Specifically, the experiments were con-

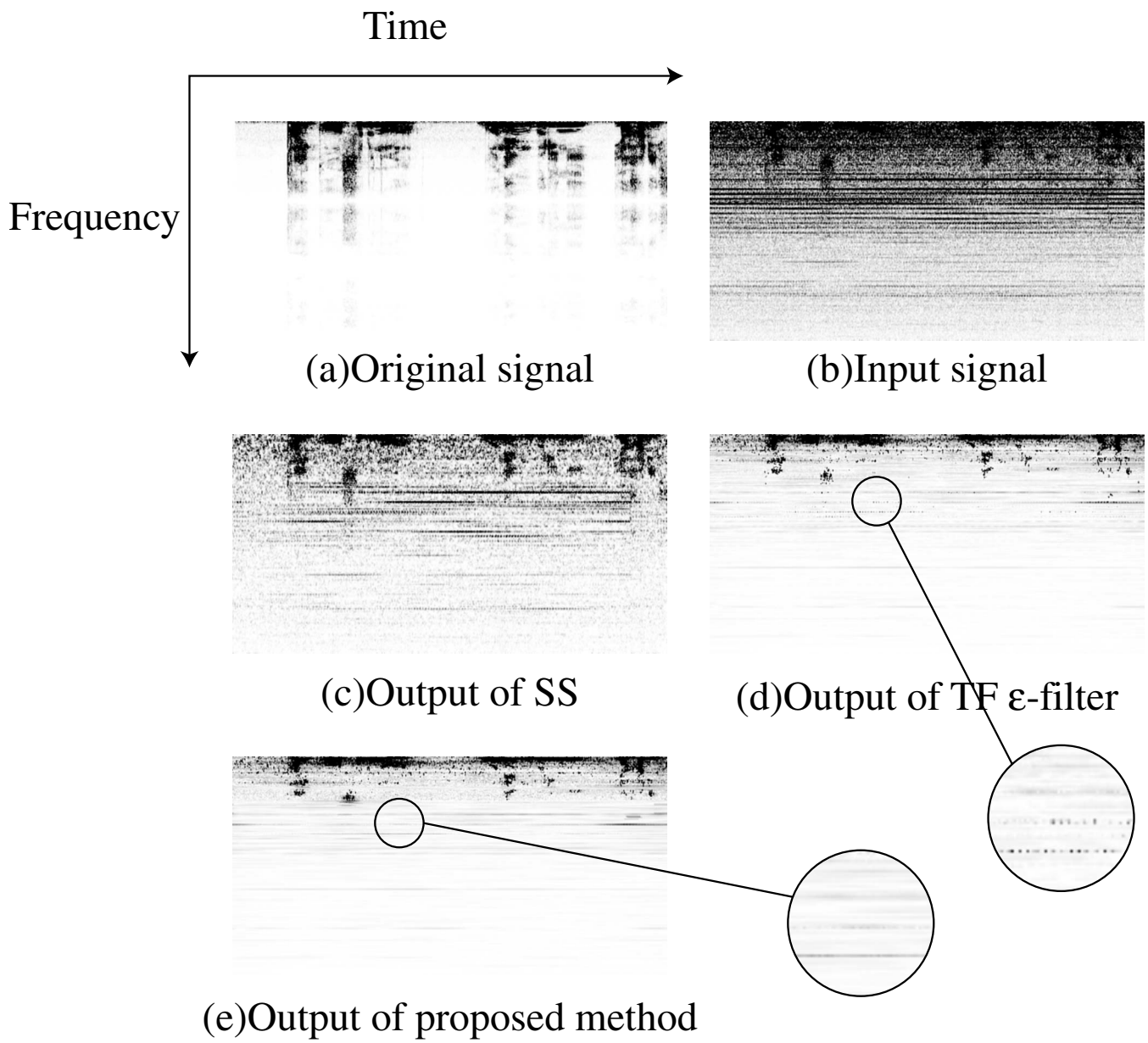


FIG. 7. Experimental results when a signal with natural noise is used.

ducted as follows: We first transformed the signal in the complex spectrum by using STFT. Block size and hop size were respectively set at 512 and 256, as shown in Table I. A Hanning window is utilized as the window function. The window was shifted point by point. We used a computer with an Intel Pentium M processor 1.73 GHz CPU. All programs were implemented by MATLAB. To evaluate the performance of the proposed methods quantitatively, we utilized the noise

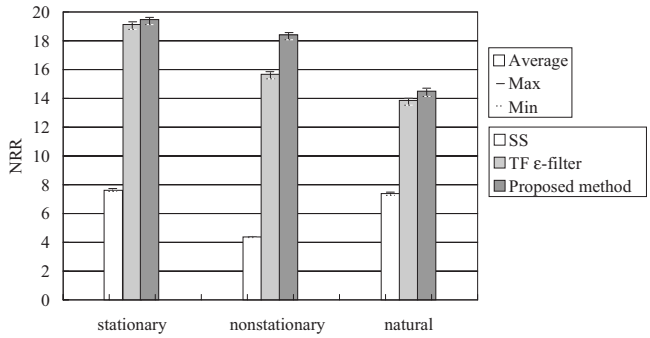


FIG. 8. NRR when various types of noise are used.

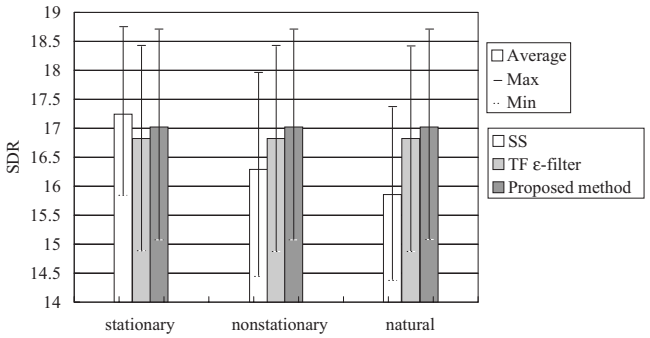


FIG. 9. SDR when various types of noise are used.

TABLE I. Common parameters.

Parameter	Value
Sampling frequency	44 100
STFT block size	512
Hop size	256
Window function	Hanning window

reduction ratio (NRR) and signal-to-distortion ratio (SDR). To calculate NRR, the signal-to-noise ratio (SNR) is defined as

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{\Sigma_{k=1}^{L} s(k)^2}{\Sigma_{k=1}^{L} n(k)^2}\right), \tag{43}$$

where $s(k)$, $n(k)$, and L represent the speech signal at time k, the noise signal at time k, and the length of signal, respectively.

NRR is defined as

$$\mathrm{NRR} = \mathrm{SNR}_{\mathrm{out}}\text{-}\mathrm{SNR}_{\mathrm{in}}, \tag{44}$$

where $\mathrm{SNR}_{\mathrm{out}}$ and $\mathrm{SNR}_{\mathrm{in}}$ represent the SNR after the process and before the process, respectively. To calculate $\mathrm{SNR}_{\mathrm{out}}$, we separately applied each method to the signal and noise, and calculated $\mathrm{SNR}_{\mathrm{out}}$ by using the obtained signal and noise. SDR can be represented as

$$\mathrm{SDR} = 10 \log_{10}\left(\frac{\Sigma_{k=1}^{L} s_{\mathrm{in}}(k)^2}{\Sigma_{k=1}^{L} (s_{\mathrm{in}}(k) - s_{\mathrm{out}}(k))^2}\right), \tag{45}$$

where $s_{\mathrm{in}}(k)$ and $s_{\mathrm{out}}(k)$ represent, respectively, the input signal and the output signal at time k when we used only the desired signal.

NRR represents how much the method reduces the noise. SDR represents how much the signal is distorted by reducing the noise.

B. Experimental results in the case of stationary noise

We first conducted the experiment using a signal with stationary noise. We prepared a speech signal as the signal and white noise as the stationary noise, respectively. We set ε_F of Eq. (31) and ε_M of Eq. (39) at 0.8 and 0.01, respectively. In addition, the window size of the TF ε-filter and that of the FM ε-filter are set at 101 and 17, respectively. We set the same parameters to the TF ε-filter as those of the proposed method. The parameters of the SS are set optimally.

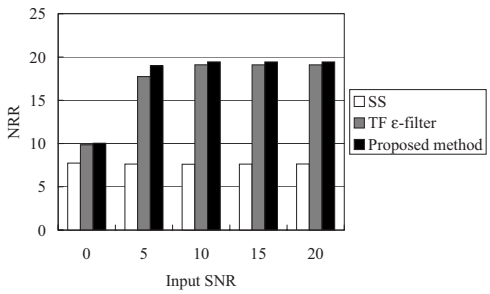


FIG. 10. Experimental results on NRR.

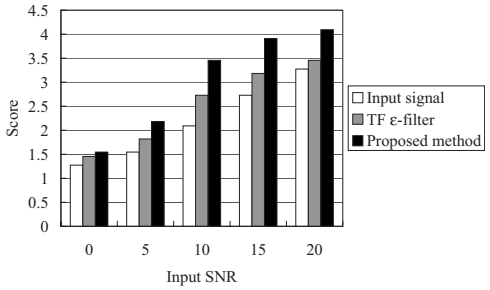


FIG. 11. Experimental results on subjective evaluation.

Figure 5(a) shows the spectrogram of the original signal. Figure 5(b) shows the spectrogram of the signal with stationary noise. Figures 5(c)–5(e) depict the spectrogram of the output of SS, the TF ε-filter, and the proposed method, respectively. As shown in Fig. 5, the proposed method can reduce the musical noise effectively, while it remains when SS and the TF ε-filter are applied.

C. Experimental results in the case of nonstationary noise

The experiment was conducted using a signal with nonstationary noise, which is the noise whose variance changes. We used the same speech signal as in Sec. V B. We prepared white noise with a variance that sometimes varied. We set ε_F of Eq. (31) and ε_M of Eq. (39) at 0.8 and 0.01, respectively. In addition, we set the window size of the TF ε-filter and that of the FM ε-filter at 101 and 17, respectively. We set the same parameters to the TF ε-filter as those used in the proposed method. The parameters of the SS are set optimally. Figure 6(a) shows the spectrogram of the original signal. Figure 6(b) shows the spectrogram of the signal with nonstationary noise. Figures 6(c)–6(e) depict the spectrogram of the output of the SS, the TF ε-filter, and the proposed method, respectively. As shown in Fig. 6, when we use the proposed method, the noise can be reduced in spite of nonstationary noise.

D. Experimental results in the case of natural noise

The experiment was conducted using a signal with natural noise. We used the same speech signal as in Sec. V B and the sound generated from the cooling fan of a personal computer as a noise source. We set ε_F of Eq. (31) and ε_M of Eq. (39) at 0.8 and 0.01, respectively. In addition, we set the window size of the TF ε-filter and that of the FM ε-filter at 101 and 17, respectively. We set the same parameters to the TF ε-filter as those of the proposed method. The parameters of the SS are set optimally. Figure 7(a) shows the spectrogram of the original signal. Figure 7(b) shows the spectrogram of the signal with natural noise. Figures 7(c)–7(e) show the spectrogram of the output of the SS, TF ε-filter, and proposed method, respectively. As shown in Fig. 7, when we use the proposed method, the noise can be reduced in spite of natural noise.

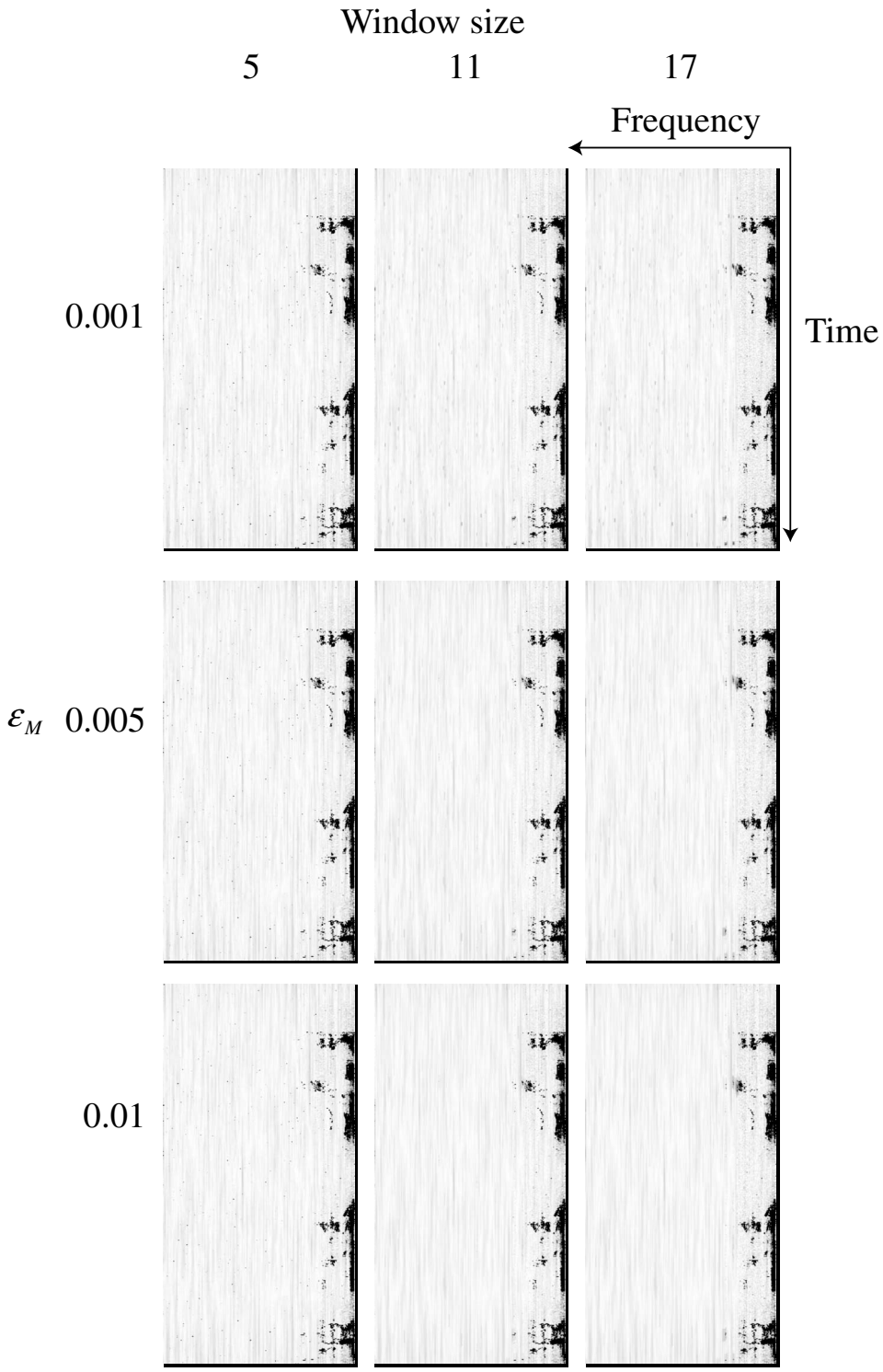


FIG. 12. Performance change depending on the window size and ε.

E. Quantitative evaluation for utilizing various noises

To evaluate the performance of the proposed methods, we calculated NRR and SDR about the stationary, nonstationary, and natural noise. We employed the voices of three males and two females as the speech signal. Figures 8 and 9 show the maximal, the minimal, and the average values of NRR and SDR, respectively. As shown in Fig. 8, NRR in the proposed method is superior to that in the other methods in all the experiments. As shown in Fig. 9, SDR in the proposed method is improved compared to that in the TF ε-filter in all cases. SDR in the proposed method is also superior to that in the other methods in all the cases of nonstationary noise and natural noise.

F. Robustness for various SNR

To evaluate the robustness for various SNR, we conducted an experiment using signals with various types of noise to confirm that the proposed methods can be employed

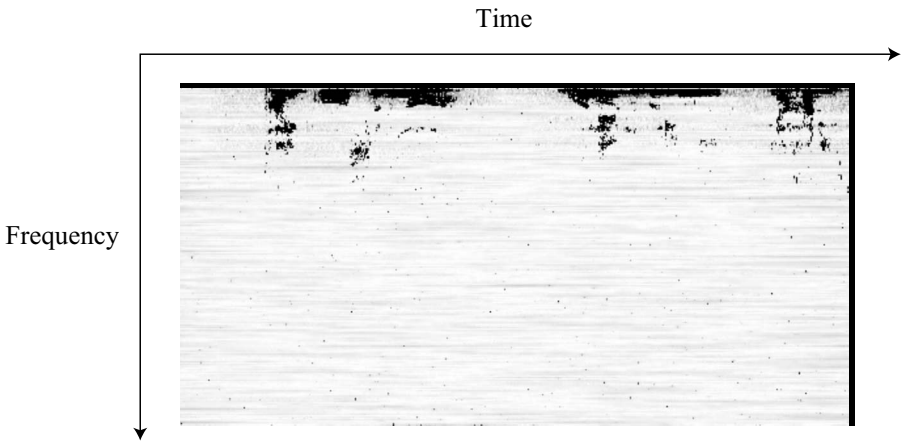


FIG. 13. Experimental results when the window size is set 5 and the ε is set 0.001.

not only for reducing small noise but also for loud noise. Robustness in this experiment means that the proposed method can reduce the noise regardless of input SNR. The window size and ε_M of the FM ε-filter are set at 17 and 0.01, respectively. We used five signals whose SNR is 0, 5, 10, 15, and 20 dB, respectively. Figure 10 shows the results on NRR. As shown in Fig. 10, the NRR of the proposed method is better than that of the TF ε-filter or SS.

G. Results of the experiment on subjective evaluation

We conducted an experiment on subjective evaluation. We used five signals, that are identical to those in Sec. V F as input signals. The examinees listened to the three signals: input signal, signal processed by the TF ε-filter, and signal processed by the proposed method for every input SNR. The examinees rated each signal on a scale of 1 to 5. Note that score 1 is the worst rating while score 5 is the best as auditory impression. The 11 examinees participated in this experiment.

Figure 11 shows the results of the experiment on subjective evaluation. As shown in Fig. 11, the signal processed by the proposed method shows better results than any other signals.

H. Performance change depending on the window size and ε

We also conducted an experiment to confirm the performance depending on the window size and ε_M of the FM ε-filter. To clarify differences depending on ε value and window size, we conducted an experiment using window sizes of 5, 11, and 17 as well as ε values of 0.001, 0.005, and 0.01. We set ε_F at 0.6. We employed the same signal and noise as in Sec. V B. Figure 2 shows the spectrogram of the signal with musical noise used in the experiment.

Figure 12 shows the results. To clarify differences in performance, we show the enlarged figure as shown in Fig. 13 when the window size and the ε are set at 5 and 0.001, respectively. We also show the enlarged figure as shown in Fig. 14 when the window size and the ε are set at 17 and 0.01, respectively. As shown in Figs. 12–14 the smaller musical noise remains as the window size and ε become larger.

VI. CONCLUSION

In this paper, we introduced an advanced M-transform for the time-frequency domain, namely TF M-transform. The proposed method is simple and can reduce musical noise effectively.

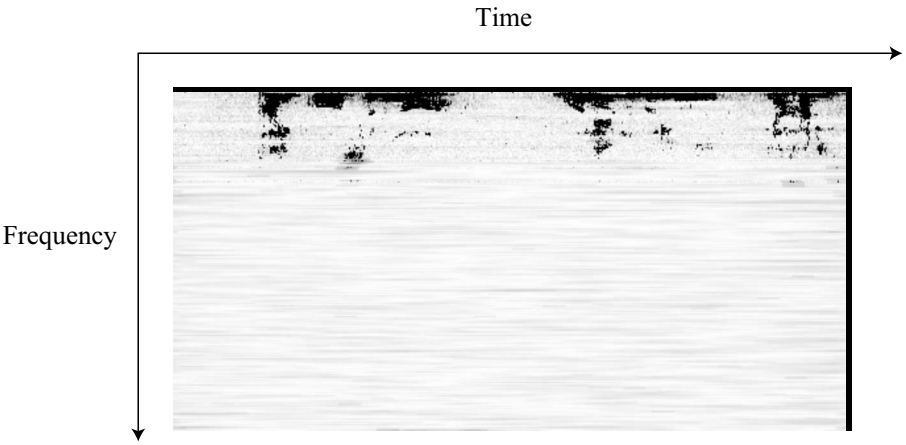


FIG. 14. Experimental results when the window size is set 17 and the ε is set 0.01.

We also proposed noise reduction combining the TF ε-filter and TF M-transform. The proposed method can reduce not only the low-level stationary noise but also high-level nonstationary noise. It also can reduce musical noise. The experimental results show that it can reduce the noise more effectively than methods using the conventional TF ε-filter and SS regardless of the input SNR. The method can be applied not only to the TF ε-filter but also to any methods in the time-frequency domain such as SS. For further developments, we would like to employ the proposed method as the preprocessing or postprocessing of some other noise reduction system. We also would like to apply the proposed method to the real robot audition and speech system.

ACKNOWLEDGMENTS

This research was supported by a research grant of the Support Center for Advanced Telecommunications Technology Research (SCAT), a research grant of the Tateisi Science and Technology Foundation, and by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 20700168, 2008. This research was also supported by Waseda University Grant for Special Research Projects(B), Nos. 2007B-142, 2007B-143 and 2007B-168, by "Establishment of Consolidated Research Institute for Advanced Science and Medical Care," Encouraging Development Strategic Research Centers Program, the Special Coordination Funds for Promoting Science and Technology, Ministry of Education, Culture, Sports, Science and Technology, Japan, by the CREST project "Foundation of technology supporting the creation of digital media contents" of JST, by the Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture, the 21st Century Center of Excellence Program, "The innovative research on symbiosis technologies for human and robots in the elderly dominated society," Waseda University, and by Project for Strategic Development of Advanced Robotics Elemental Technologies (NEDO: 06002090).

[1]K. Sasaki and K. Hirata, "3D-localization of a stationary random acoustic source in near-field by using 3 point-detectors," The Society of Instrument and Control Engineers **34**, 1329–1337 (1998).
[2]Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by the closely located four point microphone method," J. Acoust. Soc. Jpn. **10**, 101–110 (1990).
[3]M. Matsumoto and S. Hashimoto, "A miniaturized adaptive microphone array under directional constraint utilizing aggregated microphones," J. Acoust. Soc. Am. **119**, 352–359 (2006).
[4]K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, Germany, 1997, pp. 215–218.
[5]Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**, 1391–1400 (1986).
[6]K. Takao, M. Fujita, and T. Nishi, "An adaptive antenna array under directional constraint," IEEE Trans. Antennas Propag. **24**, 662–669 (1976).
[7]A. J. Bell and T. J. Sejowski, "An information maximization approach to blind separation and blind deconvolution," Neural Comput. **7**, 1129–1159 (1995).
[8]H. Saruwatari, S. Kurita, and K. Takeda, "Blind source separation combining frequency-domain ICA and beamforming," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, 2001, pp. 146–157.
[9]T. Ihara, M. Handa, T. Nagai, and A. Kurematsu, "Multi-channel speech separation and localization by frequency assignment," IEICE Trans. Fundamentals **J86-A**, 998–1009 (2003).
[10]S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, Florida, 2002, pp. 529–532.
[11]M. Aoki, Y. Yamaguchi, K. Furuya, and A. Kataoka, "Modifying SAFIA: Separation of the target source close to the microphones and noise sources far from the microphones," IEICE Trans. Fundamentals **J88-A**, 468–479 (2005).
[12]P. Daniel, W. Ellis, and R. Weiss, "Model-based monaural source separation using a vector-quantized phase-vocoder representation," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, 2006, pp. V-957–960.
[13]J. S. Lim, A. V. Oppenheim, and L. D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 419–423 (1978).
[14]R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME **82**, 35–45 (1960).
[15]M. Fujimoto and Y. Ariki, "Speech recognition under noisy environments using speech signal estimation method based on Kalman filter," IEICE Trans. Inf. Syst. **J85-D-II**, 1–11 (2002).
[16]S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 2nd ed., Wiley, New York, 2000.
[17]J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 197–210 (1978).
[18]M. Meguro, A. Taguchi, and N. Hamada, "Data-dependent weighted median filtering with robust motion information for restoring image sequence degraded by additive Gaussian and impulsive noise," IEICE Trans. Fundamentals **E84-A**, 432–440 (2001).
[19]X. Wang and D. Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images," IEEE Trans. Circuits Syst., II: Analog Digital Signal Process. **46**, 78–80 (2002).
[20]T. Abe, M. Matsumoto, and S. Hashimoto, "Noise reduction combining time-domain ε-filter and time-frequency ε-filter," J. Acoust. Soc. Am. **122**, 2697–2705 (2007).
[21]H. Harashima, K. Odajima, Y. Shishikui, and H. Miyakawa, "ε-separating nonlinear digital filter and its applications," IEICE Trans. Fundamentals **J65-A**, 297–303 (1982).
[22]K. Arakawa, K. Matsuura, H. Watabe, and Y. Arakawa, "A method of noise reduction for speech signals using component separating ε-filters," IEICE Trans. Fundamentals **J85-A**, 1059–1069 (2002).
[23]S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**, 113–120 (1979).
[24]K. Yamashita, S. Ogata, and T. Shimamura, "Improved spectral subtraction utilizing iterative processing," IEICE Trans. Fundamentals **J88-A**, 1246–1257 (2005).
[25]J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 354–358 (1978).
[26]N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech Audio Process. **7**, 126–137 (1999).
[27]Z. Goh, K. C. Tan, and T. G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," IEEE Trans. Speech Audio Process. **6**, 287–292 (1998).
[28]Y. Nomura, H. Tozawa, J. Lu, H. Sekiya, and T. Yahagi, "Musical noise reduction by spectral subtraction using morphological process," IEICE Trans. Inf. Syst. **89**, 991–1000 (2006).
[29]H. Kashiwagi, M. Liu, H. Harada, and T. Yamaguchi, "M-transform and its application to system identification," The Society of Instrument and Control Engineers **34**, 1785–1790 (1998).
[30]H. Harada, H. Kashiwagi, T. Andoh, and K. Kaba, "Impulsive noise reduction by use of M-transform," The Society of Instrument and Control Engineers **39**, 688–690 (2003).

Time reversal of flexural waves in a beam at audible frequency[a)]

Dany Francoeur and Alain Berry[b)]
Groupe d'Acoustique de l'Université de Sherbrooke, Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, Québec J1K 2R1, Canada

(Received 10 March 2007; revised 20 May 2008; accepted 21 May 2008)

There has been very limited work on the application of time Reversal to the propagation of audible frequency waves in mechanical structures. The present work concentrates on the application of time reversal to the focusing of audible range, flexural waves in an infinite beam, and to the detection of local heterogeneity in such a beam. Practical applications of time reversal of flexural waves in structures include vibration energy focusing, detection of vibratory or acoustic sources, and detection of defects in mechanical structures. An analytical model of flexural wave propagation in the beam as well as sensing and emission using piezoelectric transducers is presented. Time reversal experiments are conducted and compared to the model results in either a homogeneous beam or a beam with point mass heterogeneities. In the various situations tested, it is shown that time reversal effectively compensates the spreading in time of the impulse due to the dispersive propagation of flexural waves. One interesting aspect of this property is the generation of large amplitude impulsive responses in the beam using remote actuators. Finally, the "Décomposition de l'Opérateur de Retournement Temporel" approach is examined to detect and localize point mass scatterers in the beam. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945160]

PACS number(s): 43.60.Tj, 43.40.Cw, 43.60.Jn [DF] Pages: 1006–1017

I. INTRODUCTION

Most previous investigations of time reversal acoustics have been conducted with ultrasound because of the small spatial resolution permitted by small wavelength acoustic waves (Prada *et al.*, 2002). The application of time reversal acoustics to the audible frequency range has been the subject of several recent papers, with applications to the reduction of reverberation (Candy *et al.*, 2004), creation of a virtual sound field (Yon *et al.*, 2003), or audible sound focusing using a time reversal acoustic sink (Bavu *et al.*, 2007). On the other hand, time reversal has been also investigated for wave propagation in solids (Draeger *et al.*, 1997; Draeger *et al.*, 1999; Puckett *et al.*, 2003). Most studies have been limited to nondispersive wave propagation (wave velocity independent of frequency). Notable exceptions are the works of Ing *et al.* (1998) and Montaldo *et al.* (2001), who have concentrated on time reversal of Lamb waves in the ultrasonic range in dispersive solids. Another paper is the work of Wang *et al.* (2004), who have experimented time reversal of both extensional, nondispersive waves and flexural, dispersive waves in thin plates. To date, however, there has been very limited work on the application of time reversal to the propagation of audible frequency waves in mechanical structures (Dickey, 2000). Especially, the case of flexural waves in mechanical structures is characterized by large wavelengths and dispersive propagation, which create specific conditions for time reversal.

The present work concentrates on the application of time reversal to the focusing of audible range, flexural waves in an infinite beam (the structural wavelength is large compared to the beam thickness and width) and to the detection of local heterogeneity in such a beam. Practical applications of time reversal of flexural waves in structures include vibration energy focusing, detection of vibratory or acoustic sources, and detection of defects in mechanical structures. In time reversal acoustics, energy focusing results from constructive interference of time-compact waves processed and emitted by a time reversal mirror (TRM) (Fink, 1993). In dispersive propagation conditions, these waves spread or contract in time as they propagate in the medium. It is shown here that wave dispersion is compensated in the time reversal process and that dispersive propagation can be favorably exploited to enhance energy focusing in a structure. Detection, localization, and energy focusing on scatterers in the propagation medium are well-known applications of time reversal. This has been investigated through various approaches: iterative methods (Fink *et al.*, 2002; Montaldo *et al.*, 2004) and time reversal operator decomposition (DORT) (Prada *et al.*, 1996). In previous work, the DORT method was found to rely on several assumptions: the various scatterers in the medium should be "well resolved," and the TRM should be sufficiently compact with respect to the separation between the TRM transducers and the scatterers, in order to eliminate the direct propagation paths between emitters and receivers of the TRM.

In this study, we investigate time reversal in a long flexural beam with point mass scatterers. Although time reversal is a general concept applicable to any wave propagating medium for which the wave equation is invariant after time reversal, it is applied here to a flexural beam for sake of simplicity. The TRM used is an array of piezoelectric actuators and sensors uniformly distributed along the beam. In such a situation of a one-dimensional structure with a distrib-

[a)]Portions of this work were presented in "Time reversal in heterogeneous flexural beams," 149th ASA meeting, Canada, Vancouver, May 2005.
[b)]Electronic mail: alain.berry@usherbrooke.ca

 0001-4966/2008/124(2)/1006/12/$23.00

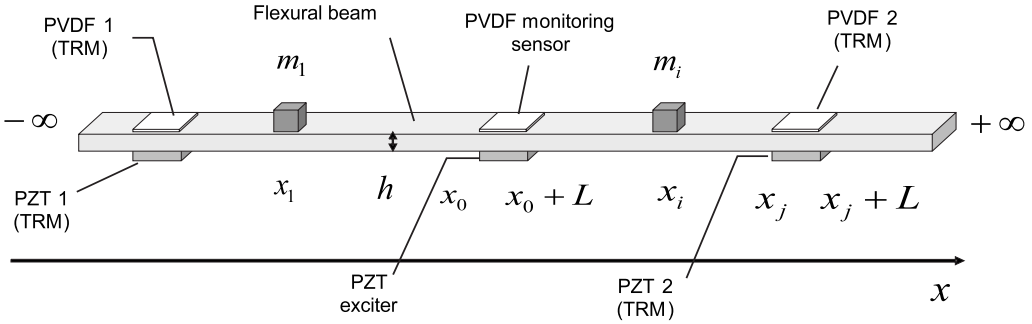


FIG. 1. Schematics of the flexural beam with point mass heterogeneities and piezoelectric TRM.

uted TRM, the DORT assumptions may not hold. One aspect of this work is therefore to investigate the application of DORT to detect and localize point mass scatterers in a flexural beam for audible frequency propagation. Section II of this paper discusses an analytical model of flexural wave propagation in the beam as well as sensing and emission using piezoelectric transducers. Section III presents the time reversal equations in the beam. Section IV details the results of time reversal experiments in either a homogeneous beam or a beam with point mass heterogeneities. In the various situations tested, it is shown that time reversal effectively compensates the spreading in time of the impulse due to the dispersive propagation of flexural waves: the backpropagation phase contributes to recontract the impulse at the point of initial emission. One interesting aspect of this property is the generation of large amplitude impulsive responses in the beam using remote actuators; this is discussed in Sec. IV C. Finally, Sec. V discusses the application of DORT to detect and localize point mass scatterers in the beam.

II. MODEL

Consider an infinite flexural beam of thickness h, mass per unit length μ, and bending stiffness EI (E, Young's modulus; I, sectional mass moment of inertia). The response to a unitary impulse transverse force at $x=x_s$ in terms of the transverse displacement $w(x,t)$ of the beam is solution of the Euler–Bernoulli equation (Graff, 1991):

$$-\mu\frac{\partial^2 w}{\partial t^2}+EI\frac{\partial^4 w}{\partial x^4}=\delta(x-x_s)\delta(t), \tag{1}$$

where δ is the delta-Dirac distribution. Applying the Fourier transform $\tilde{w}(x,\omega)=\int_{-\infty}^{+\infty}w(x,t)e^{-j\omega t}dt$ to the above equation results in

$$-\mu\omega^2\tilde{w}+EI\frac{\partial^4\tilde{w}}{\partial x^4}=\delta(x-x_s). \tag{2}$$

The solution takes the form of an evanescent wave and an outgoing flexural wave on each side of the excitation point,

$$\tilde{w}(x,\omega)=\frac{1}{4EIk^3}e^{-k|x-x_s|}+\frac{j}{4EIk^3}e^{-jk|x-x_s|}, \tag{3}$$

where the flexural wave number in the beam is $k=(\mu/EI)^{1/4}\omega^{1/2}$. Since the flexural wave number is not directly proportional to angular frequency, the wave propagation in the beam is dispersive, resulting in a spreading of the impulse response as the distance $|x-x_s|$ increases. The impulse response of the beam can be obtained with the inverse Fourier transform $w(x,t)=1/2\pi\int_{-\infty}^{+\infty}\tilde{w}(x,\omega)e^{j\omega t}d\omega$. Given the expressions (3), the inverse Fourier transform cannot be obtained in closed form, and must therefore be calculated numerically.

The theoretical model investigated in the following sections is slightly more complicated than above and is illustrated in Fig. 1. The base structure is an infinite uniform flexural beam with heterogeneity elements (modeled as point masses m_i at positions x_i, $i=1,\ldots,M$). The TRM is formed of a series of reciprocal, collocated pairs of piezoceramic (PZT) actuators and piezopolymer (PVDF) sensors distributed along the beam; each pair extends from x_j to x_j+L, $j=1,\ldots,N$, where L designates the identical length of all N PZT sources and PVDF sensors (two PZT-PVDF pairs are shown for illustration purpose on Fig. 1). In addition, the initial primary excitation source is a single PZT actuator extending from x_0 to x_0+L, and a monitoring PVDF sensor is collocated with the primary PZT exciter to observe the system response at the point of initial emission during the time reversal operation. In the following, the PZT exciter and monitoring PVDF sensor are identified with the index 0 to distinguish with the TRM transducers $j=1,\ldots,N$. A single surface-mounted PZT actuator induces both bending moments and extensional forces in the beam when subjected to an electric field in the transverse direction. The active power or power flow induced in the beam by the action of a single surface-mounted PZT can therefore be partitioned into a flexural and an extensional power flow. For the beam and PZT emitters considered in this study (see Sec. IV), the ratio between extensional and flexural power flow can be obtained using the relations given in Gibbs (1992): this ratio is found to be smaller than 0.1 for frequencies smaller than 2.5 kHz. Therefore, extensional waves in the beam are disregarded in this study. Under the assumption of perfect bonding and small actuator thickness, the bending excitation of the PZT is a pair of opposite bending moments concentrated at the actuator ends x_j and x_j+L (Gibbs, 1992),

$$M_{\mathrm{PZT}}(\omega)=\frac{EIK_f d_{31}U^{\mathrm{PZT}}(\omega)}{h_{\mathrm{PZT}}}, \tag{4}$$

where h_{PZT} is the PZT actuator thickness, d_{31} is the piezoelectric strain coefficient, U_j^{PZT} is the actuation voltage, and K_f is a coefficient given by

$$K_f = \frac{6EE_{\mathrm{PZT}}hh_{\mathrm{PZT}}(h+h_{\mathrm{PZT}})}{E^2h^4 + EE_{\mathrm{PZT}}hh_{\mathrm{PZT}}(4h^2+6hh_{\mathrm{PZT}}+4h_{\mathrm{PZT}}^2) + E_{\mathrm{PZT}}^2h_{\mathrm{PZT}}^4}. \tag{5}$$

The forced response of a homogeneous beam under the action of a single PZT actuator j in the frequency domain is therefore the solution of

$$-\mu\omega^2\tilde{w} + EI\frac{\partial^4\tilde{w}}{\partial x^4} = M_{\mathrm{PZT}}(\omega)[\delta'(x-x_j) - \delta'(x-x_j-L)], \tag{6}$$

where δ' denotes the derivative of the Dirac function. In the case of the heterogeneous beam with point masses m_i under the action of a single PZT actuator, the vibration solution in the beam takes the form of a system of propagative and evanescent bending waves in each interval $x_l \leqslant x \leqslant x_{l+1}$ between consecutive elements on the beam (masses and ends of active piezoceramic sources),

$$\tilde{w}_l(x,\omega) = A_{+l}e^{-jk(x-x_l)} + A_{-l}e^{jk(x-x_{l+1})} + B_le^{-k(x-x_l)} + C_le^{k(x-x_{l+1})} \quad \text{for } x_l \leq x \leq x_{l+1}. \tag{7}$$

The wave amplitudes A_{+l}, A_{-l}, B_l, C_l are solved from appropriate continuity conditions at x_l and x_{l+1}. The solution procedure is detailed in the Appendix.

Once the beam response is determined, the electrical charge response of each PVDF sensor in the TRM under the action of a single PZT actuator j can be determined from the following equation (Lee and Moon, 1990):

$$q_k(\omega) = -\frac{h+h_{\mathrm{PVDF}}}{2}b_{\mathrm{PVDF}}e_{31}\int_{x_i}^{x_i+L}\frac{\partial^2\tilde{w}_k}{\partial x^2}dx, \tag{8}$$

where h_{PVDF} and b_{PVDF} are, respectively, the PVDF thickness and width, e_{31} is the piezoelectric charge coefficient of the PVDF material, and x_k, x_k+L are the limits of the PVDF sensor. Inserting Eq. (7) in Eq. (8),

$$q_k(\omega) = -\frac{h+h_{\mathrm{PVDF}}}{2}b_{\mathrm{PVDF}}e_{31}[jk(A_{+l}+A_{-l})(1-e^{-jkL}) + k(B_l+C_l)(1-e^{-kL})]. \tag{9}$$

The voltage response of the PVDF sensors is given by $U_k^{\mathrm{PVDF}} = q_k/C$, where C is the sensor capacitance. The input voltage of the PZT actuators U^{PZT} and the output voltage of the PVDF sensors U^{PVDF} are the signals that form the reciprocal transducers of the TRM. The frequency response functions between individual PZT actuators and PVDF sensors are defined as

$$H_{jk}(\omega) = \frac{U_k^{\mathrm{PVDF}}(\omega)}{U_j^{\mathrm{PZT}}(\omega)}. \tag{10}$$

This includes frequency response functions H_{0k} between the PZT exciter on the beam and the TRM sensors, and the mutual paths $H_{jk}(\omega)$, $j \geqslant 1$ between the TRM actuators and sensors. The impulse response of the PVDF sensor k, $H_{jk}(t)$, under the action of a single PZT actuator j driven by a unitary impulse voltage is obtained by the inverse Fourier transform of Eq. (10). Similarly, the impulse response of the heterogeneous beam $G_{jl}(x,t)$ under the action of a single PZT actuator j driven by a unitary impulse voltage is obtained by the inverse Fourier transform of $G_{jl}(x,\omega) = \tilde{w}_l(x,\omega)/U^{\mathrm{PZT}}(\omega)$.

Note that the extension to flexural wave propagation in a plate (instead of a beam) follows similar lines, starting with the usual Love–Kirchoff plate equation instead of the Euler–Bernoulli beam equation (1).

III. TIME REVERSAL IN A FLEXURAL BEAM

Time reversal simulations were conducted with the model described above. The time reversal process is classical and consists of the following steps:(1) compute the impulse responses between the PZT exciter and each of the TRM PVDF sensors, $H_{0k}(t)$; (2) time invert the impulse responses, $H_{0k}(-t)$; (3) reemit the time inverted responses through all the corresponding PZT actuators of the TRM. The heterogeneous beam response after the reemission process is given by $\Sigma_k H_{0k}(-t) * G_{kl}(x,t)$, where $*$ denotes the time convolution. Also, the response of the monitoring PVDF sensor after reemission is given by $\Sigma_k H_{0k}(-t) * H_{k0}(t) = \Sigma_k H_{0k}(-t) * H_{0k}(t)$, since the $H_{jk}(t)$ satisfy the reciprocity property $H_{jk}(t) = H_{kj}(t)$.

If the PZT exciter is initially driven by the arbitrary input signal $s(t)$, the heterogeneous beam response after the reemission process is given by

$$w_l(x,t) = s(-t) * \sum_k H_{0k}(-t) * G_{kl}(x,t), \tag{11}$$

and the response of the monitoring PVDF sensor after reemission is given by

$$U_0^{\mathrm{PVDF}}(t) = s(-t) * \sum_k H_{0k}(-t) * H_{0k}(t). \tag{12}$$

The corresponding frequency domain relations are

$$\tilde{w}_l(x,\omega) = \tilde{s}^*(\omega)\sum_k H_{0k}^*(\omega)G_{kl}(x,\omega), \tag{13}$$

$$\tilde{U}_0^{\mathrm{PVDF}}(\omega) = \tilde{s}^*(\omega)\sum_k |H_{0k}(\omega)|^2, \tag{14}$$

where $\tilde{s}(\omega)$ is the frequency spectrum of the input signal $s(t)$ and * denotes the complex conjugate. Equations (12) and (14) are similar to those obtained with the matched-filter approach (Fink, 1992) or the phase conjugation techniques (Jackson and Dowling 1991), and show that the time reversal operation results in the maximum response $U_0^{\mathrm{PVDF}}(t)$ at the location of the initial excitation. Moreover, Eq. (15) shows that after reemission, the signal measured by the sensor at the location of the initial excitation depends only of the magnitude of the transfer functions $H_{0k}(\omega)$. In other words, the phase of the initial excitation is preserved by the time reversal operation (with a sign change). As it was anticipated, these developments show that the usual time reversal properties still hold for the case of dispersive propagation in a flexural beam. Note that the evanescent waves created by discontinuities in the beam (point masses and active PZT exciter), corresponding to the last two terms in Eq. (7) are not reproduced by the time reversal process, when propaga-

TABLE I. Properties of PVDF sensors.

Length L	0.075 m
Width b_{PVDF}	0.0254 m
Thickness h_{PVDF}	28 μm
Piezoelectric charge coefficient e_{31}	0.055 C/m^2
Capacitance C	380 pF/cm^2

tion occurs over several wavelengths. Therefore, only propagating bending waves are time reversed far from active sources.

IV. TIME REVERSAL EXPERIMENTS

Time reversal experiments were conducted for audible range bending wave propagation in a beam. A 5 m long steel beam (Young's modulus $E=210$ GPa; structural damping $\eta=1\%$; mass density 7850 kg/m^3) of thickness $h=6$ mm and width $b=25.4$ mm was used in the experiments. The structural damping is applied in the form of complex Young's modulus $E^*=E(1+j\eta)$. The beam is suspended by tree nylon threads at different positions in a plane perpendicular to the bending vibration. In order to ensure anechoic terminations and clearly illustrate dispersive propagation in the beam, the beam ends were submerged in large boxes filled with sand. Initial measurements showed that reflected waves from the beam terminations were very small above 100 Hz. Various point masses between 0.2 and 1.31 kg were added at various positions on the beam in order to induce structural heterogeneity. Note that since PZT and PVDF are not punctual, they can theoretically interact with elastic waves as scatterers. PVDF sensors being very thin and light (Table I), their effect on wave propagation can be disregarded. PZT emitters used in this study add significant local mass and stiffness to the beam. To evaluate the dynamic effect of a PZT, a conservative approximation is to consider the PZT as a small variation of thickness and to calculate the reflection coefficient and transmission coefficients in response to an incident wave (Doyle and Kamle 1985). The resulting reflection coefficient (independent of frequency) is thus about 4% and the transmission coefficient is 99.9% (independent of frequency). In comparison, the minimum reflection coefficient (at 100 Hz) of a mass of 0.228 kg is about 28% increasing to 59% at 5 kHz. In consequence, the interaction of PZT emitters with flexural waves in the beam has been ignored.

For the time reversal experiments, the TRM is formed by two pairs of collocated PZT and PVDF transducers on each side of the initial excitation point, as shown in Fig. 1. In a one-dimensional system, two PZT-PVDF pairs are sufficient to observe and reemit all the energy generated by the primary source. The properties of the PZT actuators and PVDF sensors are listed in Tables I and II. An additional PZT actuator was installed on the beam to serve as the primary source. It has the same properties as the TRM sources except that its length is 0.0375 m. The length of the PZT actuators and PVDF sensors of the TRM should be smaller than half the structural wavelength of interest in the beam in order to effectively excite and observe bending waves. The chosen value $L=7.5$ cm is appropriate for frequencies below approximately 2.5 kHz for the beam considered. Above this frequency, the spatial integration inherent to the piezoelectric transducers induces a rapid cutoff of the transducer sensitivity.

TABLE II. Properties of PZT actuators.

Length L	0.075 m
Width b_{PZT}	0.0254 m
Thickness h_{PZT}	0.001 m
Piezoelectric strain coefficient d_{31}	1.71×10^{-10} m/V
Young's modulus E_{PZT}	63 GPa

TABLE III. Experimental configuration for the heterogeneous beam.

Element	Position x (m)
TRM transducer 1 (PZT1, PVDE1)	0
Point mass $m_1=0.228$ kg	1
PZT exciter	3
Point mass $m_2=0.233$ kg	3.5
TRM transducer 2 (PZT2, PVDF2)	3.7

Note also that in reality, the TRM transducers have a transverse sensitivity and they therefore excite and sense small bending vibrations in the width direction of the beam. The previous analytical model is purely one dimensional and thus excludes this effect. Given the relatively narrow beam chosen ($b=25.4$ mm), the transverse TRM sensitivity essentially results in a magnitude deviation between the modeled and measured TRM sensitivities.

The positions of the various elements on the beam are shown on Table III. The TRM transducers were located at approximately 0.5 m from the beam terminations. In the experiments, all input and output signals were sampled at 5 kHz; the useful frequency range is therefore limited by the Nyquist frequency at 2.5 kHz, which corresponds to the sensitivity cutoff of the TRM transducers, as explained previously. A compact time signal (a rectangular signal of duration 0.4 ms) was initially used as an input $s(t)$ to the PZT exciter. This signal was passed through a 0.1–2.49 kHz bandpass filter and amplified to 180 V before driving the PZT exciter. The structural wavelength in the beam calculated from the Euler–Bernoulli theory is 75 cm at 0.1 kHz and 15 cm at 2.49 kHz. These values being large in comparison to both the beam thickness and width, the Euler–Bernoulli assumption for the structure is valid in the frequency range of interest. Moreover, the smaller wavelength is bigger than any added mass, so the "point mass" is also validated. The cutoff frequency at 0.1 kHz is used to filter the low frequency bending waves that are not adequately absorbed by the sand terminations. The response of the two PVDF sensors of the TRM to the primary excitation was measured, stored, time inverted, and used as an input to the two PZT actuators of the TRM. Prior to being time inverted, the sensor outputs were passed through a 0.1–2.49 kHz bandpass filter and amplified. Similarly, the time-inverted inputs to the PZT actuators were bandpassed and amplified through a high voltage piezoelectric amplifier. In order to monitor the beam response under both the primary PZT exciter and the TRM sources, two

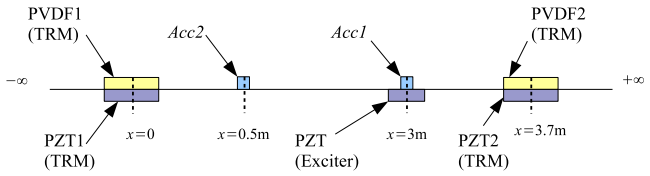


FIG. 2. (Color online) Homogeneous beam configuration for time reversal experiments.

inductively coupled plasma accelerometers were used to measure the transverse acceleration $\ddot{w}_l$ at two locations on the beam. The first accelerometer was collocated with the primary PZT exciter at $x=3$ m, and the second accelerometer was located at $x=0.5$ m or at $x=1.5$ m, between the first TRM transducer and the PZT exciter.

A. Homogeneous beam

Experiments were first conducted on the homogeneous beam (no added point masses) in order to verify the time reversal of dispersive bending waves. In this case, the second monitoring accelerometer was located at $x=0.5$ m. Figure 2 shows the experimental configuration and Figs. 3 and 4 show the measured and predicted filtered acceleration at the first monitoring accelerometer position (collocated with the PZT exciter), and at the second monitoring accelerometer position, under the impulse excitation of the primary PZT exciter. The theoretical results were obtained from multiplication of Eq. (7) with the frequency response of the actual rectangular input signal of the PZT, and using the actual properties of the beam and PZT. In addition, the theoretical results were bandpass filtered using the same filter parameters as in the experiments. Note, however, that the experimental gains of the PZT input and accelerometer output were not taken into account in the model. The theoretical results of Figs. 3 and 4 as well as upcoming figures were thus adjusted such that they match experimental results in magnitude. Also, note that a delay of 0.01 s (the same for all the figures) in the experimental data has been added in the acquisition for triggering. In addition, since the entire process from the direct propagation to the time reversal is done in a single step, additional delays cause the focus instant to be around 0.395 s in all time reversal experiments.

Comparison of Figs. 3(a) and 4(a) clearly shows the dispersive propagation of the pulse between the excitation point and the second monitoring accelerometer, 2.5 m from the source. The low frequency comb filter features of the acceleration frequency spectra are due to the wave reflections from the beam terminations, which are not accounted for in the model. Moreover, the relatively high cutoff of the antialiasing filters (2.49 kHz) as compared to the sampling frequency (5 kHz) results in significant frequency aliasing in high frequency. Despite these imperfections, the agreement between the experimental results and the model is very satisfactory.

Figure 5 shows the time response of PVDF sensor 1 under the pulse excitation of the primary source, as well as the result of the time reversal experiment (after time inversion of the two PVDF signals and reemission by the two PZT

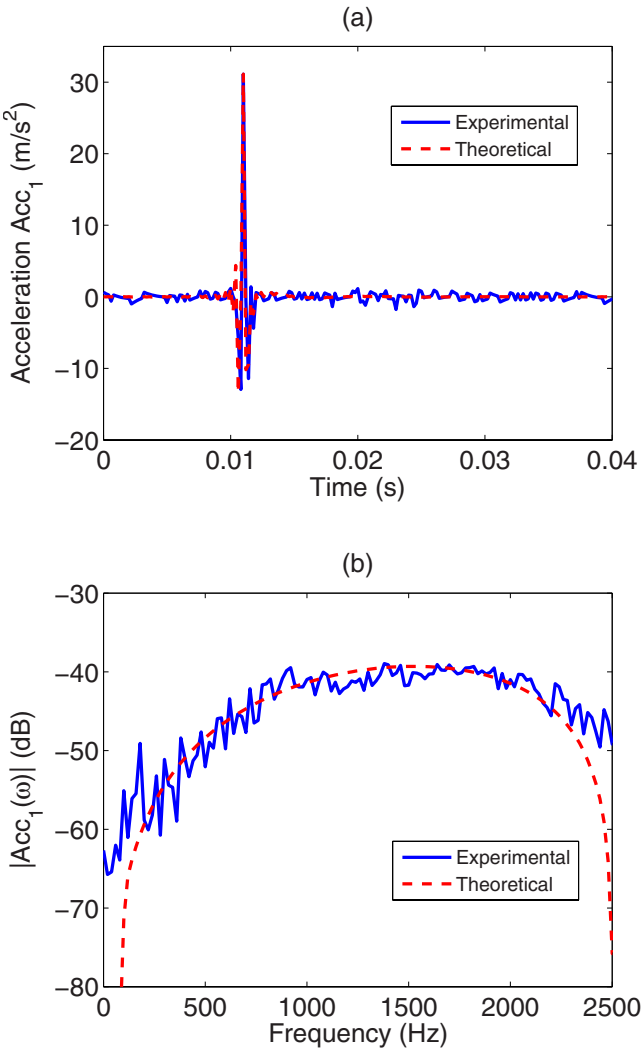


FIG. 3. (Color online) Measured and predicted filtered acceleration of homogeneous beam at the first monitoring accelerometer position (collocated with the PZT exciter), under the impulse excitation of the primary PZT exciter. (a) Time response; (b) frequency response.

actuators) in terms of the accelerations measured at the two monitoring accelerometer locations. The theoretical PVDF response was obtained as the bandpass filtered version of Eq. (9), and the predicted time reversed propagation is obtained as the bandpass filtered version of Eq. (11). While the spreading of the initial pulse is clearly visible in the PVDF sensor signal, the time reversal operation successfully inverts the bending wave dispersion and recontracts the beam response at the location of the initial excitation. In contrast, the acceleration measured by the second monitoring accelerometer close to the TRM transducer 1 shows the two distinct dispersed pulses originating from the two PZT sources of the TRM. The time reversal simulations provided by the model are again in good agreement with the experimental data.

B. Heterogeneous beam

Experiments were conducted on the beam with the added point masses described in Table III in order to verify the time reversal operation in the presence of point scatterers on the structure. In this case, the point masses are located

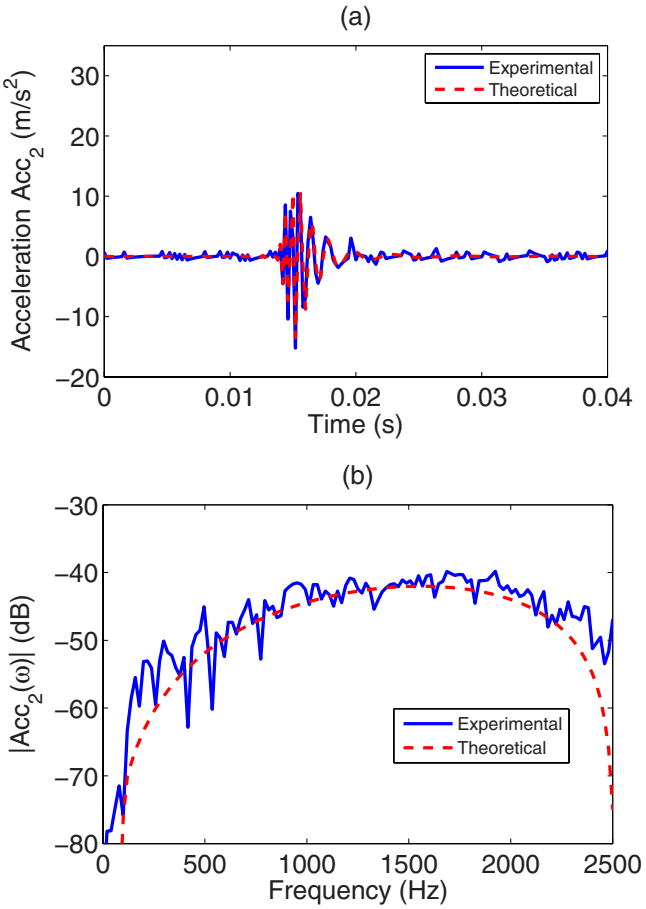


FIG. 4. (Color online) Measured and predicted filtered acceleration of homogeneous beam at the second monitoring accelerometer position (2.5 m from the PZT exciter), under the impulse excitation of the primary PZT exciter. (a) Time response; (b) frequency response.

between the two TRM transducers and the second monitoring accelerometer is located at $x=1.5$ m, between m_1 and the PZT exciter (Fig. 6). Figures 7 and 8 show the beam response under the same impulse excitation of the primary PZT as previously, as well as the response resulting from reemission of time reversed PVDF signals by the two TRM transducers. In contrast to Fig. 3(a), the impulse response at the primary source position on Fig. 7(a) shows multiple reflections due to the point mass discontinuities on the beam. This results in a much longer impulse response of the structure. The signal of PVDF sensor 1 under the pulse excitation of the primary source [Fig. 7(b)] also shows a more complex time response as compared to the homogeneous beam, because of the superposition of wave dispersion and reflections from mass discontinuities.

The response measured after reemission of the time reversed signals shows the significant energy focusing in the time domain at the initial source position [Fig. 8(b)]. The time reversal operation is thus capable of compensating for both the wave dispersion and multiple scattering effects in the beam. In contrast, the signal measured at the second monitoring accelerometer location [Fig. 8(c)] does not reveal such an effective focusing. In contrast to the homogeneous beam situation, the time response at the initial source position contains energy before and after arrival of the large pulse. This can be qualitatively interpreted by the presence of the term $\Sigma_k|H_{0k}(\omega)|^2$ in Eq. (14). This term multiplies the source term $\tilde{s}^*(\omega)$ by transfer functions that inherently contain the multiple reflections due to the point scatterers in the propagation medium.

C. Time reversal as a technique to generate large amplitude impulses in flexural beams at low voltage

This section demonstrates how time reversal can exploit dispersive propagation in flexural beams to generate large amplitude impulses at a given time and position using remote sources. This is illustrated through the experiment depicted on Fig. 9: the configuration is similar to Fig. 6, except that point masses $m_3=1.03$ kg and $m_4=1.31$ kg are now exterior to the TRM limits. Such masses are used to create strong wave reflections and therefore long impulse responses between the initial source and the TRM sensors.

The measured and predicted filtered acceleration at the monitoring accelerometer (collocated with the PZT exciter), under impulse excitation of the primary PZT, is shown in Fig. 10(a). This response was obtained for a 150 V peak input voltage of the PZT exciter; the early impulse response shows distinct dispersed pulses that are reflected by the two point masses. The filtered acceleration at the same location during the time reversed propagation is shown on Fig. 10(b). In this case, the time reversed PVDF signals were amplified to a maximum input voltage of 50 V peak before being applied to the two TRM piezoceramics. Despite the lower actuation of the sources and their larger distance to the monitoring accelerometer during the retropropagation phase, the resulting maximum acceleration on Fig. 10(b) is significantly amplified with respect to Fig. 10(a) (passing from about 30 to 125 m/s^2). The bending wave dispersion in the beam causes the input signals driving the TRM sources to have a much longer duration than the signal driving the PZT exciter; the ability of the time reversal process to invert wave dispersion ensures that the longer impulses originating from the TRM sources recontract in time and space to give a large amplitude, short impulse at the initial source position. Time reversal combined with natural wave dispersion can thus be exploited to generate large amplitude pulses using remote actuators.

V. DETECTION OF POINT SCATTERERS USING TIME REVERSAL

This section examines the detection and localization of heterogeneity in a flexural beam (modeled here as point masses) using a modified version of the DORT method ("Décomposition de l'opérateur de retournement temporel"). The DORT method is extensively discussed in the literature to selectively focus energy on individual scatterers in an acoustic medium (Prada *et al.*, 1996). Scatterer detection has practical applications in detection of vibratory or acoustic sources or detection of defects in structures.

A. The DORT method

To locate a scatterer, the basic time reversal method proposes to measure the time responses of scattered waves with

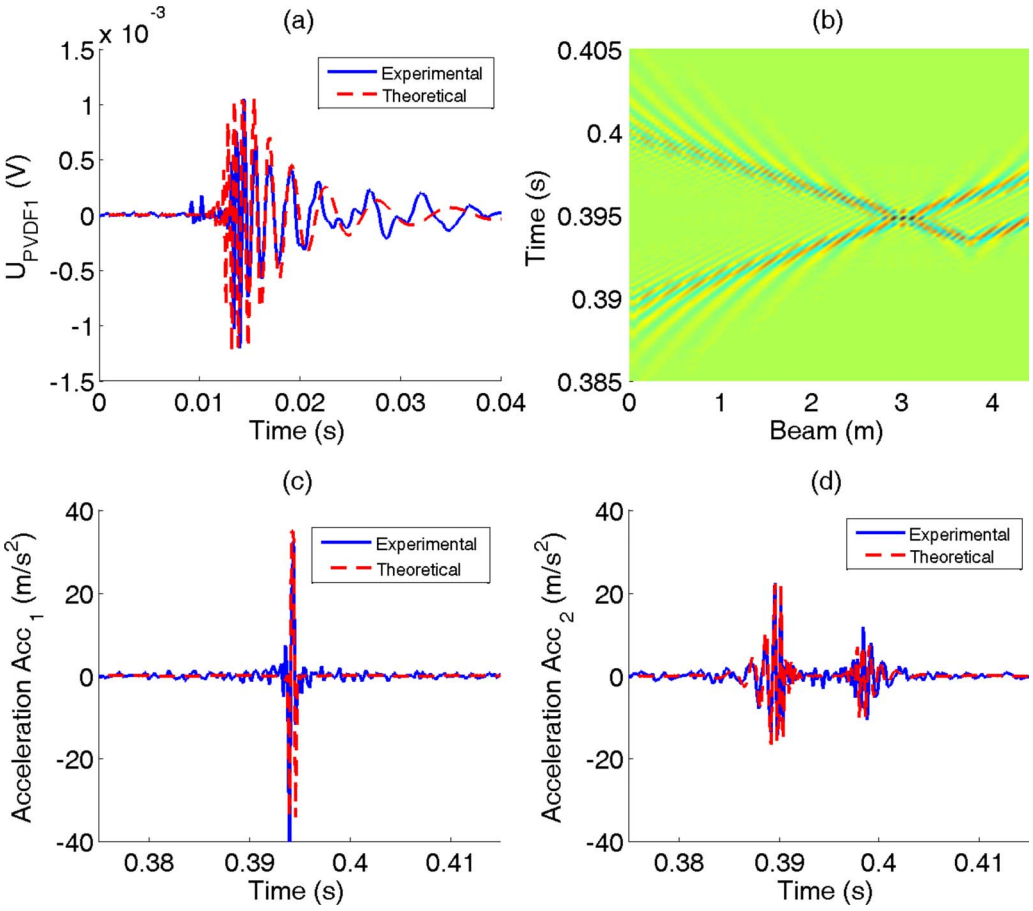


FIG. 5. (Color online) Measured and predicted time reversal results in the homogeneous beam. (a) Time response of PVDF sensor 1 under the pulse excitation of the primary source. Time reversal results. (b) Theoretical temporal and spatial acceleration. (c) Filtered acceleration at the first monitoring accelerometer position (collocated with the PZT exciter at $x=3$ m). (d) At the second monitoring accelerometer position ($x=0.5$ m).

the TRM, and to reemit the time-reversed responses in the medium. This last step can be done experimentally or virtually if a numerical model of wave propagation in the medium is available. The constructive interference of reemitted waves generates a maximum response at the position of the scatterer. This approach, however, has limited ability to resolve various scatterers in the medium. In contrast, the DORT method in general starts with a $N \times N$ matrix $\mathbf{H}(\omega)$ of transfer functions $H_{jk}(\omega)$ between an individual emitter $j=1,\ldots,N$ and an individual sensor $k=1,\ldots,N$ distributed in the propagation medium. The time reversal operator is defined by Prada *et al.* (1996) $\mathbf{O}(\omega)=\mathbf{H}(\omega)\mathbf{H}^*(\omega)$. The eigenvalue decomposition of $\mathbf{O}$ leads to

$$\mathbf{O}(\omega) = \mathbf{Q}(\omega)\mathbf{\Lambda}(\omega)\mathbf{Q}^H(\omega), \tag{15}$$

where $\mathbf{Q}(\omega)$ is the matrix of eigenvectors of $\mathbf{O}$, $\mathbf{Q}^H(\omega)$ is the Hermitian transpose of $\mathbf{Q}$, and $\mathbf{\Lambda}(\omega)$ $=\mathrm{diag}(\lambda_1(\omega),\lambda_2(\omega),\ldots,\lambda_N(\omega))$ is the diagonal matrix of eigenvalues of $\mathbf{O}$. Since $\mathbf{O}$ is Hermitian, the eigenvalues λ_i are real and positive. It has been established (Prada *et al.*, 1996) that for well-resolved scatterers in the propagation medium, each eigenvalue and corresponding eigenvector are associated with a distinct scatterer. More precisely, the eigenvector $\mathbf{q}_i(\omega)$ provides the frequency response of input signals to the N emitters in order to create a focalization on the ith scatterer.

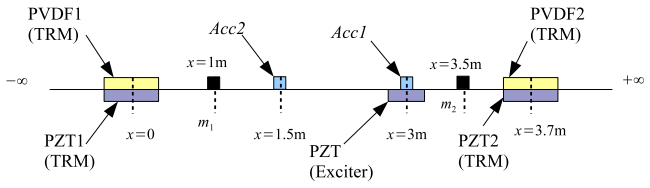


FIG. 6. (Color online) Heterogeneous beam configuration for time reversal experiments.

B. The TRM for scatterer localization

In order to conduct experiments of scatterer localization in the flexural beam, we installed a TRM formed of eight regularly spaced, collocated pairs of piezoceramic (PZT) actuators and piezopolymer (PVDF) sensors. The TRM is depicted on Fig. 11. The PZT actuators and PVDF sensors have the same parameters as those listed in Tables I and II except that their length was reduced to 25 mm. This dimension was chosen such that the sensitivity cutoff of the TRM transducers extends to 5 kHz. The separation between PZT-PVDF pairs was about 60 cm and the distance between the rightmost and leftmost pairs to the anechoic sand terminations was about 30 cm. A point mass $m_1=0.228$ kg was attached to the beam to create a discrete scatterer. In the experiments, the PZT emitters were driven with a short time impulse, and the input and output signals were sampled at 10 kHz with a

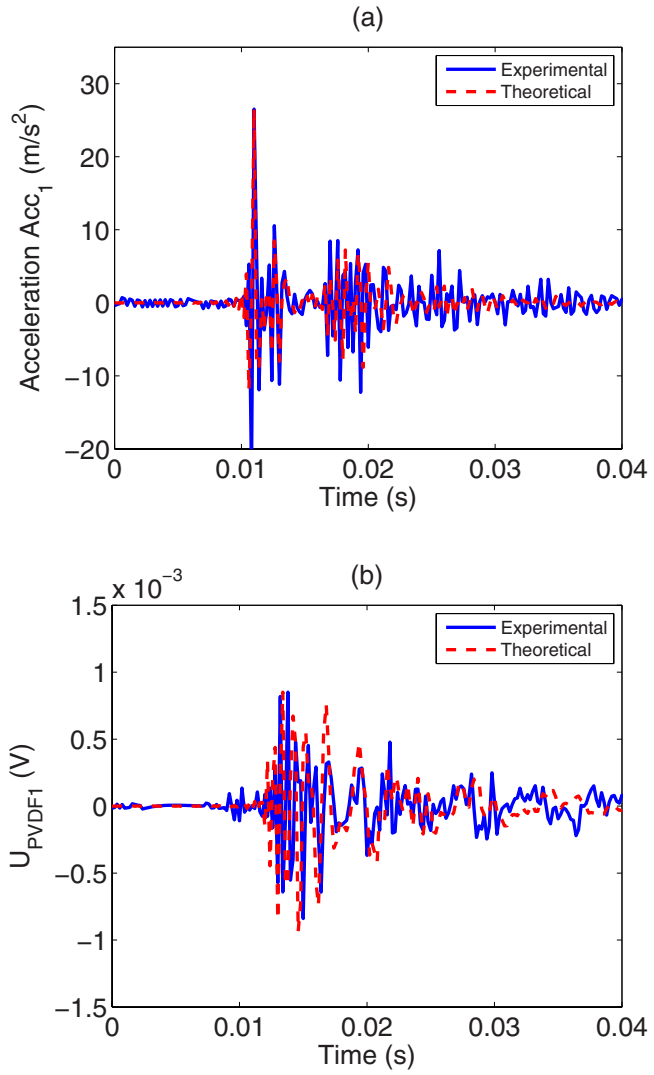


FIG. 7. (Color online) Measured and predicted results in the heterogeneous beam. (a) Filtered acceleration at the first monitoring accelerometer (collocated with the PZT exciter), under impulse excitation of the primary PZT. (b) Time response of PVDF sensor 1 under the pulse excitation of the primary source.

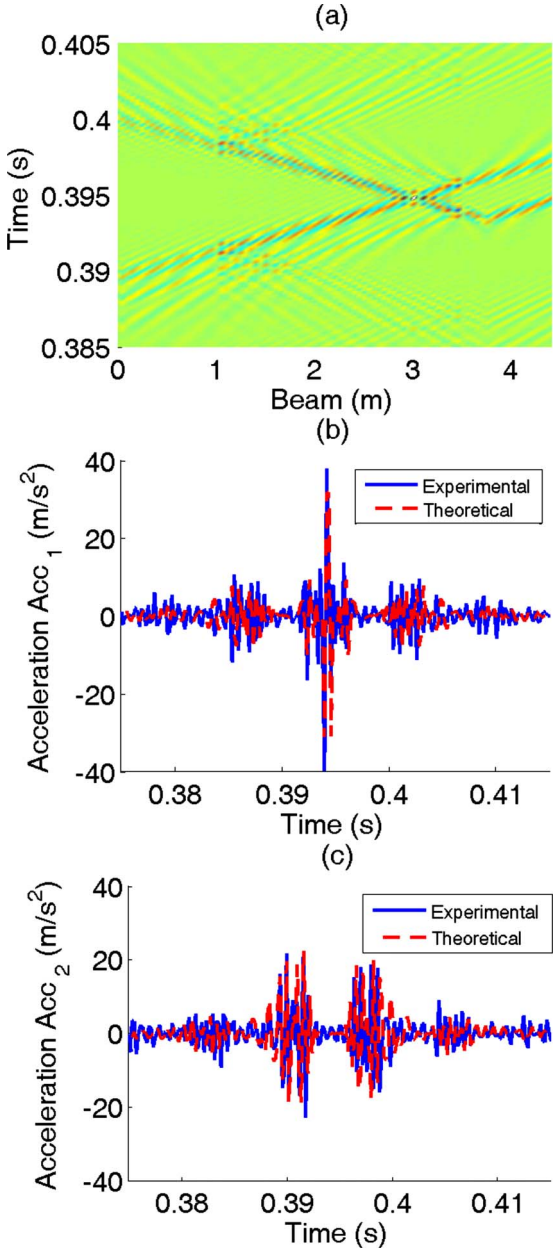


FIG. 8. (Color online) Measured and predicted time reversal results in the heterogeneous beam. (a) Theoretical temporal and spatial acceleration. (b) Filtered acceleration at the first monitoring accelerometer position (x =3 m). (c) At the second monitoring accelerometer position (x=0.5 m).

Brüel&Kjær PULSE Analyzer; the useful frequency range is therefore limited by the Nyquist frequency at 5 kHz. At this frequency, the flexural wavelength in the beam is approximately 10 cm; therefore, the half-wavelength diffraction limit imposes a spatial resolution of no less than 5 cm for localization experiments. The PULSE system has its own antialiasing filter which is automatically set when a useful frequency range is chosen. A high-pass filter was also used for the same reason as in the time reversal experiments.

C. Modified DORT method

In previous work, scatterer localization and selective focusing using the DORT method were typically done in circumstances where the size of the TRM is smaller than the propagation distance between the TRM and the scatterers (Prada *et al.*, 1996). In such situations, the direct propagation path between an individual emitter and an individual receiver of the TRM is significantly shorter than the indirect path caused by the presence of the scatterer. The direct path component is usually omitted in $\mathbf{H}(\omega)$; this can be done easily in experiments using appropriate time windowing of the time responses in order to remove the early, direct propagation paths between elements of the TRM.

In the case of a series of transducers integrated into a

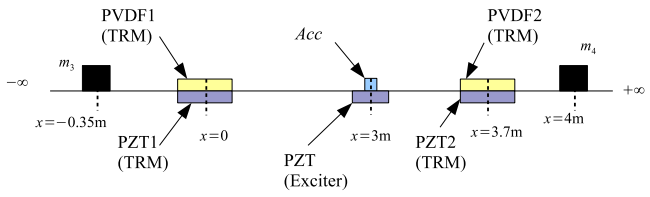


FIG. 9. (Color online) Configuration for the generation of a large amplitude impulse at x=3 m.

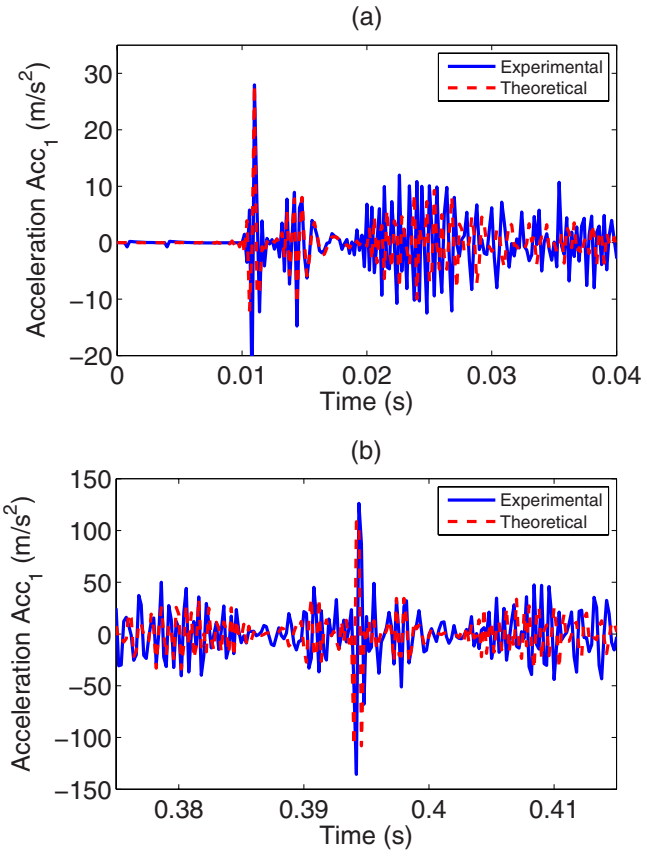


FIG. 10. (Color online) Measured and predicted filtered acceleration at the monitoring accelerometer (collocated with the PZT exciter). (a) Under impulse excitation of the primary PZT. (b) During the time reversed propagation.

beam, this technique is more difficult to implement especially if an impulse excitation is used. Indeed, to locate a scatterer in a one-dimensional medium, the TRM must extend over the beam on both sides of the scatterer in order to exploit constructive interference mechanisms; the direct paths are then not necessarily shorter than the indirect, scattered paths. Furthermore, wave dispersion in a flexural beam can cause low-frequency direct waves to superimpose with high-frequency scattered waves. It is thus impossible to easily remove the direct path contribution in the impulse responses of the TRM.

To correct this problem, a modification of the conventional DORT method, similar to the one proposed by (Derveaux *et al.*, 2007) is employed. The transfer function matrix $\mathbf{H}(\omega)$ is replaced by

$$\mathbf{H}'(\omega) = \mathbf{H}(\omega) - \mathbf{H}_0(\omega), \tag{16}$$

where $\mathbf{H}_0(\omega)$ is the transfer function matrix of the system without scatterers. $\mathbf{H}_0(\omega)$ can be experimentally identified in

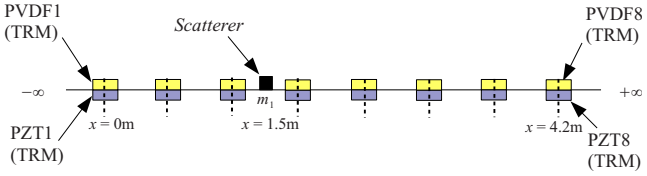


FIG. 11. (Color online) The TRM for scatterer localization experiments on the flexural beam.

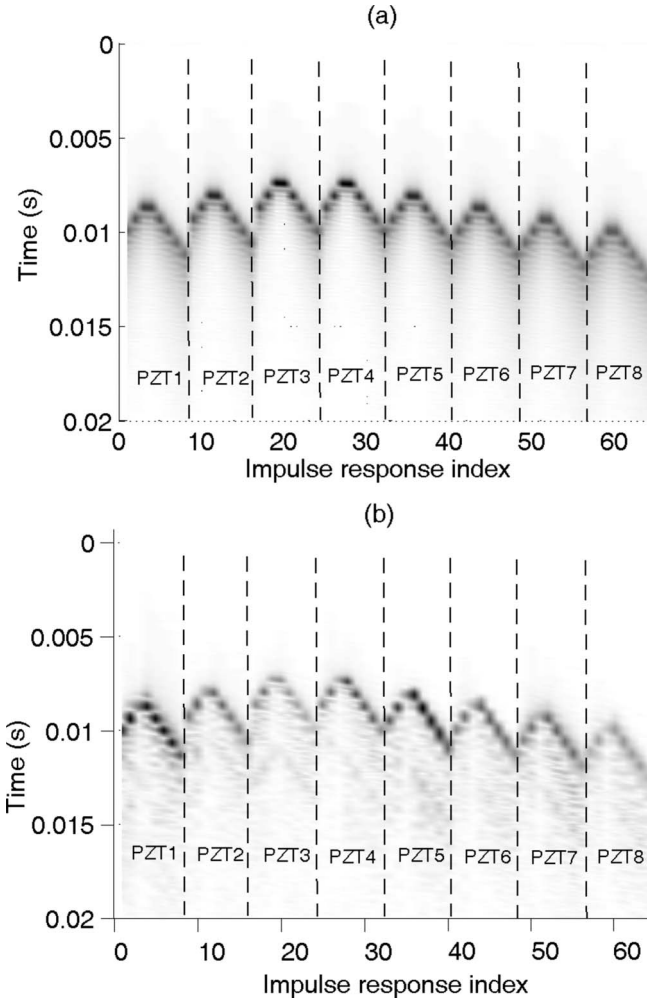


FIG. 12. Impulse responses of the TRM shown in Fig. 11, after subtraction of the homogeneous beam impulse responses. (a) Theoretical; (b) experimental.

a preliminary step or deduced from a numerical model of the system. The subtraction $\mathbf{H}(\omega)-\mathbf{H}_0(\omega)$ removes the direct propagation paths from $\mathbf{H}(\omega)$ and leaves only the scattered paths in $\mathbf{H}'(\omega)$.

D. Results for one scatterer

Figure 12 shows the theoretical and experimental impulse responses $h'_{jk}(t)$ obtained for the heterogeneous beam of Fig. 11. These impulse responses were obtained after subtraction of the homogeneous beam situation, according to Eq. (16), and therefore contain only the waves scattered by the point mass. In Fig. 12, the impulse responses 1 to 8 are relative to the responses of receivers PVDF1 to PVDF8 when the emitter PZT1 is active, impulse responses 9–17 show the responses of PVDF1 to PVDF8 when the emitter PZT2 is active, and so on. The impulse responses are aligned in time such that the successive activations of the PZT correspond to the same time. The results clearly show that the PZT emitters closest to the scatterer (PZT 3 and 4) create early time responses; likewise, PVDF receivers closest to the scatterer (PVDF 3 and 4) record also early time responses. These impulse responses are therefore intrinsic to the scatterer only and do not contain the direct propagation paths.

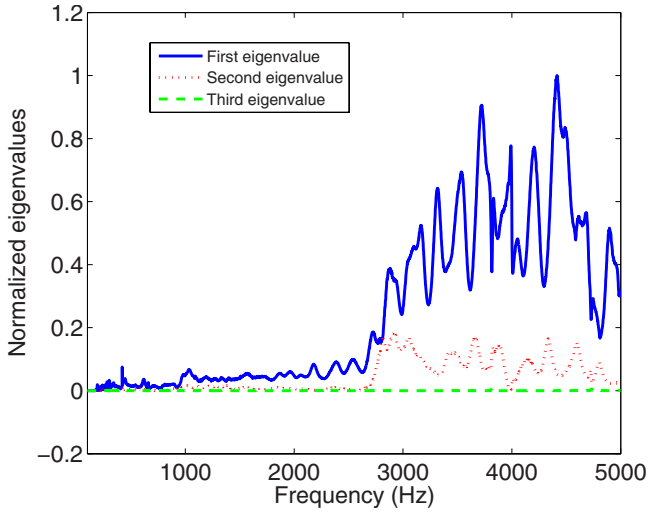


FIG. 13. (Color online) Experimental eigenvalues of the time reversal operator for the beam with one scatterer.

Moreover, the responses relative to emitters far from the point mass (such as PZT8) are more spread in time, because of the dispersion experienced by the initial impulse as it reaches the point mass.

Figure 13 shows the three first eigenvalues $\lambda_i(\omega)$ of the time reversal operator $\mathbf{O}(\omega)=\mathbf{H}(\omega)\mathbf{H}^*(\omega)$ over the frequency range between 0 and 5 kHz. In the case of a single scatterer in the medium, the theory predicts only one nonzero eigenvalue at all frequencies. Eigenvalues deduced from experimental results reveal additional small eigenvalues, referred to as "noise eigenvalues" in the literature (Prada *et al.*, 1996).

Figures 14 and 15 show the propagation in the beam of the first eigenvector $\mathbf{q}_1(\omega)$ of $\mathbf{O}(\omega)$ calculated over the frequency range between 0 and 5 kHz. This result is obtained by calculating the homogeneous beam response to the eigenvector $\mathbf{q}_1(\omega)$ in the frequency domain and applying an inverse Fourier transform. In theory, $\mathbf{q}_1(\omega)$ provides the frequency response of input signals to the eight PZT emitters in order to create a focalization on the scatterer. In our case, this propagation was simulated using the *homogeneous* beam model described in this section and the theoretical or experimental eigenvector $\mathbf{q}_1(\omega)$ as inputs to the PZT emitters. In Figs. 14 and 15, the backpropagation of $\mathbf{q}_1(\omega)$ is shown in terms of the beam transverse acceleration as a function of time [Figs. 14(a) and 14(b)] and position [Figs. 15(a) and 15(b)] along the beam. Figures 14(a) and 15(a) are relative to the theoretical eigenvector $\mathbf{q}_1(\omega)$ whereas Figs. 14(b) and 15(b) are relative to the experimental eigenvector. The position of the scatterer corresponds theoretically to the maximum acceleration amplitude over time and position along the beam. In both the theoretical and experimental results, this maximum is found at $x=1.5$ m, which corresponds to the actual position of the scatterer. The transverse acceleration of the beam as a function of time shows constructive interference of the signals generated by the TRM, resulting in a maximum at the scatterer position [Fig. 14(a)]. This interference is less clear in the experimental results [Fig. 14(b)], but still provides a maximum acceleration at the scatterer location [Fig. 15(b)].

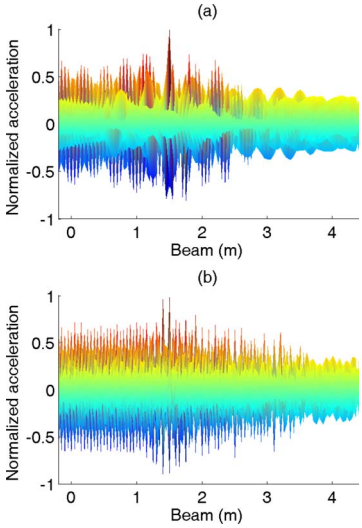


FIG. 14. (Color online) Simulated transverse acceleration of the homogeneous beam vs position, after propagation of (a) the theoretical eigenvector $\mathbf{q}_1(\omega)$ and (b) the experimental eigenvector $\mathbf{q}_1(\omega)$ for the frequency range between 0 and 5 kHz.

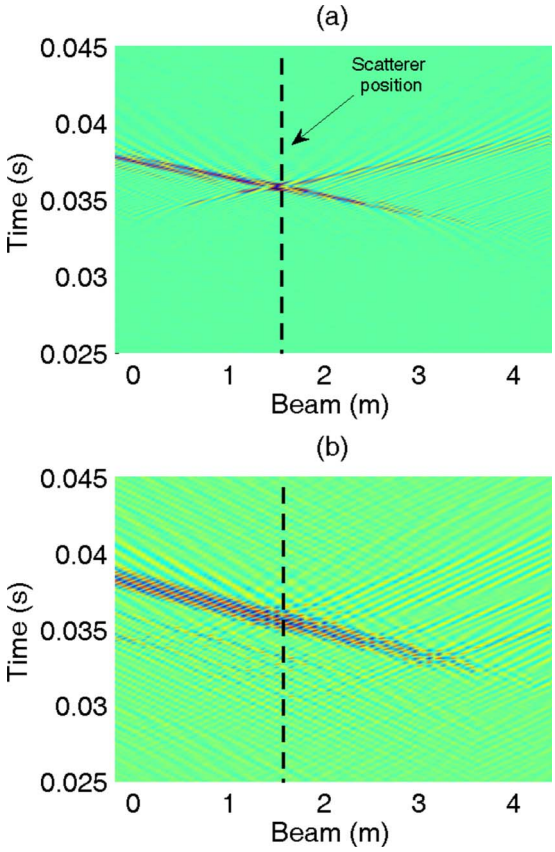


FIG. 15. (Color online) Simulated transverse acceleration of the homogeneous beam vs time and position, after propagation of (a) the theoretical eigenvector $\mathbf{q}_1(\omega)$ and (b) the experimental eigenvector $\mathbf{q}_1(\omega)$ for the frequency range between 0 and 5 kHz.

The DORT method was also investigated to detect and localize two point masses on the beam. In this case, an additional mass of 1.026 kg was mounted on the beam at $x=3.7$ m. The results were inconclusive, as neither scatterer could be detected with the method presented. A possible ex-

planation of this failure is the underlying assumption in the DORT method that distinct scatterers in the medium should be well resolved, meaning that waves scattered by a given scatterer should not be influenced by other scatterers (Prada *et al.*, 1996). This assumption becomes highly questionable for propagation in one-dimensional media such as the flexural beam investigated in this study. One way to evaluate this, suggested by Derveaux *et al.* (2007), is to assess the orthogonality between the scatterers in the frequency range of interest. Since the eigenvectors provide the frequency responses of the TRM, this computation can be done by testing the condition $1/N_\omega \Sigma_{\omega_i} |\mathbf{q}_1^{\mathbf{T}}(\omega_\mathbf{i})\mathbf{q}_2(\omega_\mathbf{i})|/|\mathbf{q}_1(\omega_\mathbf{i})||\mathbf{q}_2(\omega_\mathbf{i})| \ll 1$, where N_ω is the number of frequencies, $\mathbf{q}_1$ and $\mathbf{q}_2$ are the first and second eigenvector, respectively. The theoretical model gives $1/N_\omega \Sigma_{\omega_i} |\mathbf{q}_1^{\mathbf{T}}(\omega_\mathbf{i})\mathbf{q}_2(\omega_\mathbf{i})|/|\mathbf{q}_1(\omega_\mathbf{i})||\mathbf{q}_2(\omega_\mathbf{i})| = 0.2802$, while the experimental values gives $1/N_\omega \Sigma_{\omega_i} |\mathbf{q}_1^{\mathbf{T}}(\omega_\mathbf{i})\mathbf{q}_2(\omega_\mathbf{i})|/|\mathbf{q}_1(\omega_\mathbf{i})||\mathbf{q}_2(\omega_\mathbf{i})| = 0.3581$. These values clearly show that the two scatterers are not well-resolved, making the DORT method not applicable to our situation.

VI. CONCLUSION

This paper investigated time reversal of flexural waves in a beam at audible frequency. A specific objective was to study time reversal under the dispersive conditions and long flexural wavelengths propagation created by vibration of large mechanical structures in the audio range. The TRM used in this study is an array of collocated pairs of piezoelectric actuators and sensors. An analytical model of the beam vibrations and piezoelectric transducers response was developed in order to physically interpret time reversal results and compare to experimental results.

Vibration focusing experiments were successfully conducted on a long flexural beam using time reversal. As predicted by theory, time reversal compensates for the effect of wave dispersion in the beam at the initial source position. Wave dispersion in structures can also be exploited in combination with time reversal to generate high-energy impulse response at a given location on the structures using remote actuators. Detection and localization of heterogeneity in a flexural beam (modelled as point masses) using the DORT method was also examined. To this end, a modification of the original DORT method was proposed in order to eliminate direct propagation paths between the TRM emitters and sensors and enhance the scattered components. This approach successfully identified and located a single point mass scatterer on the beam. However, results for multiple scatterers were inconclusive: In our view, in the case of a one-dimensional medium, the DORT method assumption that individual scatterers should be well resolved is likely to fail.

ACKNOWLEDGMENT

This work was supported by the National Science and Engineering Research Council (NSERC),, Canada.

APPENDIX: WAVE SOLUTION IN THE INFINITE, HETEROGENEOUS FLEXURAL BEAM

In order to solve the flexural response of the infinite heterogeneous beam, it is partitioned into homogeneous segments comprised between the ends of an active PZT source (at x_j and x_j+L) or a point mass discontinuity (at x_i). Figure 16 illustrates a representative situation.

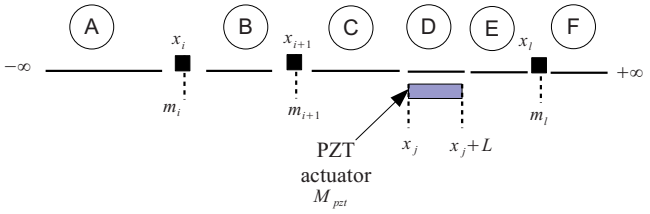


FIG. 16. (Color online) Segmentation of a heterogeneous beam into homogeneous segment at each discontinuities (point mass and PZT ends).

In a given homogeneous segment $x_l \leqslant x \leqslant x_{l+1}$, the transverse displacement is given by

$$\tilde{w}_l(x,\omega) = A_{+l}e^{-jk(x-x_l)} + A_{-l}e^{jk(x-x_{l+1})} + B_l e^{-k(x-x_l)} + C_l e^{k(x-_{l+1})} \quad \text{for } x_l \leq x \leq x_{l+1}. \tag{A1}$$

The following displacement and slope continuity equations apply at $x=x_l$ for all l,

$$w_{l-1}(x_l,\omega) = w_l(x_l,\omega), \tag{A2}$$

$$\frac{\partial w_{l-1}}{\partial x}(x_l,\omega) = \frac{\partial w_l}{\partial x}(x_l,\omega). \tag{A3}$$

The bending moment is continuous at the point mass locations x_i, while it is discontinuous at the PZT ends x_j and x_j+L,

$$EI\frac{\partial^2 w_{i-1}}{\partial x^2}(x_i,\omega) = EI\frac{\partial^2 w_i}{\partial x^2}(x_i,\omega) \quad \text{for all } i, \tag{A4}$$

$$EI\frac{\partial^2 w_j}{\partial x^2}(x_j,\omega) - \frac{\partial^2 w_{j-1}}{\partial x^2}(x_j,\omega) = -M_{\mathrm{PZT}}(\omega)EI\frac{\partial^2 w_{j+1}}{\partial x^2} \times (x_j+L,\omega) - \frac{\partial^2 w_j}{\partial x^2}(x_j+L,\omega) = M_{\mathrm{PZT}}(\omega) \quad \text{for all } j. \tag{A5}$$

Also, the transverse shear force is continuous at the PZT ends, while it is discontinuous at the point mass locations,

$$EI\frac{\partial^3 w_{j-1}}{\partial x^3}(x_j,\omega) = EI\frac{\partial^3 w_j}{\partial x^3}(x_j,\omega)EI\frac{\partial^3 w_j}{\partial x^3}(x_j+L,\omega) = EI\frac{\partial^3 w_{j+1}}{\partial x^3}(x_j+L,\omega) \quad \text{for all } j, \tag{A6}$$

$$EI\frac{\partial^3 w_{i-1}}{\partial x^3}(x_i,\omega) - EI\frac{\partial^3 w_i}{\partial x^3}(x_i,\omega) = -m\omega^2\tilde{w}(x_i,\omega) \quad \text{for all } i. \tag{A7}$$

Finally, the anechoic conditions at the beam ends enforce the conditions $A_{+l}=B_l=0$ at the leftmost beam segment and $A_{-l}=C_l=0$ at the rightmost beam segment. These four conditions, together with conditions (A2)–(A7), provide the equations necessary to solve the unknown wave amplitudes A_{+l}, A_{-l}, B_l, C_l in Eq. (A1).

Bavu, E., Besnainou, C., Gibiat, V., Rosny, J., and Fink, M. (**2002**). "Subwavelength Sound Focusing Using a Time-Reversal Acoustic Sink," Acta. Acust. Acust. **93**, 706–715.
Candy, J. V., Meyer, A. W., Poggio, A. J., and Guidry, B. L. (**2004**). "Time-reversal processing for an acoustic communications experiment in a highly reverberant environment," J. Acoust. Soc. Am. **115**, 1621–1631.
Derveaux, G., Papanicolaou, G., and Tsogka, C. (**2007**). "Time reversal imaging for sensor networks with optimal compensation in time," J. Acoust. Soc. Am. **121**, 2071–2085.
Dickey, J. (**2000**). "Determining impact source location in structural networks," J. Acoust. Soc. Am. **107**, 2884 (**2000**).
Doyle, J. F., and Kamle, S. (**1985**). "An Experimental Study of the Reflection and Transmission of Flexural Waves at Discontinuities," J. Appl. Mech. **52**, 669–673.
Draeger, C., Cassereau, D., and Fink, M. (**1997**). "Theory of the time-reversal process in solids," J. Acoust. Soc. Am. **102**, 1289–1295.
Draeger, C., Aime, J.-C., and Fink, M. (**1999**). "One-channel time-reversal in chaotic cavities: Experimental results," J. Acoust. Soc. Am. **105**, 618–625.
Fink, M. (**1992**). "Time reversal of ultrasonic fields—Part I: Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566.
Fink, M. (**1993**). "Time-reversal mirrors," J. Phys. D **26**, 1333–1350.
Fink, M., Kuperman, W. A., Montagner, J.-P., Tourin, A. (**2002**). *Imaging of Complex Media with Acousic and Seismic Waves* (Springer-Verlag, Heidelberg).
Gibbs, G. P., and Fuller, C. R. (**1992**). "Excitation of thin beams using asymmetric piezoceramic actuators," J. Acoust. Soc. Am. **92**, 3221–3227.
Graff, K. F. (**1991**). *Wave Motion in Elastic Solids* (Dover, New York).
Ing, R. K., and Fink, M. (**1998**). "Time-reversed Lamb waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1032–1043.
Jackson, D. R., and Dowling, D. R. (**1991**). "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181.
Lee, C. K., and Moon, F. C. (**1990**). "Modal sensors/actuators," ASME J. Appl. Mech. **57**, 434–441.
Montaldo, G., Roux, P., Derode, A., Negreira, C., Fink, M. (**2001**). "Generation of very high pressure pulses with 1-bit time reversal in a solid waveguide," J. Acoust. Soc. Am. **110**, 2849–2857.
Montaldo, G., Tanter, M., and Fink, M. (**2004**). "Revisiting iterative time reversal processing: Application to detection of multiple targets," J. Acoust. Soc. Am. **115**, 776–784.
Prada, C., Manneville, S., Spoliansky, D., and Fink, M. (**1996**). "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2076.
Prada, C., Kerbrat, E., Cassereau, D., and Fink, M. (**2002**). "Time reversal techniques in ultrasonic nondestructive testing of scattering media," Inverse Probl. **18**, 1761–1773.
Puckett, A. D., and Peterson, M. L. (**2003**). "A time-reversal mirror in a solid circular waveguide using a single, time-reversal element," ARLO **4**, 31–36.
Wang, C. H., Rose, J. T., and Chang, F.-K. (**2004**). "A synthetic time-reversal imaging method for structural health monitoring," Smart Mater. Struct. **13**, 415–423.
Yon, S., Tanter, M., and Fink, M. (**2003**). "Sound focusing in rooms: The time-reversal approach," J. Acoust. Soc. Am. **113**, 1533–1543.

Prediction of the acoustic form function by neural network techniques for immersed tubes

A. Dariouchy[a] and E. Aassif
Laboratoire de Métrologie et Traitement de l'Information, Département de Physique, Université Ibn Zohr Faculté des Sciences, B.P. 8106, 80000 Agadir, Morocco

G. Maze[b] and D. Décultot
Laboratoire d'Acoustique Ultrasonore et d'Electronique, LAUE UMR CNRS 6068, Université du Havre, 76610 Le Havre, France

A. Moudden
Laboratoire de Métrologie et Traitement de l'Information, Département de Physique, Université Ibn Zohr Faculté des Sciences, B.P. 8106, Agadir, Morocco

(Received 13 June 2007; revised 21 March 2008; accepted 22 May 2008)

A new approach is used to predict the acoustic form function (FF) for an infinite length cylindrical shell excited perpendicularly to its axis using the artificial neural network (ANN) techniques. The Wigner–Ville distribution is used like a comparison tool between the FF calculated by the analytical method and that predicted by the ANN techniques for a stainless steel tube. During the development of the network, several configurations are evaluated for various radius ratios b/a (a: outer radius: b: inner radius of the tube). The optimal model is a network with one hidden layer. It is able to predict the FF with a mean relative error about 1.61% for the cases studied in this paper.
 [DOI: 10.1121/1.2945164]

PACS number(s): 43.60.Np, 43.20.Fn, 43.60.Hj, 43.20.Ks [WMC] Pages: 1018–1025

I. INTRODUCTION

Many studies, theoretical and experimental, show that acoustic resonances of a target are related to its physical and geometrical properties. Conversely, starting from these resonances, we can characterize the material constituting a target with a known geometry. Among all the targets of simple geometrical forms (plate, cylinder, sphere, tube, etc.), tubes are the subject of a few studies of characterization.[1–12] If an air-filled tube immersed in water is excited by a plane acoustic wave perpendicularly to its axis, circumferential waves are generated in the shell and in the water/shell interface.[13,14] For some frequencies, these circumferential waves form stationary waves on the circumference of the tube constituting resonances.[14] The mode n of a resonance is the number of wavelengths around the circumference. These resonances are observed on the spectrum of the acoustic pressure backscattered (or the form function) by the tube.[3,5] The circumferential waves can belong to two wave types that are equivalent to the Lamb waves on a plate if the shell wall is thin: the antisymmetric (A_i) and symmetric (S_i) circumferential waves ($i=0,1,2,\ldots$: index of wave).[5,9,13] For a tube made of a given material, the reduced resonance frequencies of these waves essentially depend on the radius ratio b/a (a: outer radius; b: inner radius of tube).[6,8] The resonance modes n of a surface wave line up on trajectories (n as function of the reduced frequency called Regge trajectories).[5,15] For $i \geqslant 1$, these trajectories have reduced cutoff frequencies, which depend on the physical characteristics of the tube.[16] The aim of this paper is to compare the form function predicted by the neural network techniques with that obtained analytically. The Wigner–Ville time-frequency distribution was used as a comparison tool to check the validity of the use of the neural network model to predict the form function.[16] The analysis of this time-frequency distribution takes into account both the time parameter and the frequency parameter, leading to synthetic images that allow us to follow the evolution of the frequential content of a wave echo as a time function.[17,18] The artificial neural networks (ANNs) are a tool of statistical analysis that builds a model of behavior starting from a certain number of examples.[19–21] Major benefits using a neural network are excellent management of uncertainties, noisy data, and nonlinear relationships. In this study, the ANN model predicts the form functions of tubes without the use of the analytical method for various radius ratios b/a (a: outer radius; b: inner radius). A model at seven entries corresponding to radius ratio b/a and the six previous values of the form function calculated analytically is developed. This model is able to predict the FF for stainless steel tubes of various radius ratios b/a between 0.9 and 0.99.

II. BACKSCATTERED ACOUSTIC SIGNAL FROM A TUBE

The scattering of a plane wave by an infinite length cylindrical shell with a radius ratio b/a is investigated through the solution of the wave equation and the associated boundary conditions. Figure 1 shows the direction of a plane wave incident on an infinite length cylindrical shell in a fluid medium. The axis of the cylindrical shell is identical to the z-axis of the cylindrical coordinate system (r,θ,z).

[a]Electronic mail: abdelilah_dariouchy@yahoo.fr
[b]Electronic mail: gerard.maze@univ-lehavre.fr

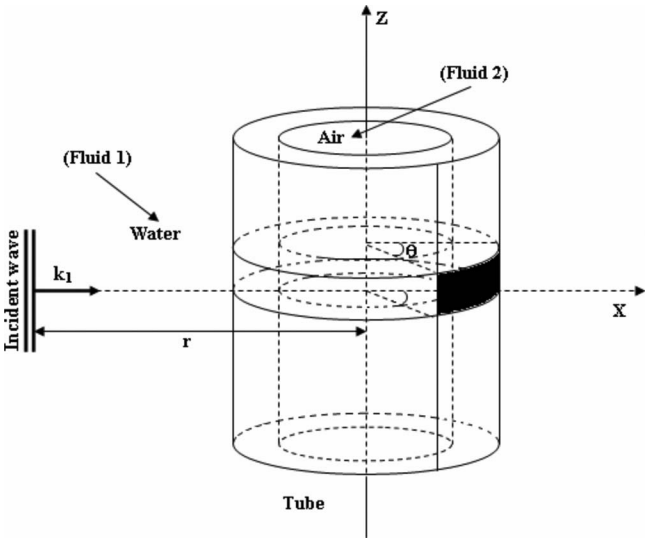


FIG. 1. The geometry used to calculate the form function for an elastic tube.

The acoustic plane wave with frequency ω ensonifies the cylindrical shell normally to the z-axis. Fluid 2 filling the cavity of the shell has a density ρ_2, and the velocity of longitudinal wave in this fluid is c_2. In general, the outer fluid, labeled fluid 1, is different and its density is ρ_1 and the velocity of wave is c_1.

In this study, an air-filled tube immersed in water is excited by a plane acoustic wave perpendicularly to its axis (Fig. 1). The complex pressure P_{scat}^{∞} backscattered by the tube in a far field is the sum of the normal modes, which takes into account the effects of the incident wave, the reflective wave (1), circumferential waves in the shell (2) [A_i, S_i ($i=0,1,2,\ldots$): equivalent to Lamb waves if the tube wall is thin], and an antisymmetrical interface Scholte A wave labeled also A_0^- wave[22] (3) connected to the geometry of the object (Fig. 2). In our case, the flexural waves A, A_1, and A_2 and the compression waves S_0, S_1, and S_2 can be observed.

The general analytical form of the backscattered pressure P_{scat}^{∞} in far field at normal incidence can be expressed as

$$P_{\text{scat}}^{\infty}(\omega) = P_0 \frac{1-i}{\sqrt{\pi k_1 r}} \exp i(k_1 r - \omega t) \sum_{n=0}^{\infty} \varepsilon_n \frac{D_n^1(\omega)}{D_n(\omega)} \cos(n\theta), \tag{1}$$

where ω is the angular frequency, k_1 is the wave number with respect to the wave velocity in the external fluid, P_0 is

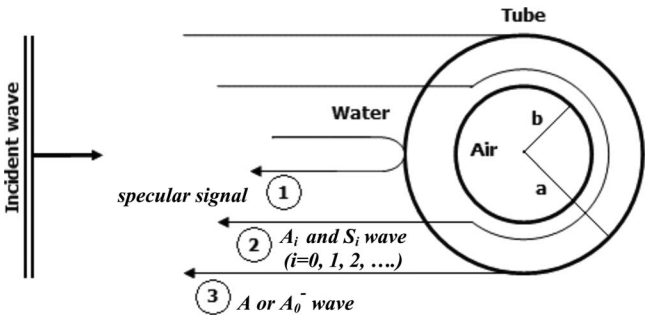


FIG. 2. Mechanisms of the echo formation (1: specular reflection, 2: circumferential shell waves, and 3: Scholte A or A_0^- wave).

TABLE I. Physical parameters.

	Density ρ (kg/m^3)	Longitudinal velocity C_L (m/s)	Transversal velocity C_T (m/s)
Steel	7800	5880	3140
Water	1000	1470	⋯
Air	1.29	334	⋯

the amplitude of the incident plane wave, r is the distance of pressure measurement, θ is the azimuthal angle ($\theta=180°$ for the backscattering spectrum), $D_n^1(\omega)$ and $D_n(\omega)$ are determinants computed from the boundary conditions of the problem (continuity of stress and displacement of both interfaces)—the terms of these determinants are given in Ref. 7 (in the present study, the normal incidence is only considered)—and ε_n is the Neumann coefficient ($\varepsilon_n=1$ if $n=0$ and $\varepsilon_n=2$ if $n \neq 0$).

The physical parameters used in the calculation of the backscattered complex pressure are illustrated in Table I.

Usually this backscattered pressure is presented as the FF defined by[23]

$$\left| \frac{P_{\text{scat}}^{\infty}}{P_0} \right| = \sqrt{\frac{a}{2r}} \text{FF}. \tag{2}$$

This form function is a function of the reduced frequency $x_1=k_1 a$ (a: outer radius of tube) given by

$$x_1 = k_1 a = \frac{2\pi \nu a}{c_1}, \tag{3}$$

where $\nu=\omega/2\pi$ is the frequency of a wave in hertz.

Figure 3 shows an example of results computed for a stainless steel tube ($b/a=0.96$). The backscattered pressure spectrum is presented in Fig. 3(a). It is obtained from the computation of the FF [Eq. (2)]. From the backscattered complex pressure, the impulse time response is calculated with an inverse Fourier transform [Eq. (4)]; the function $h(\omega)$ is the bandpass of transducers used in the experiments,

$$P_{\text{scat}}^{\infty}(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} h(\omega) P_{\text{scat}}^{\infty}(\omega) e^{-i\omega t} d\omega. \tag{4}$$

Examples of results are given in Refs. 6 and 22. To obtain a resonance spectrum using "resonance scattering theory,"[2,3] the theoretical studies indicate that it is possible to suppress the rigid background for the thick cylindrical shells, the soft background for the very thin shells, and the intermediate background for the other tubes.[24] This calculus allows us to suppress the reflected signal on the tube. It is possible to obtain the resonance spectrum using another method developed for the first time in experiment[25] and applied to theoretical results. After we have calculated the FF with Eq. (2), the impulse time signal is obtained with an inverse Fourier transform. Then a new Fourier transform is applied to the impulse time signal in which the specular echo (reflected echo) is suppressed and replaced by zeros. This technique supresses the real specular reflection, and the obtained spectrum is the resonance spectrum presented in Fig. 3(b). The amplitude transitions on the backscattering spectrum [Fig.

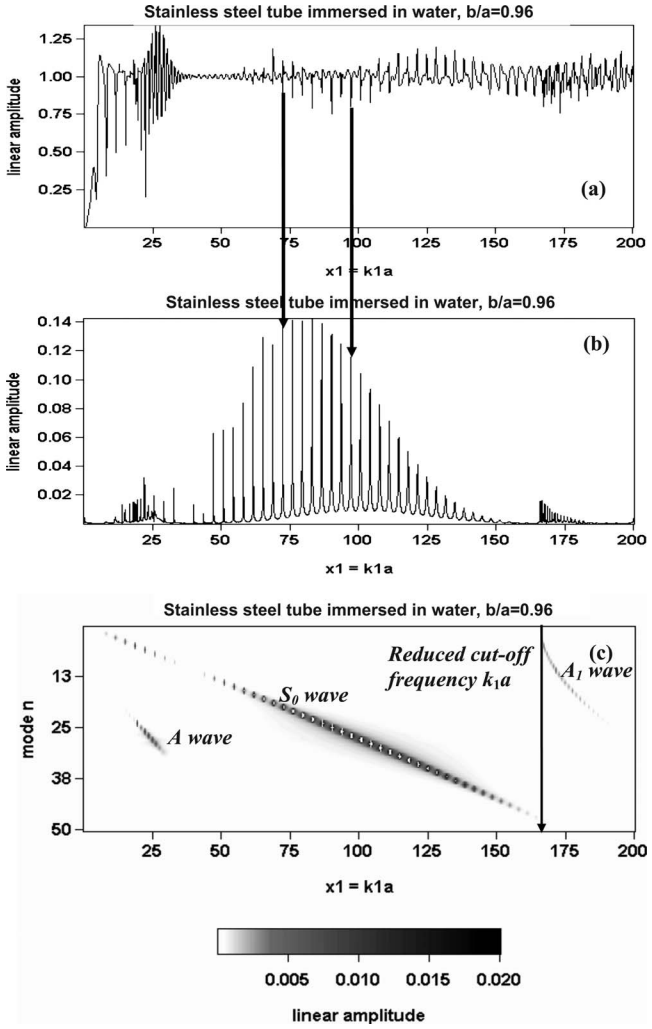


FIG. 3. (a) Form function spectrum, (b) resonance spectrum, and (c) trajectories of modes n for an air-filled stainless steel cylindrical shell (b/a =0.96) immersed in water. Two arrows indicate the relation between transition (a) and peak (b). The reduced cutoff frequency of the A_1 wave is indicated by a vertical line (c).

3(a)] are in front of the peaks observed in Fig. 3(b) (two arrows give examples of relation). These peaks are due to the resonances of modes n, they are connected with the propagation of circumferential waves (A, S_0, and A_1 waves).[13,26] When a circumferential wave is generated on the shell, a resonance is established when the number of wavelengths on the circumference of the tube is an integer; this integer is mode n. To plot Fig. 3(c), the resonance spectra are calculated for each value of n with the previous technique and associated in plane "n-k_1a" with amplitude in the gray level. The trajectories of modes n of resonances in the same frequency domain ($0<x_1<200$) are Regge trajectories.[5,15] Each trajectory is related to a wave type (A,S_0,A_1). The trajectory in relation to the A_1 wave has a cutoff frequency for the small values of mode n indicated by an arrow in Fig. 3(c).

III. WIGNER–VILLE TIME-FREQUENCY DISTRIBUTION

The Wigner–Ville distribution (WVD) presents some convenient properties of applications such as energy conservation, time and frequency shift invariance, and preservation of the time duration and bandwidth.[18]

The time-frequency analysis takes into account both the time parameter and the frequency parameter, leading to an image that allows us to follow the evolution of the frequency content of acoustic echoes as a function of time.[27]

Among the time-frequency techniques, WVD is used for its interesting properties in terms of acoustic applications.[16] Other time-frequency methods (wavelets, for example) would be used in our application, but we had good results with the WVD.[17] The WVD is applied to obtain the dispersion curves of the group and phase velocities of circumferential waves propagating around a tube with different radius ratios.[16] The WVD can be applied to the backscattered time signal obtained from the computation and/or the experiment. This allows us to determine the reduced cutoff frequencies.[16]

The WVD of time complex signal $u(t)$ displays the energy distribution of this signal in the time-frequency plane [Eq. (5)].

$$W_u(t,\nu)=\int_{-\infty}^{+\infty} u\left(t+\frac{\tau}{2}\right)u^*\left(t-\frac{\tau}{2}\right)\exp(-i2\pi\nu\tau)d\tau, \quad (5)$$

with $\nu=\omega/2\pi$ as the frequency, t as the time of the signal, and $u^*(t)$ as the conjugated complex time signal of $u(t)$.

IV. ARTIFICIAL NEURAL NETWORKS APPROACH: MULTILAYER PERCEPTRON (MLP)

The ANN technology is an alternate computational approach inspired by studies of the brain and nervous systems.[19] The advantage of neural networks is that they are capable of linear and nonlinear modelings of systems.[28,29]

In the present study, we use a multilayer perceptron (MLP) trained with a backpropagation algorithm of gradient to predict the form function for stainless steel tubes.[30,31] The MLP architecture comprises mainly parallel adaptive processing elements with hierarchical structured interconnected networks. Each processing unit of a MLP has multiple inputs and a single output.

To be able to find a radius ratio b/a corresponding to an unknown form function versus the reduced frequency k_1a ($0<k_1a<200$, $\Delta k_1a=0.1$), the analytical method requires days to calculate a set of the form functions and to compare them thereafter with the unknown form function. However with the neural network method in only a few minutes we can achieve the same work in a very precise way for radius ratio to three or four decimal places (e.g., $b/a=0.956$).

The relationship between the input and output signals is usually formulated as follows:[32]

$$O_j=f(y_j)=\frac{1}{1+\exp(-y_j)}, \quad (6)$$

where f is a sigmoidal function defined by the relation

$$f(x)=\frac{1}{1+e^{-x}}.$$

The variable x depends on the network input to the node. $\Sigma_{i=1}^{l} w_{ij}I_i$ is a bias term; hence

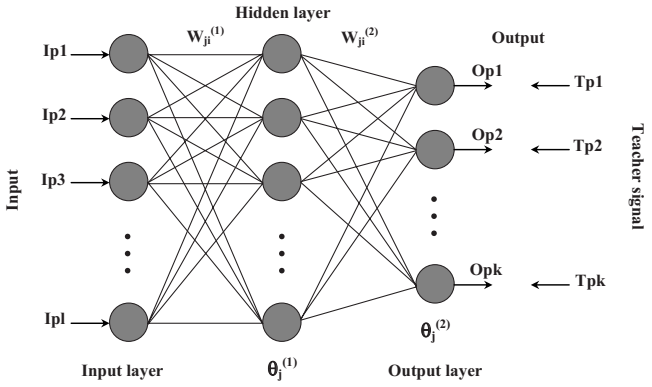


FIG. 4. Layout of the three-layer back propagation neural network.

$$y_j = \sum_{i=1}^{l} w_{ij} I_i + \theta_j,$$

where O_j is the output signal of the jth unit, y_j the potential of the jth unit, w_{ij} are the connection weights between ith and jth units, θ_j is the threshold value or the bias of the jth unit, and l is the number of input signals.

Figure 4 shows a three-layer neural network. All the units are formed into a multiple layers, i.e., an input layer, a hidden layer, and an output layer. The basic idea of training a neural network is as follows. First, the square error of the pth training pattern E_p is defined as

$$E_p = \frac{1}{2}\sum_{k=1}^{m} (T_{pk} - O_{pk})^2, \tag{7}$$

where T_{pk} is the teacher signal (desired output) of the kth output unit for the pth training pattern, O_{pk} is the output signal of the kth output unit for the pth training pattern, and m is the number of output units. In the training process, w_{ji} and θ_i are modified repeatedly based on the gradient descent method to minimize the above error. This modification proceeds downward. Through such an iterative process, the network attains the ability to promptly output the same signal to the teacher's. This training algorithm is called back-propagation neural network.[21] When the network synaptic weights of the hidden layers are not in the linear part of the sigmoidal function, there will be an explosion of the weights.

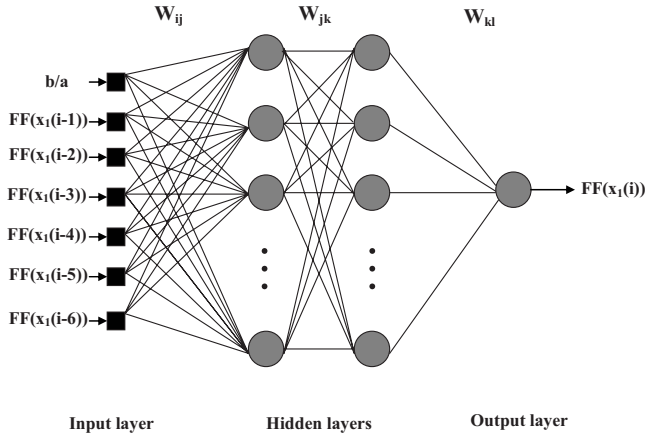


FIG. 5. Architecture of the neural network model used in this study.

Not to leave the linear interval of sigmoid, the synaptic weights are initialized randomly between −0.5 and +0.5.

V. METHODOLOGY

A. ANN model and training phase

The neural network method requires for its training a set of form functions calculated by the analytical method or obtained by experiments. Before any training, it is essential to normalize all the data set. The input and output data of the network are normalized between 0.1 and 0.9 using the following equation:[32]

$$y = 0.1 + 0.8\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right), \tag{8}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of a particular parameter in the entire data set, respectively.

The neuron number in each hidden layer can vary according to the complexity of the problem and the data set. In this work, the data set is constituted from the form functions calculated by the analytical method. This data set is divided into two sets. The training and testing sets are made up respectively of 60% (6 FF) and 40% (4 FF) of the collected data set. The input vector of the MLP is represented by the radius ratio b/a and the FF values for the preceding six indices of dimensionless frequency $[x_1(i)]$, [i.e., $x_1(i-1)$, $x_1(i-2)$, $x_1(i-3)$, $x_1(i-4)$, $x_1(i-5)$, and $x_1(i-6)$] (Fig. 5). Accordingly, the output vector of the MLP represents the expected FF for frequency index $x_1(i)$.[33–36]

The ANN model can be represented by the following compact form (Fig. 5):

$$(\mathrm{FF})_{x_1(i)} = \mathrm{ANN}\left[\frac{b}{a}, (\mathrm{FF})_{x_1(i-1)}, (\mathrm{FF})_{x_1(i-2)}, (\mathrm{FF})_{x_1(i-3)}, (\mathrm{FF})_{x_1(i-4)}, (\mathrm{FF})_{x_1(i-5)}, (\mathrm{FF})_{x_1(i-6)}\right], \tag{9}$$

Once the data set is normalized, it remains to train the

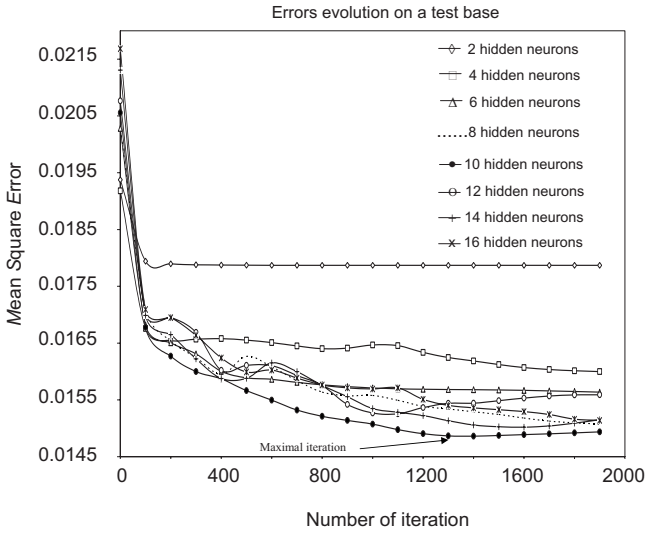


FIG. 6. Visualization of training and testing errors as a function of the number of iterations for various ANN models on a test base.

TABLE II. Statistical accuracy measures of the ANN model at the testing phase.

Number of neurons in the hidden layer	MAE ($10^{-3}k_1a$)	MRE (%)	SE ($10^{-4}k_1a$)	R
2	8.73	2.13	2.00	0.910
4	7.45	1.77	1.79	0.933
6	6.99	1.70	1.75	0.935
8	6.92	1.68	1.69	0.940
10	6.72	1.61	1.67	0.941
12	7.87	1.85	1.74	0.936
14	7.38	1.75	1.69	0.940
16	6.79	1.63	1.69	0.940

multilayer perceptron using the back-propagation algorithm of the gradient to predict the form function.

The training phase is made only once and requires a few minutes. Once the network is trained, the synaptic weights are fixed and will be used thereafter to predict any form function included in the b/a interval [0.9–0.99].

B. Selection of the optimal model

The dilemma bias variance makes it possible to find a model that carries out the best compromise between the capabilities of training and generalization. The bias term expresses the average on all the possible bases of training of the difference square between the predictions of the model, while the variance term expresses the sensitivity of the model to the whole of data used for the training.[37] The selection of models is done, comparing the errors of each ANN configuration, calculating the mean relative error (MRE), the mean absolute error (MAE), and the standard error (SE) of the form function. The coefficient of correlation (R) is used like measures of the performance of the model between the predicted outputs by ANN and the anlytically calculated outputs. The following relations give the errors and the coefficient of correlation:

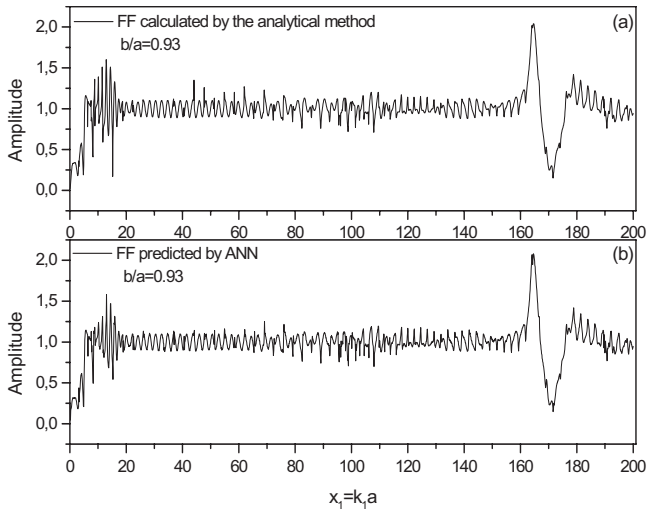


FIG. 7. Form function (a) calculated by the analytical method and (b) predicted by the ANN method (for $b/a=0.93$).

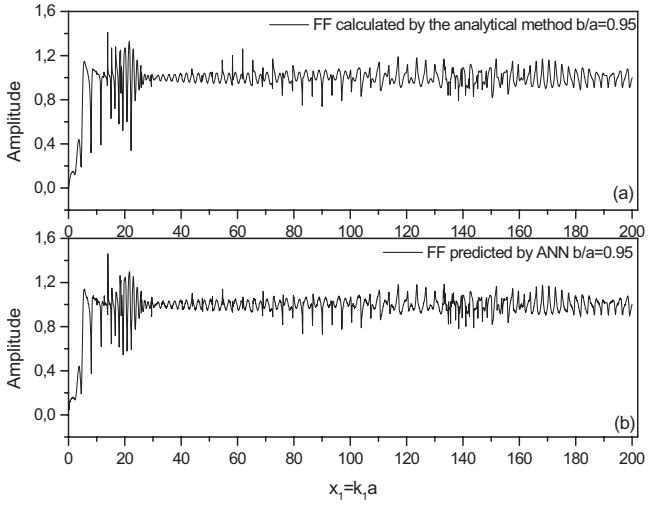


FIG. 8. Form function (a) calculated by the analytical method and (b) predicted by the ANN method (for $b/a=0.95$).

$$\mathrm{MRE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|C_i - O_i|}{C_i} \times 100, \tag{10}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|C_i - O_i|, \tag{11}$$

$$\mathrm{SE} = \frac{\sqrt{\Sigma_{i=1}^{N}(C_i - O_i)^2}}{N-1}, \tag{12}$$

$$R = 1 - \frac{\Sigma_{i=1}^{N}(C_i - O_i)^2}{\Sigma_{i=1}^{N}(C_i - O_m)^2}, \tag{13}$$

where N is the number of data, C_i and O_i are, respectively, the analytically calculated form function values and the predicted form function values by ANN and O_m is the mean of the predicted form function values by ANN.

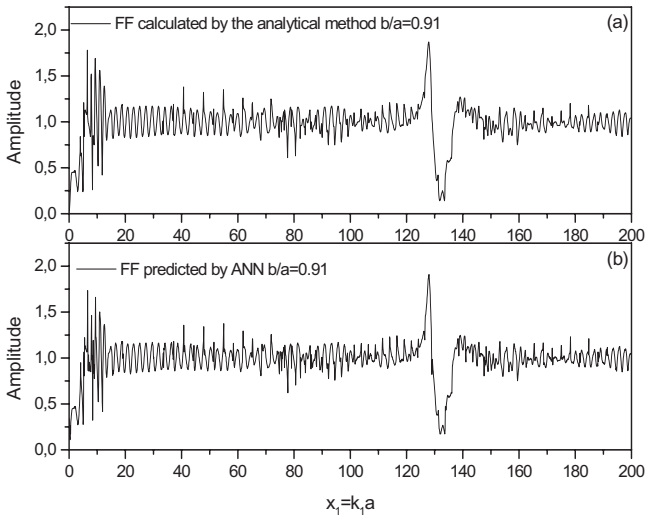


FIG. 9. Form function (a) calculated by the analytical method and (b) predicted by the ANN method (for $b/a=0.91$).

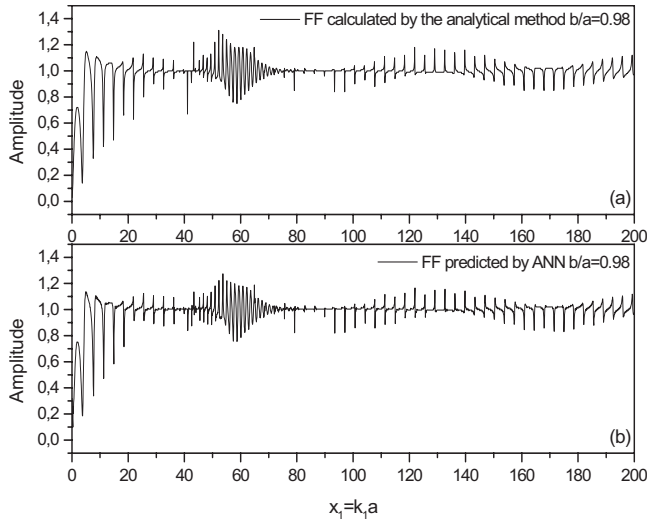


FIG. 10. Form function (a) calculated by the analytical method and (b) predicted by the ANN method (for $b/a=0.98$).

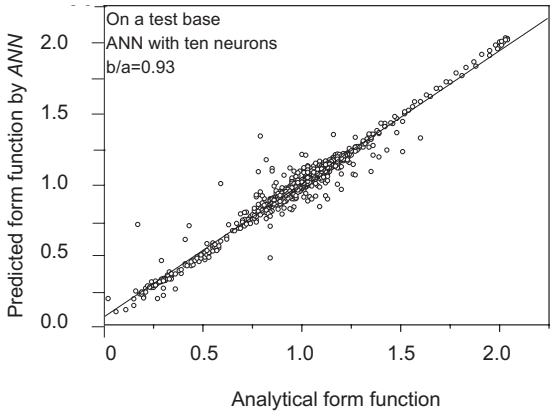


FIG. 11. Correlation between calculated and predicted values of form function on a test data set for a radius ratio of $b/a=0.93$.

VI. RESULTS AND DISCUSSION

The neural network is trained by using randomly 60% (6 FF) of the analytically computed form function data set, while the remaining data set, 40% (4 FF), is reserved as a test base. The training phase is done for several ANN architectures until the testing error values begin to increase; then the training phase is stopped. Indeed, the values of errors are measured for each neural network architecture. The evolution of the mean quadratic errors of training and testing during the training phase is represented in Fig. 6. Starting from

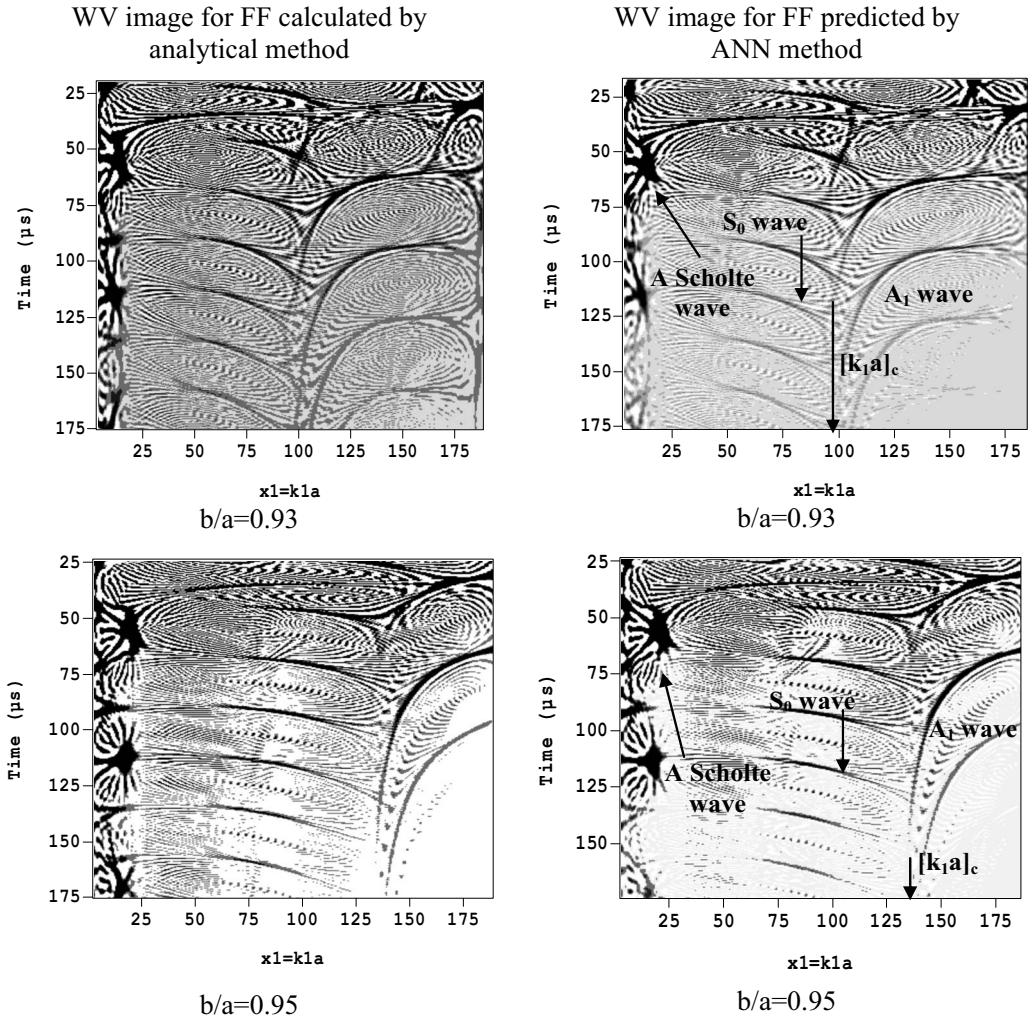


FIG. 12. Comparison between the WV images for the form function calculated by the analytical method and predicted by the ANN method ($b/a=0.93$ and 0.95).

TABLE III. Comparison between the reduced frequencies computed by the eigenmode and starting from the Wigner–Ville (WV) image.

Circumferential wave A_1	$(k_1a)_c$ computed by the eigenmode method	$(k_1a)_c$ estimated (WV) by analytical method	$(k_1a)_c$ estimated (WV) by ANN
$b/a=0.91$	74.5	74.0 ± 0.3	74.1 ± 0.3
$b/a=0.93$	95.8	95.2 ± 0.3	95.0 ± 0.3
$b/a=0.95$	134.1	134.7 ± 0.2	134.6 ± 0.2

this figure, the best configuration is found for a network with one hidden layer, composed of ten neurons. The MRE and the SE for the optimal model are, respectively, 1.61% and $(1.67\times10^{-4})k_1a$ (Table II).

Figures 7–10 show the comparison between the form functions predicted by the ANN model on a test database and that calculated by the analytical method for the radius ratios $b/a=0.91$, 0.93, 0.95, and 0.98. Figure 11 shows the good agreement between the predicted form function values by ANN and the analytically calculated form function values. The optimal configuration is found by a model of ten hidden neurons with a coefficient of correlation R about 0.941. The increase in the number of neurons in the hidden layers can especially generate sharp variations of the parameters of the network that can cause an overtraining in the case of two hidden layers.[38–40]

However, the Wigner–Ville time-frequency representation is used as a tool of comparison to check the validity of the neural network model because the studied signals are not stationary. Using a time-frequency representation can allow us to follow the evolution of the frequency content of acoustic echoes as a function of time. Figure 12 shows the good agreement between the Wigner–Ville image of the predicted form function by ANN and the analytically calculated form function for a stainless steel tube of radius ratios $b/a=0.93$ and 0.95. Moreover, starting from the Wigner–Ville image of the predicted signal, the reduced cutoff frequency ($[k_1a]_c$) of the tube corresponding to the A_1 wave can be determined.

The circumferential antisymmetric A_1 wave propagates around the circumference of the tube only for frequencies superior to the reduced cutoff frequency.[38] The reduced cutoff frequency values ($[k_1a]_c$) obtained from the Wigner–Ville image of the test database ($b/a=0.91$, 0.93, and 0.95) and analytically calculated are presented in Table III. This table also presents the values computed with the eigenmode method.[2,3,13,14,38] A good agreement is shown in this table.

With the optimal model developed in this paper, two problems can be resolved.

(1) In the direct operating mode of this optimal model, the form function of a cylindrical shell made of a specified material with a particular radius ratio b/a can be predicted.
(2) In the reverse operating mode of the optimal model, we can inject a form function for a cylindrical shell made of a specified material in the output layer to predict its radius ratio b/a.

If another material is considered, it is necessary to use supplementary inputs in the model with a new training.

VII. CONCLUSION

This paper presents the application of an ANN model to predict the form function for a stainless steel tube. The optimal model is composed of one hidden layer with ten neurons. This model is able to predict the FF with a MRE about 1.61%. In this article, the ANN approach can be used as a new tool for a nondestructive characterization of a thin elastic tube. To check the credibility of the ANN model in the time and frequency domains, Wigner–Ville time-frequency representation was used like a tool of comparison between the FF calculated by the analytical method and that predicted by the ANN technique. The reduced cutoff frequencies estimated by the ANN method are in good concordance with those computed by the eigenmode method. The comparison of the predicted results by ANN and the analytically calculated results indicates that the artificial neural network method is suitable to predict the backscattered pressure. In this paper, we have worked on stainless steel tubes, but our ANN model can also be applied to various types of materials by taking account of the material characteristics (density, phase velocities, and attenuations) as relevant inputs of the model.

[1] R. Hickling, "Analysis of echoes from a hollow metallic sphere in water," J. Acoust. Soc. Am. **36**, 1124–1137 (1964).
[2] L. Flax, L. R. Dragonette, and H. Überall, "Theory of elastic resonance excitation by sound scattering," J. Acoust. Soc. Am. **63**, 723–731 (1978).
[3] J. D. Murphy, E. D. Breitenbach, and H. Überall, "Resonance scattering of acoustic waves from cylindrical shells," J. Acoust. Soc. Am. **64**, 677–683 (1978).
[4] G. Maze and J. Ripoche, "Visualization of acoustic scattering by elastic cylinders at low ka," J. Acoust. Soc. Am. **73**, 41–43 (1983).
[5] G. Maze, J.-L. Izbicki, and J. Ripoche, "Resonances of plates and cylinders: Guided waves," J. Acoust. Soc. Am. **77**, 1352–1357 (1985).
[6] G. Maze, "Acoustic scattering from submerged cylinders, MIIR lm/Re: Experimental and theoretical study," J. Acoust. Soc. Am. **89**, 2559–2566 (1991).
[7] F. Léon, F. Lecroq, D. Décultot, and G. Maze, "Scattering of an obliquely incident acoustic wave by an infinite hollow cylindrical shell," J. Acoust. Soc. Am. **91**, 1388–1397 (1992).
[8] J. Ripoche and G. Maze, "A new acoustic spectroscopy: The resonance scattering spectroscopy by the MIIR," *Acoustic Resonance Scattering*, edited by R. H. Überall (Gordon and Breach, New York, 1992), Sec. V, pp. 69–103.
[9] D. Décultot, F. Lecroq, G. Maze, and J. Ripoche, "Acoustic scattering from a cylindrical shell bounded by hemispherical endcaps: Resonance explanation with surface waves propagating in cylindrical and spherical shells," J. Acoust. Soc. Am. **94**, 2916–2923 (1993).
[10] N. Veksler, G. Maze, J. Ripoche, and V. Porochovskii, "Scattering of obliquely incident plane acoustic wave by circular cylindrical shell: Results of computations," Acta Acust. **82**, 689–697 (1996); G. Maze, F. Léon, and N. Veksler, "Scattering of an obliquely incident plane acoustic wave by circular cylindrical shell: Experimental results," *ibid.* **84**, 1–11 (1997).
[11] S. F. Morse, P. L. Marston, and G. Kaduchak, "High-frequency back-

scattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation and calculations," J. Acoust. Soc. Am. **103**, 785–794 (1998).
[12]L. Haumesser, D. Décultot, F. Léon, and G. Maze, "Acoustic scattering from a finite cylindrical shell at oblique incidence: Experimental identification along the shell length," J. Acoust. Soc. Am. **111**, 2034–2039 (2002).
[13]M. Talmant, G. Quentin, J. L. Rousselot, J. V. Subrahmanyam, and H. Überall, "Acoustic resonances of thin cylindrical shells and the resonance scattering theory," J. Acoust. Soc. Am. **84**, 681–688 (1988).
[14]G. V. Frisk, J. W. Dickey, and H. Überall, "Surface wave modes on elastic cylinders," J. Acoust. Soc. Am. **58**, 996–1008 (1975).
[15]J. L. Izbicki, G. Maze, and J. Ripoche, "Influence of the free modes of vibration on the acoustic scattering of a circular cylindrical shell," J. Acoust. Soc. Am. **80**, 1215–1219 (1986).
[16]R. Latif, E. H. Aassif, G. Maze, D. Décultot, and A. Moudden, "Determination of group and phase velocities from time-frequency representation of Wigner-Ville," NDT & E Int. **32**, 415–422 (1999); R. Latif, E. H. Aassif, A. Moudden, D. Décultot, B. Faiz, and G. Maze, "Determination of the cutoff frequency of an acoustic circumferential wave using a time-frequency analysis," *ibid.* **33**, 373–376 (2000).
[17]N. Yen, L. Dragonette, and S. K. Numrich, "Time-frequency analysis of acoustic scattering from elastic objects," J. Acoust. Soc. Am. **87**, 2359–2370 (1990).
[18]T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner Distribution - A tool for time-frequency signal analysis," part 2: Discrete-Time Signals, Philips J. Res. **35**, 276–300 (1980).
[19]W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull. Math. Biophys. **5**, 115–133 (1943).
[20]F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton," Technical report No. 85-460-1, Project PARA, Cornell Aeronautical Laboratory, New York, 1957.
[21]M. L. Minsky and S. A. Papert, *Perceptrons* (MIT, Cambridge, MA, 1990).
[22]G. Maze, F. Léon, J. Ripoche, and H. Überall, "Repulsion phenomena in the phase-velocity dispersion curves of circumferential waves on elastic cylindrical shells," J. Acoust. Soc. Am. **105**, 1695–1701 (1999); G. Maze, N. Touraine, A. Baillard, D. Décultot, V. Latard, L. Derbesse, P. Pernod, and A. Merlen, "A_0-wave and A-wave in cylindrical shell immersed in water: Influence on the acoustic scattering," 1999 ASME Design Engineering Technical Conferences, DECTC99/VIB-8090, Las Vegas, NV, USA (12–15 September, 1999), pp. 1–11, on CD-Rom.
[23]L. Flax and W. G. Neubauer, "Acoustic reflection from layered elastic absorptive cylinders," J. Acoust. Soc. Am. **61**, 307–312 (1977).
[24]N. Veksler, *Resonance Acoustic Spectroscopy*, Springer Series on Wave Phenomena (Springer-Verlag, Berlin, 1993).
[25]P. Pareige, P. Rembert, J.-L. Izbicki, G. Maze, and J. Ripoche, "Méthode impulsionnelle numérisée (MIN) pour l'isolement et l'identification des résonances de tubes immergés, (digital impulse method for isolation and identification of resonances of immersed tubes)," Phys. Lett. A **135**, 143–146 (1989).
[26]N. H. Sun and P. L. Marston, "Ray synthesis of leaky Lamb wave contributions to backscattering from thick cylindrical shells," J. Acoust. Soc. Am. **91**, 1398–1402 (1992).
[27]D. H. Hughes and P. L. Marston, "Local temporal variance of Wigner's distribution function as a spectroscopic observable: Lamb wave resonances of a spherical shell," J. Acoust. Soc. Am. **94**, 499–505 (1993).
[28]K. Levenberg, "A method for the solution of certain non-linear problems in least squares," Q. Appl. Math. **2**, 164–168 (1944).
[29]D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," J. Soc. Ind. Appl. Math. **11**, 431–441 (1963).
[30]D. Plaut, S. Nowlan, and G. E. Hinton, "Experiments on learning by back propagation," Technical report No. CMU-CS-86-126, Department of Computer Science, Carnegie-Mellon University, 1986.
[31]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error back propagation," *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT, Cambridge, MA, 1986), pp. 318–362.
[32]S. K. Singh, K. Srinivasan, and D. Chakraborty, "Acoustic characterization and prediction of surface roughness," J. Mater. Process. Technol. **152**, 127–130 (2004).
[33]S. Riad, J. Mania, L. Bouchaou, and Y. Najjar, "Rainfall-runoff model using an artificial neural network approach," Math. Comput. Modell. **40**, 839–846 (2004).
[34]S. Riad, J. Mania, L. Bouchaou, and Y. Najjar, "Predicting catchment flow in a semi-arid region via an artificial neural network technique," Hydrolog. Process. **18**, 2387–2393 (2004).
[35]P. J. Brokwell and R. A. Davis, *Times Series Theory and Methods*, 2nd ed. (Springer-Verlag, New York, 1991).
[36]G. E. P. Box and G. Jenkins, *Times Series Analysis, Forecasting and Control* (Holden-Day, San Fransisco, 1970).
[37]S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," Neural Comput. **4**, 1–58 (1992).
[38]A. Dariouchy, E. H. Aassif, D. Decultot, and G. Maze, "Acoustic characterization and prediction of the cut-off dimensionless frequency of an elastic tube by neural networks," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1055–1064 (2007).
[39]K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks **2**, 359–366 (1989).
[40]G. Cybenko, "Approximation by superposition of a sigmoidal function," Math. Control, Signals, Syst. **2**, 303–314 (1989).

Removing additive noise via neuro-fuzzy-based reinforcement learning

Ching-Shun Lin
Department of Electronic Engineering, National Taiwan University of Science and Technology, 43, Section 4, Keelung Road, Taipei 106, Taiwan

Chris Kyriakakis
Immersive Audio Laboratory, USC Ming-Shieh Department of Electrical Engineering, EEB432, 3740 McClintock Avenue, Los Angeles, California 90089

(Received 5 September 2007; revised 9 March 2008; accepted 28 May 2008)

In this paper, a systematic treatment for developing a noise removal system based on the fundamental principle of reinforcement learning and fuzzy cerebellar model articulation controller (FCMAC) is presented. The proposed system improves its performance over time through two mechanisms. First, the modified stochastic real-valued algorithm, learning from its own mistakes via the reinforcement signal and reinforcing its action to improve future performance, is used for searching the optimal noise spectrum for the overall training system. Second, system states associated with the positive reinforcement are memorized by FCMAC-based neurons, where, in the future, similar states will share the experiences already stored there and then lead the action to a more positive situation. In this work, FCMAC's intrinsically poor approximation of rapidly varying functions is solved by taking the complex semicepstrum. In addition, the FCMAC provides an improvement in accuracy of function approximation without losing the property of generalization, which makes the high fidelity digital signal processing possible. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945794]

PACS number(s): 43.60.Np, 43.60.Lq, 43.60.Ac, 43.60.Cg [DOS] Pages: 1026–1037

I. INTRODUCTION

In daily communication, the background noise causes a signal degradation, which could lead to distortion of audio and unintelligibility of speech. The purpose of this topic is to develop a novel method for removing uncorrelated, stationary, additive noise based on reinforcement learning. Conventional approaches use adaptive filters or stochastic processes to reduce noise with given frequencies,[1,2] whereas ignoring the fact that different audio/speech signals in a specific venue could be colored by noise with uncertain characteristics. Wiener filtering, for example, is a commonly used approach for adaptively decomposing the noisy speech cepstrum into estimates of clean speech ceptrum and noise cepstrum.[3] Requiring *a priori* knowledge of the noise characteristics is the major drawback of such methods. Owing to system noise and environmental disturbance, these reference models are not always available, and thus are not suitable for applications in practice. Moreover, analytical models tend to become complicated and nonlinear so that they cannot be solved by conventional methods. Although several noise-removing approaches, such as spectral subtraction, can substantially reduce the noise buried in speech, they also introduce signal distortion.[4] For information transfer purposes many features of original sound can be omitted or degraded as long as the important information, such as intelligible speech, is preserved. However, the general goal for audio issue is to reproduce all aspects that can improve sound quality and contribute influential experience, and therefore those methods cannot provide a listener with an experience similar enough to the original sound event and environment.[5]

An alternative way to tackle this problem is to use neural networks to remove the unwanted components; however, in most cases the exact desired signals are unavailable for training. A major characteristic of our proposed system is the capability of using incomplete knowledge of the recorded signals to predict their inputs. To achieve this, we adopt a preprocessor and a quantizer based on complex semicepstrum and fuzzy cerebellar model articulation controller (FCMAC), respectively.[6] Because the original FCMAC, a supervised learning algorithm, is not applicable to an ill-defined objective, the adaptive heuristic critic (AHC) and the stochastic real valued (SRV) are also applied to meet this requirement.[7–9] In other words, the synaptic weights of FCMAC are updated by a stochastic reinforcement signal that represents the difference between the actual and the predictive failures. From the standpoint of the modified SRV, FCMAC is employed as a signal quantizer. On the other hand, FCMAC uses the modified SRV as its weight-update rule. Therefore, this self-organizing FCMAC can cancel noise resulting from a given electroacoustic system including both the electrical and the acoustical domains.

One of the spectral-domain algorithms proposed by Ephraim and Malah,[10] log spectral amplitude estimator (LSAE), is used to compare with our algorithm owing to their similar preprocessors and segmented approaches.[10,11] In the LSAE approach, signal is segmented into overlapping blocks with a certain overlapping rate, and then a Hanning window is applied to the signal blocks, which are respec-

 0001-4966/2008/124(2)/1026/12/$23.00

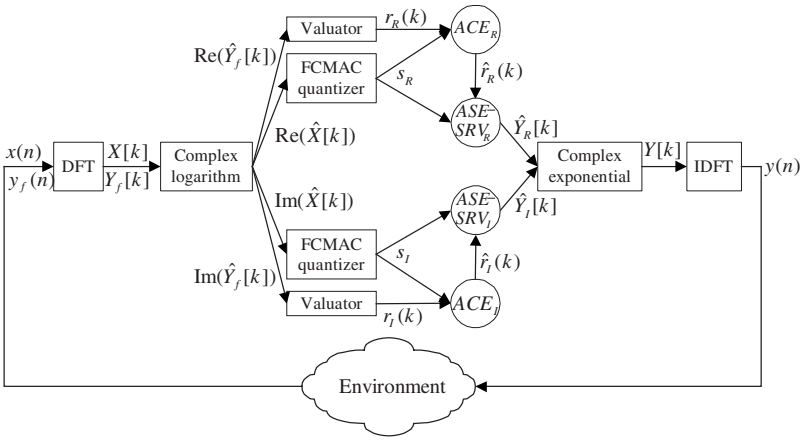


FIG. 1. Block diagram of the noise removal system.

tively similar to the ways that FCMAC-based quantizer uses overlapped receptive fields, and then adopts membership functions to extract features. More specifically, a modified estimator proposed by Cohen[12] for the *a priori* signal-to-noise ratio (SNR) and the *a priori* speech absence probability that turns out to be the speech presence probability estimated for each frequency bin and each frame by a soft-decision approach, which is similar to our approach that uses a FCMAC-based quantizer to judge whether a state variable belongs to a certain spectral group. In addition, both mechanisms process signals in the logarithmic spectrum (or semiquefrency) domain, and remove noises based on the stochastic models.

The rest of this paper is organized into four sections. In Sec. II, we first introduce the related algorithms, which are immediately followed by the proposed approaches in each subsection. The experimental results and their corresponding discussions are illustrated in Sec. III. Finally, conclusions and future research are shown in Sec. IV.

II. ARCHITECTURE AND SYSTEM IMPLEMENTATION

The noise removal system is composed of three modules, of which the first is a preprocessor using the complex semicepstrum to extract the representative features from input $x(n)$ and fedback output $y_f(n)$, the second is FCMAC-based reinforcement learning used for identifying similar features as well as removing noises, and finally, the inverse complex semicepstrum is taken to synthesize $\hat{Y}_R[k]$ and $\hat{Y}_I[k]$ into output signal $y(n)$. The architecture of this noise removal system is shown in Fig. 1, and each module is described as follows.

A. Feature extraction

1. Complex cepstrum

The cepstral processing, originally developed by Bogert *et al.*[13] has been proven useful in various fields such as homomorphic filtering, seismic analysis, and audio signal processing.[13–15] Consider a stable sequence $x(n)$ whose discrete Fourier transform (DFT) is defined as

$$X[k] = X(e^{j\omega})|_{\omega=(2\pi/N)k} = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn}. \quad (1)$$

The finite complex cepstrum corresponding to $x(n)$ is defined as the sequence $\hat{x}(n)$ and can be represented by using the inverse discrete Fourier transform (IDFT):

$$\hat{x}_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}[k]e^{j(2\pi/N)kn}, \quad (2)$$

where $\hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN)$ is the time-aliased version of $\hat{x}(n)$, and

$$\hat{X}[k] = \log(X(e^{j\omega}))|_{\omega=(2\pi/N)k} = \log(X[k]) \quad (3)$$

is its DFT, where $\log(\cdot)$ denotes the complex logarithm.

2. Complex semicepstrum

Although the cepstrum transformation of a real sequence is real, the magnitude-based real cepstrum is not applied here because signals cannot be recovered without the phase information. In addition, the imaginary part of the output of FCMAC is an approximate estimate rather than an actual value; it causes the imaginary part of IDFT to not necessarily be zero. The imaginary part of DFT cannot; therefore, be omitted even though the input is a real sequence owing to these reasons. Also, if real signals are transformed into the quefrency domain, the output will be a real signal as usual. This may cause some information to be lost when the IDFT is taken, so only parts of complex cepstrum procedure and its inverse are adopted in this application. The shadow parts of Fig. 2 illustrate how the complex semicepstrum and its inverse are implemented.

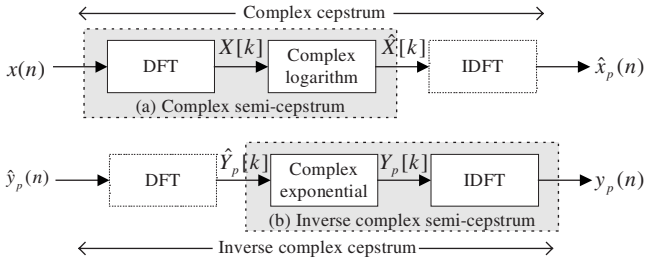


FIG. 2. (a) Realization of the complex semicepstrum and (b) its inverse.

When the time domain signal sequences as well as the fedback outputs are sent into the system, the neural network will operate like a surface-fitting system. The problem is that we do not have any idea how long it will take to make each incoming training sequence to have sufficient patterns captured for training. If the signal sequences are transferred to the semiquefrency domain and the transformed signals are used as the features, these problems can be circumvented. Moreover, as reinforcement learning requires that the excited neurons obtain sufficient credits via repetitious processes, processing the audio signals in the semiquefrency domain instead of the time domain will be more suitable for our application. The trade-off is that on-line training is no longer possible because the patterns must be processed at the same time, i.e., our approach waits for each completely incoming sequence and then sends it into FCMAC for pattern quantization.

B. Pattern quantization

1. Fuzzy cerebellar model articulation controller

The cerebellar model articulation controller (CMAC) is a linear combination of overlapped basis functions. It is a type of local neural network, i.e., only a local set of neurons is activated by a particular input. Due to the properties of local generalization, rapid computation, function approximation, and output superposition, it has been widely used in robot control, pattern recognition, and signal processing.[16] This neural network can be integrated with fuzzy logic theory.[17] However, unlike a conventional fuzzy-rule-based system, a neuro-fuzzy system does not need to extract inferential rules. Learning is based on observations of the input/output relationship of the system. Both fuzzy and CMAC systems can perform function approximation in an interpolation look-up table manner with the principles of dichotomy and generalization. Further, the nonlinear mapping of CMAC can be regarded as the subset of the aggregation operation on fuzzy sets. The identification of these similarities between CMAC and fuzzy logic has led to the theory of the fuzzy cerebellar model articulation controller.[6,18]

2. Refinement of FCMAC

The architecture of FCMAC consists of two processing stages. First a membership function, $\mu_i(\hat{X}) \rightarrow [0,1]$, associates each state variable $\hat{X}$ with a number between 0 and 1. This function $\mu_i(\hat{X})$ represents the degree of belonging of $\hat{X}$ to a certain class, and μ_i denotes a sparse vector in which up to the number of C (generalization parameter), the ratio of generalization width to quantization width, of its components are excited. In this way, inputs close to each other within the receptive field produce similar outputs. These membership functions have a peak value at the center of the excited region, and the output decreases as the input moves toward the edge of the excited region. Figure 3 shows the organization of the overlapped receptive fields in the input space with the generalization parameter $C=3$. Notice that different grades of membership will be assigned to their corresponding cells μ_i if one of the quantization regions is excited by a given state variable, which will be replaced with the complex semicepstrum signal $\hat{X}[k]$ later.

The mapping is realized by feeding the membership vector μ_i on the sensor layer into the T-norm units. By taking the algebraic product, for instance, the diagonal ν_i of a two-dimensional system can be formulated as

$$\nu_i = \mu_{ai}\mu_{bi}^T, \quad i = 1, \ldots, C, \tag{4}$$

where T denotes the matrix transpose, ν_i is the excited subset on the hyperspace, and μ_{ai} and μ_{bi} are the membership vectors in the dimension a and dimension b, respectively. The off-diagonal ν_i is defined as

$$\nu_i = \begin{cases} \mu_{aj}\mu_{b(C+1-j)}^T, & \text{for } i = C+1, \ldots, C+\lfloor C/2 \rfloor; \\ & j = 1, \ldots, \lfloor C/2 \rfloor \\ \nu_j^T, & \text{for } i = C+\lfloor C/2 \rfloor + 1, \ldots, 2C-1; \\ & j = C+1, \ldots, C+\lfloor C/2 \rfloor \end{cases} \tag{5}$$

where $\lfloor \cdot \rfloor$ denotes the floor operation. These equations shape the excited subsets ν_i to be uniformly arranged over the hyperspace.[19,20] Then each T-norm is sparsely interconnected to the next logic T-conorm unit by the algebraic sum:

$$\begin{aligned} &\bar{g}_0(\hat{X}_a, \hat{X}_b) = \bar{0} \\ &\bar{g}_i(\hat{X}_a, \hat{X}_b) = \bar{g}_{i-1}(\hat{X}_a, \hat{X}_b) + \nu_i + \nu_{i+1} - \nu_i \cdot \nu_{i+1}, \\ &for\ i = 1, \ldots, 2C-2 \end{aligned} \tag{6}$$

where $(\cdot)$ denotes the element-by-element multiplication, and $\bar{g}(\hat{X}_a, \hat{X}_b) \leftarrow \bar{g}_{2C-2}(\hat{X}_a, \hat{X}_b)$ is a $(2C-1)\times(2C-1)$ nonlinear mapping matrix. Figure 4 depicts a schematic diagram of the two-dimensional FCMAC operations that each subset is relatively offset to the others in the hyperspace. The mapping of FCMAC is implemented by replacing the logic *AND* and *OR* operators in CMAC with the commonly used T-norm and T-conorm, respectively. Because the algebraic product of T-norm and algebraic sum of T-conorm produce smoother outputs than the other approaches, they are used in this work to make system analysis possible. Figure 5 illustrates a portion of the nonlinear mapping result $\bar{g}(\hat{X}_a, \hat{X}_b)$ for a couple of inputs. The FCMAC weighted elements are, in general, initially empty and the memory elements have a null response to any input until they have been trained.

C. Reinforcement learning

1. Conventional RL algorithms

As the weight-update rule of FCMAC needs the desired output, its supervised learning rule is not suitable for a reinforcement approach. Therefore, a self-organizing FCMAC implementing a reinforcement-learning algorithm is developed to deal with this problem. The SRV algorithm is an alternative form of self-organizing learning algorithm. It randomly selects actions according to the probability distribution and receives feedback from the environment for evaluation, and then the probability distribution is adjusted based

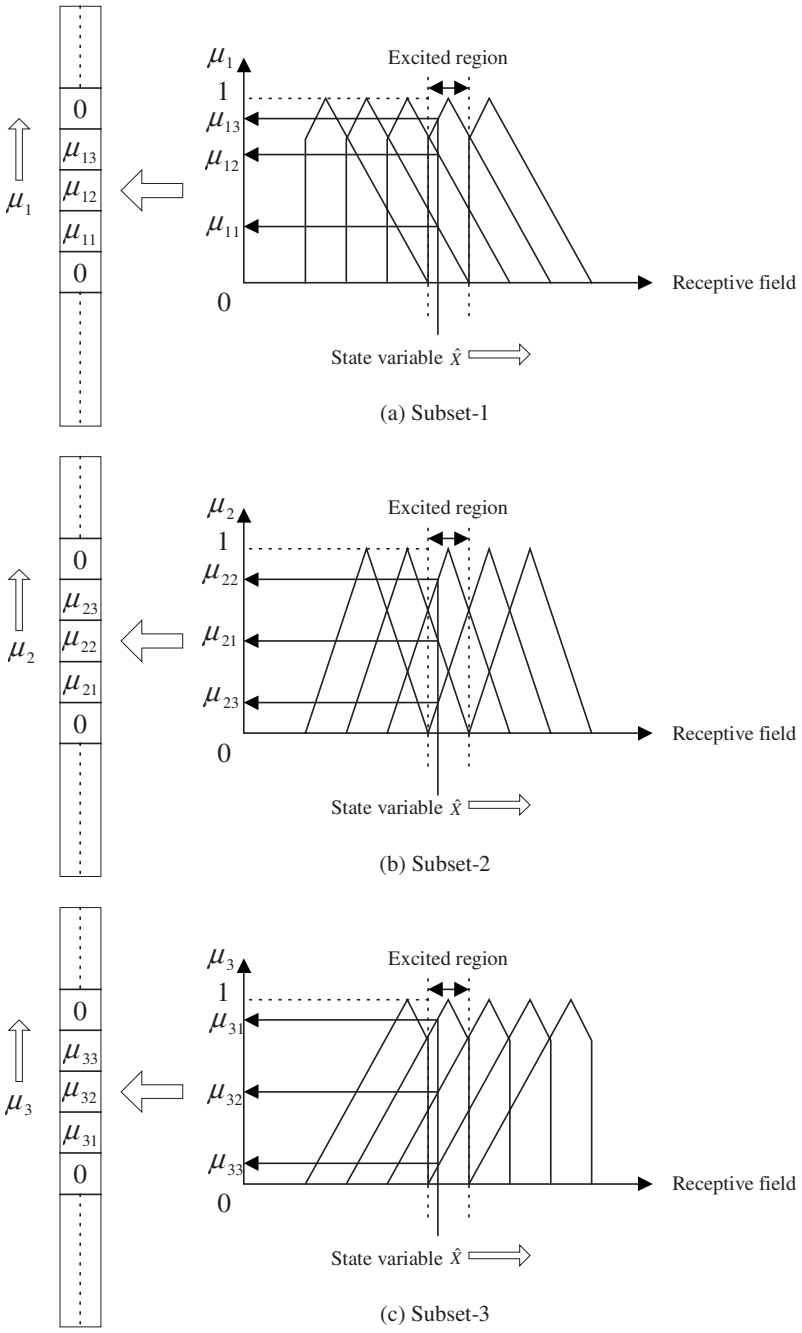


FIG. 3. Membership function with generalization parameter $C=3$ for different subsets.

on this feedback. The original SRV follows BOXES scheme in quantizing state space and constructs a neural network controller to achieve the peg-in-hole insertion.[9] In this work, we apply this FCMAC-SRV neural network to a signal-processing problem.

As shown in Fig. 1, after taking the discrete Fourier transform and the complex logarithm, we use the membership functions of FCMAC to quantize the vector $\hat{\mathbf{X}}$. As depicted in Fig. 3, different grades of membership will be assigned to their corresponding cells μ_i if one of the quantization regions is excited by a given state variable $\hat{X}[k]$. In addition, the frequency index k is used as the other feature for increasing the precision of output, which is adjusted by the following equation:

$$\hat{Y}_k(t+1) = \hat{X}_k(t) + \Delta\hat{X}_k(t), \tag{7}$$

where $\Delta\hat{X}_k(t)$ is the signal-adjusted value resulting from a Gaussian process. The actor–critic configuration proposed by Barto *et al.*[7] is also adopted in this experiment. This reinforcement learning approach can be divided into two tasks, including action function and evaluation function. The action function is implemented by integrating the associative search element (ASE) and SRV, whereas the evaluation function is achieved by an adaptive critic element (ACE). The SRV estimates the mean $m(t)$ and the standard deviation $\sigma(t)$ of the normal distribution $N(m(t),\sigma(t))$ to generate $\Delta\hat{X}_k(t)$, i.e.:

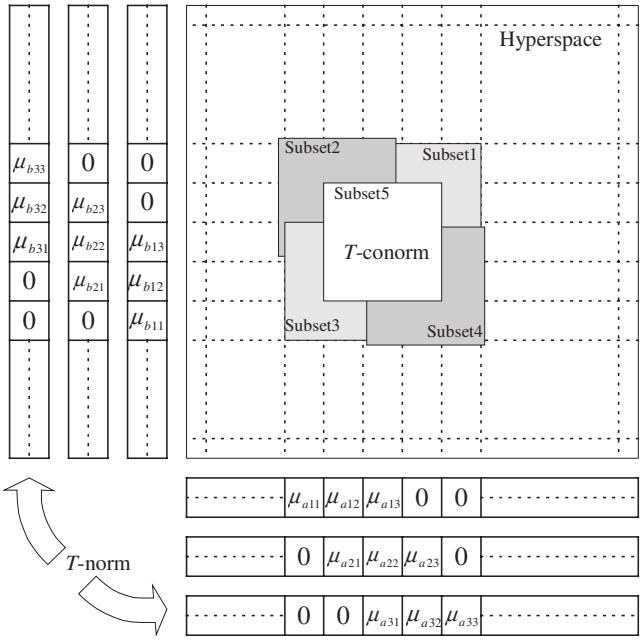


FIG. 4. Nonlinear mapping of the 2D FCMAC.

$$\Delta\hat{X}_k(t) \sim N(m(t), \sigma(t)) \tag{8}$$

where $m(t)$ is an estimate of the optimal output. It is computed as a weighted sum of the excited inputs of the unit:

$$m(t) = \sum_{j=1}^{2C-1} \left(\sum_{i=1}^{2C-1} w_{ij}(t) s_{ij}(t) \right) \tag{9}$$

where $s_{ij}(t)$ the element of FCMAC nonlinear mapping matrix assigned by $\bar{g}_{ij}$, and $w_{ij}(t)$ is the weight of ASE-SRV module. On the other hand, the standard deviation $\sigma(t)$ depends on how close the currently expected output is to the optimal output for a given input. This closeness is indirectly measured by an external reinforcement signal $r(t)$, which is generally assigned to 1 for reward or 0 for punishment. Because $r(t)$ is a performance index, $\sigma(t)$ should depend on it through the internal reinforcement signal. More specifically, the higher the expected reinforcement signal is, the better the unit is performing. In this case, $\sigma(t)$ should be shrunk to

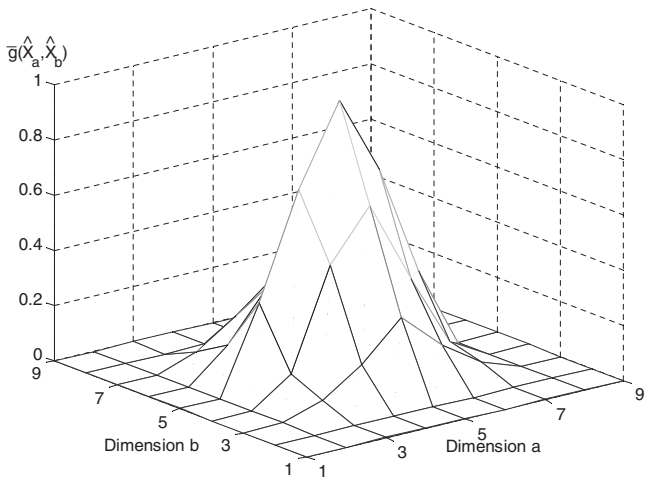


FIG. 5. Nonlinear mapping result of the 2D FCMAC.

narrow the searching scope.[21] Therefore, $\sigma(t)$ is designed as

$$\sigma(t) = 1 - \hat{r}(t), \tag{10}$$

where the internal reinforcement $\hat{r}(t)$ is calculated as follows:

$$\hat{r}(t) = \frac{1}{1 + e^{-p(t)}}, \tag{11}$$

where $p(t)$, the output of ACE evaluation unit, is the summation of the element-by-element products of the unit's weighting matrix and FCMAC mapping matrix:

$$p(t) = \sum_{j=1}^{2C-1} \left(\sum_{i=1}^{2C-1} v_{ij}(t) s_{ij}(t) \right). \tag{12}$$

The ACE not only provides an estimate of problem state, but also functions as a prediction of failure. The following equation illustrates the update rule of ACE:[9]

$$v_{ij}(t+1) = v_{ij}(t) + \beta(r(t) - \hat{r}(t)) \frac{s_{ij}(t)}{M}, \tag{13}$$

where β is the positive learning rate and M is the number of excited states to average the effects when different amount of $s_{ij}(t)$ are involved. On the other hand, the update rule of ASE-SRV is expressed as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(r(t) - \hat{r}(t)) \left(\frac{\Delta\hat{X}_k(t) - m(t)}{\sigma(t)} \right) \frac{s_{ij}(t)}{M}, \tag{14}$$

where α is the positive learning rate. The action adjustment is based on the distribution of training data, whereas the evaluation adjustment is decided by how many credits the previous action can obtain from the critic element.

One of the objectives of this work is to incorporate FCMAC with SRV algorithm to reduce noise buried in signal. The SRV predicts the degree of the event failure and resolves the compensating signal while the quantizing signals are fed in. A major characteristic of the proposed system is the capability of using the experience with only incomplete knowledge about the environment to predict its behavior. However, to overcome the inherent drawback of lengthy convergence of SRV, we integrate AHC algorithm into the origin system. The modified schemes are listed in Sec. II C 2.

2. Unsupervised prediction of modified SRV

A single action during the noise-removing attempt can be rewarded or punished by comparing the estimates of the state preceding the action with those following the action. The capability allows the learning to occur not merely on the trial but also upon failure. Unlike the single-step prediction or the supervised learning method, either of which assigns credit to the difference between the predicted output and actual output, the temporal difference (TD) methods assign credit according to the difference between temporally successive predictions.[22,23] The goal of the critic is to predict the following quantity:

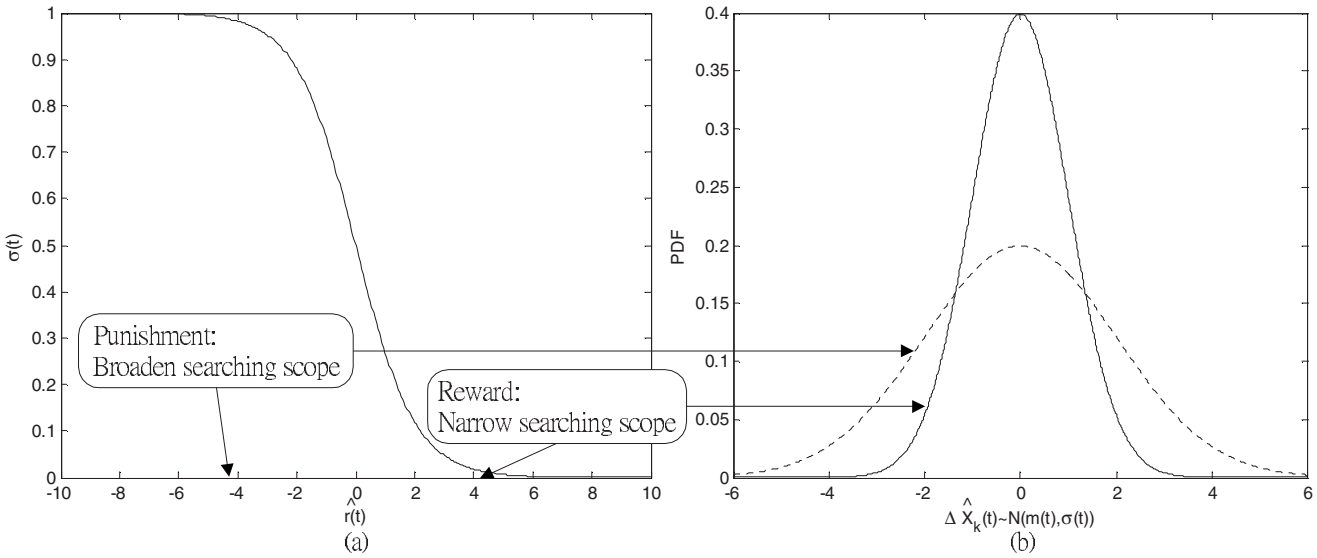


FIG. 6. Illustration of the adaptive searching scope in the modified SRV approach.

$$\hat{p}(t) = \sum_{l=0}^{\infty} \gamma^l r(t+l+1) \tag{15}$$

where $r(t+l+1)$ is the external reinforcement signal at time $t+l+1$, and $\gamma \in [0,1)$ is a discount factor determining the level of influence of future reinforcements on the selection of the current action. In other words, $\hat{p}(t)$ is the prediction of infinite discounted cumulative outcomes that ACE tries to approach. Assume the reinforcement predictions are correct; then $\hat{p}(t)$ can be rewritten as

$$\hat{p}(t) = r(t+1) + \gamma \sum_{l=0}^{\infty} \gamma^l r(t+l+2) = r(t+1) + \gamma \hat{p}(t+1). \tag{16}$$

In most studies on TD procedures, AHC's weights are updated after either the presentation of new signals or a complete training sequence. Nevertheless, if the changes of the signal $s_{ij}(t)$ and the weight of ACE $v_{ij}(t)$ induce an alteration to the internal reinforcement signal $\hat{r}(t)$ within a sequence, it probably leads to instability in the training process. Hence, the internal reinforcement signal is an appropriation to the temporal difference error[22]:

$$\hat{r}(t+1) = r(t+1) + \gamma p(s_{ij}(t+1), v_{ij}(t)) - p(s_{ij}(t), v_{ij}(t)). \tag{17}$$

This modification ensures that the changes in prediction due to $s_{ij}(t)$ and $s_{ij}(t+1)$ are effective in causing weight alteration.

3. Weight update of modified SRV

Updating the weights of ASE-SRV and ACE used for computing the expected evaluation is relatively straightforward. As both the mapping matrix and the internal reinforcement signal are supplied to the unit, we use the least mean square rule to learn these associations and the rules are now transformed into:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \hat{r}(t) \left(\frac{\Delta \hat{X}_k(t) - m(t)}{\sigma(t)} \right) \frac{s_{ij}(t)}{M} \tag{18}$$

and

$$v_{ij}(t+1) = v_{ij}(t) + \beta \hat{r}(t) \frac{s_{ij}(t)}{M}, \tag{19}$$

respectively. These functions are learnt via TD method, which adjusts the weights of the modified SRV network in proportion to the difference between reinforcement predictions on the consecutive steps.

4. Adaptive searching of modified SRV

With external reinforcement signal $r(t)$ assigned to be 1 or 0, the standard deviation for the normal distribution activation of SRV is monotonically decreased with increasing $\hat{r}(t)$ by using a nonnegative function:

$$\sigma(t) = \frac{e^{-\hat{r}(t)}}{1 + e^{-\hat{r}(t)}}. \tag{20}$$

If $\hat{r}(t)$ approaches a right prediction, it will cause the standard deviation $\sigma(t)$ to reduce, and thereby to restrict the search to a smaller neighborhood of the current mean $m(t)$. Conversely, in order to find a better solution, this monotonically decreasing function will broaden the searching scope if a punishment was issued in the previous step. Introducing this stochastic searching mechanism makes the network have the chance to jump out of the trapped local minimum.[24] Figure 6(a) shows how this function is used for adaptively adjusting the searching scope and increasing the convergence rate, whereas in Fig. 6(b) the width of the confidence interval reflects the uncertainty of the algorithm's knowledge and guides further exploration.

5. Fractional grading of modified SRV

Unlike the conventional reinforcement learning in which the valuation is a dichotomous value, the critic signal here is

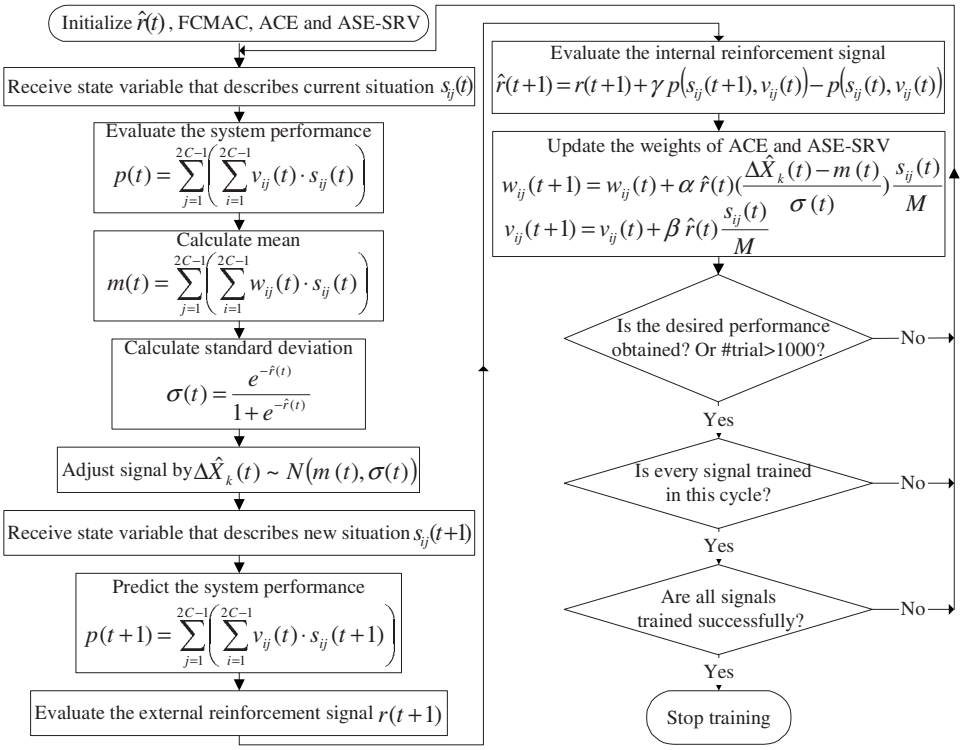


FIG. 7. Flowchart of FCMAC-based reinforcement learning.

further divided into more specific grades, i.e., a fractional expression between 0 and 1. The mean square error (MSE) $E\{|\mathrm{Re}(\hat{Y}_f[k])-\mathrm{Re}(\hat{Y}[k])|^2\}$, instead of any desired response, is also involved to construct a valuator module. Therefore the critic for each signal can be described as the following piecewise function:

$$r(t)=\begin{cases}1 & \text{if } R(t)<\rho\\ \dfrac{1-R(t)}{1-\rho} & \text{if } \rho\le R(t)\le 1\\ 0 & \text{if } R(t)>1\end{cases} \tag{21}$$

where $R(t)$ is the ratio of the current MSE to the previous MSE, and ρ is an input-dependent parameter between 0 and 1. Failure occurs when the MSE exceeds its previous value, and the critic signal for the imaginary part is also formulated similarly. The objective is to reduce the unwanted components in the semiquefrency domain by punishing the incorrect prediction, whereas change nothing on the proper signals by giving it a reward.[25] Finally, the flowchart of FCMAC-based reinforcement learning is illustrated in Fig. 7.

TABLE I. Parameters for the modified SRV system.

Generalization parameter C	3
Quantized resolution for state variable k	1.0
Quantized resolution for state variable $\hat{X}[k]$	0.042
Training tolerance	0.01
AES-SRV learning rate α	0.5
ACE learning rate β	0.5
Discount factor γ	0.3
Critical threshold ρ	0.9965

III. EXPERIMENTAL RESULTS

Additive noise is always present in the environment, and it contributes both to phase and amplitude fluctuations. Its contribution comes from the filtering of the phase component and the conversion of the amplitude component. Let $z(n)$, $n(n)$, and $x(n)$ be the clean signal, the uncorrelated additive background noise, and the contaminated signal, respectively. They can be represented in the temporal domain as $x(n)=z(n)+n(n)$. In the first experiment, we arbitrarily choose audio signals colored by high-frequency noise with 2.76 and 3.68 kHz sinusoidal waves. In theory, the frequency and the number of noises are unrestricted as long as they are not too close to each other in frequency. These narrow-band noise have the same frequency characteristics in the semiquefrency domain as they do in the frequency domain. In addition, the training pattern $\hat{X}[k]$ and its fedback output $\hat{Y}_f[k]$ are assigned as the outputs of the complex logarithm module. Because the close resemblance between $\hat{X}[k]$ and $\hat{Y}_f[k]$ can be observed by analyzing their correlation functions, $\hat{X}[k]$ and its corresponding index k are sent to FCMAC for quantization. In other words, we initialize the nonlinear mapping matrix by the procedure $s_{ij}(0)\leftarrow\bar{g}_{ij}(k,\hat{X}[k])$. On the other hand, $\hat{Y}_f[k]$ is only sent to valuator modules for comparison. As the weight training-based SRV uses Gaussian distribution to select the adjusted value, we have to verify whether a directly

TABLE II. Numbers of average attempts before success.

Part	Random selection	SRV	Modified SRV
Real	506.69	450.59	358.91
Imaginary	724.11	544.16	445.52

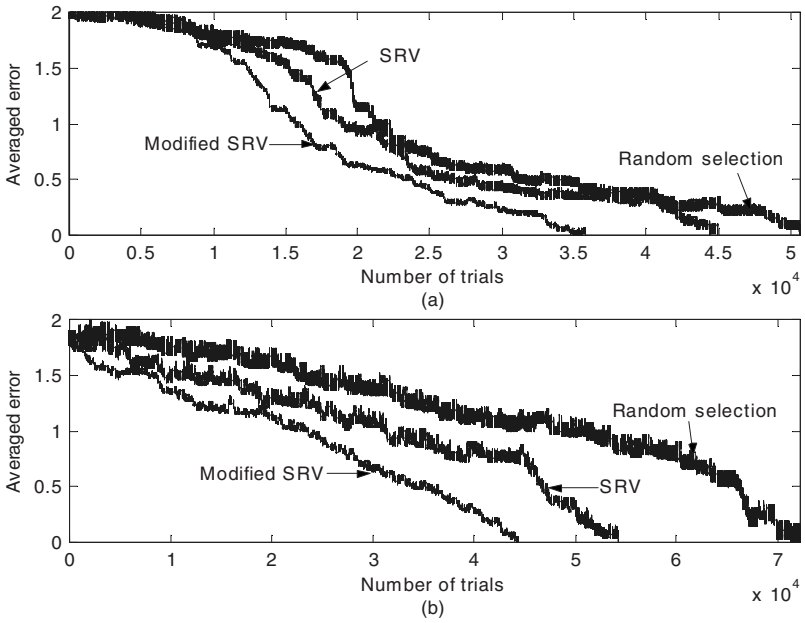


FIG. 8. Comparison of convergence rate for different approaches. (a) Real part and (b) imaginary part.

random selection from the normal distribution is more efficient owing to its low memory requirement and light computing necessity. In other words, to guarantee that our algorithms work systematically instead of just guessing a better output, a convergence comparison is made based on the same experimental parameters (see Table I for details). As shown in Fig. 8, the training error is reduced after a few trials, and the performances of both SRV and modified SRV are better than those randomly selected from the normal distribution function $N(0,1)$. On average, the modified SRV has a shorter convergence time than other approaches. The numbers of average attempts before success for different approaches are tabulated in Table II. The modified SRV approach, in general, needs the fewest tries to successfully train a signal. Figure 9(a) illustrates that most signals are punished in the first cycle because they do not meet the requirements, whereas Fig. 9(b) indicates the averaged error decreases dramatically at the same time. The modified SRV takes 3 training cycles with 1000 trials per cycle to achieve the goal in the real-part signal training; it means all trained signals can be located within the training tolerance before the 1000th trial, and the system does not receive punitive signals anymore. Figure 10 shows the overall training for the imaginary part is finished within 2 cycles, because every attempt at each signal is successful before hitting the boundary of 1000. The input and its corresponding fedback output near the high-frequency component are indicated in Fig. 11. To make Fig. 11 clear, only 100 signals are shown in this simulation. Their error after training is much smaller than that of the original signal scale. The similar results for the imaginary part can also be found in Fig. 12. Figure 13(a) shows the overall comparison between colored and original signals by

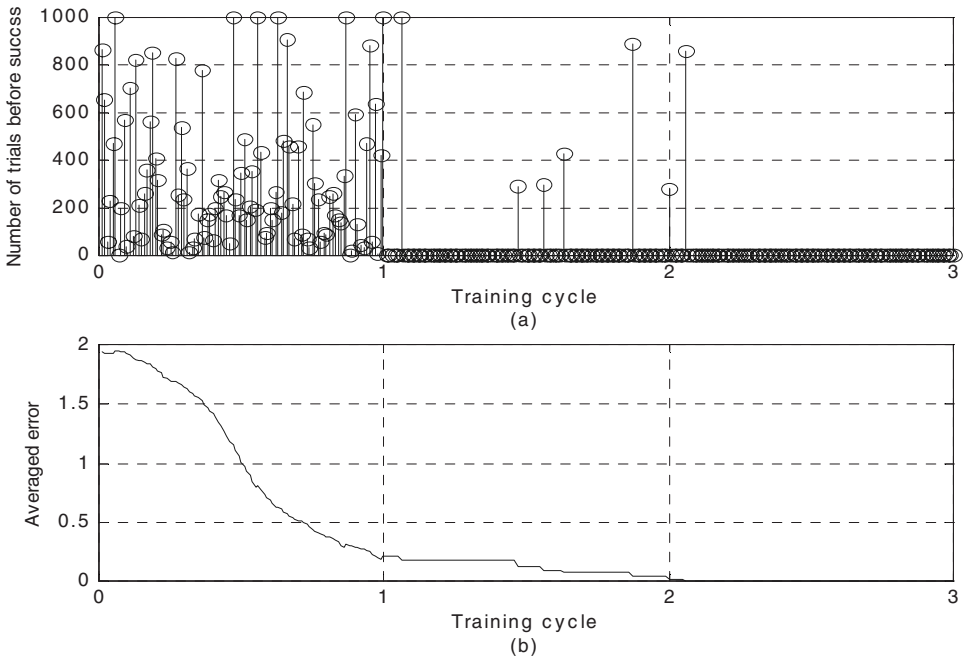


FIG. 9. (a) Number of trials for the modified SRV (real part) and (b) its corresponding averaged error.

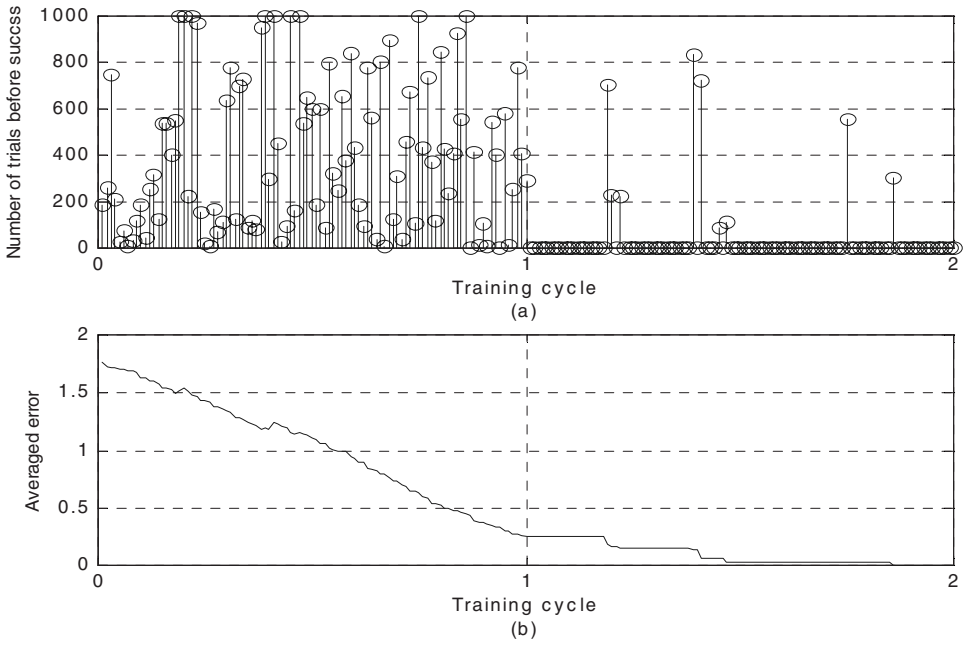


FIG. 10. (a) Number of trials for the modified SRV (imaginary part) and (b) its corresponding averaged error.

1/3 octave smoothing,[26,27] whereas Fig. 13(b) indicates the results after processing by our approach. Figures 13(a) and 13(b) demonstrate that colored signals have been successfully removed without any information about noise in advance. In general, the convergence of the reinforcement learning is improved by computing the temporal difference error instead of reinforcement error itself over several time steps.

Owing to the close relation between our algorithm and LSAE, we also apply noisy speech signal to them for comparison. In the second experiment, the additive noise signals include car interior noise and white Gaussian noise (WGN). The sampling frequency of speech signal is 19.98 kHz, which is contaminated by various noises with SNR ranging from −5 to 10 dB. The comparison is based on objective measures including SNR gain and log-spectral distance (LSD). Because the segment and synthesis mechanisms in both approaches are not exactly the same, to make a fair comparison, the SNR gain is defined as the ratio of total signal power to total noise power. Moreover, the LSD is defined as the distance between log-scaled DFT spectra for the clean and the processed signals averaged over all frequencies under the assumption that the change in phase is insignificant compared to that in log magnitude. As depicted in Fig. 14, although our method does not outperform the LSAE approach, it still shows the acceptable improvement in noise reduction. Similarly, as illustrated in Fig. 15, the norm of the log-spectral error is decreased by both approaches, but the processed signal is only slightly affected by our algorithm, especially when it was originally corrupted by WGN. One of the factors that could deteriorate the performance is the sparse resolution of quantization with respect to the generalization parameter $C=3$ in the speech enhancement. If the subsequent training input is chosen in the same neighbor-

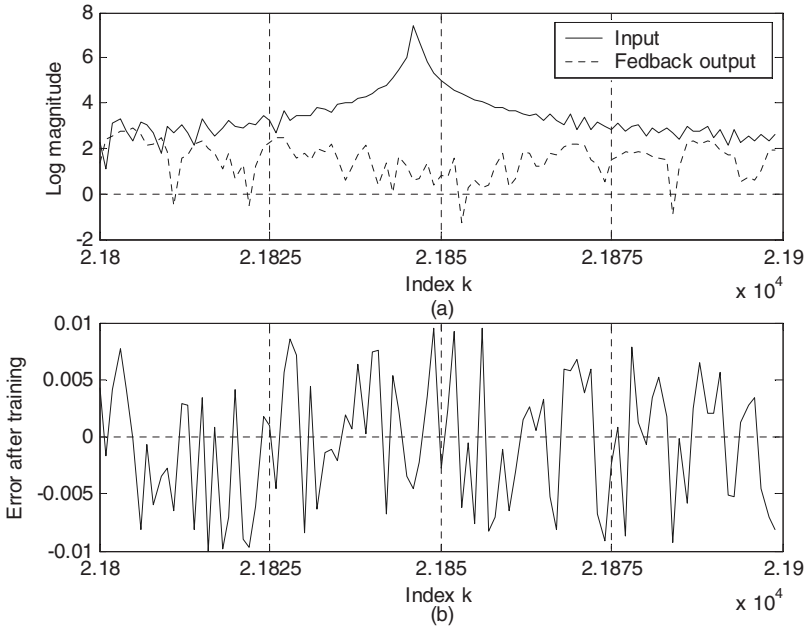


FIG. 11. (a) Comparison of the training data (real part) and (b) their error after training.

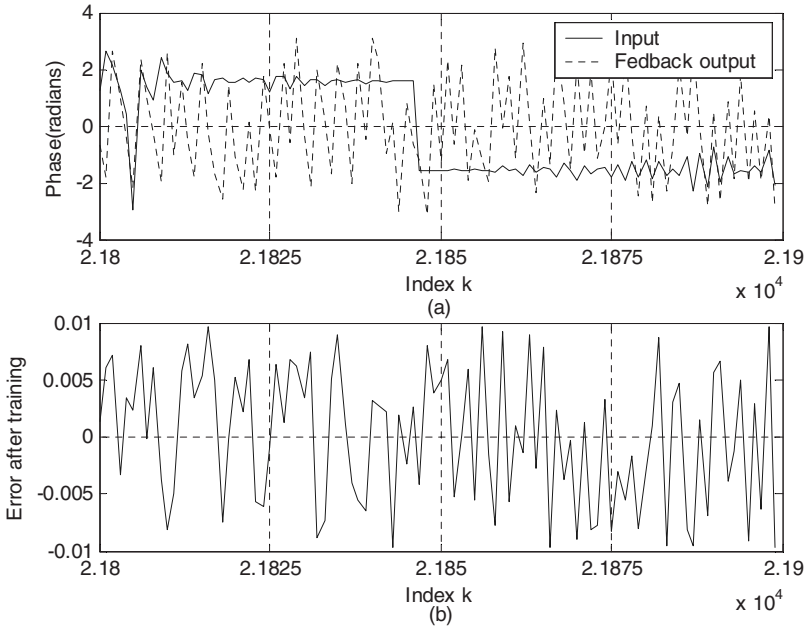


FIG. 12. (a) Comparison of the training data (imaginary part) and (b) their error after training.

hood as the previous one, some memory elements adjusted in the previous training sessions might be improperly altered.[28] This phenomenon is called learning interference and FCMAC-based algorithm is vulnerable to WGN owing to its dense distribution in semi-quefrency domain. Although using sparser quantization can solve this problem, doing so also leads to a null output when the signals cannot obtain sufficient credits in the ACE module. Therefore, quantization in the reinforcement learning is a more important issue than that in the supervised learning. In addition, one must face the trade-off between generality capacity and memory requirement when deciding how to partition the signals. Reorganizing the basis functions of FCMAC into a uniform fashion also leads to a better performance than that distributed sparsely in a diagonal fashion on the hyperspace. A fine quantization with many tiny regions promises more accurate approximation of nonlinear functions, but is time consuming in learning the correct output for each region. Obviously, an adaptive learning algorithm based on experience is needed for the quantization of signals. Moreover, utilizing the reinforcement learning in multiinput and multioutput systems presents special challenges. First, one has to design a critic element to consider all signals simultaneously, although only a reward or a punishment was fedback from the environment. Second, learning time is dramatically extended as the training data increase owing to reinforcement learning's inherently lengthy search for each state-action pair.

Most speech estimators use a gain function to modify the spectral amplitude whereas the phase remains unchanged. Conventional gain function such as spectral subtraction depends only on the measured signal level of the current frame and the estimated noise level, whereas the modified SRV uses the new situation $s_{ij}(t+1)$ and current critic weight $v_{ij}(t)$ to estimate the internal reinforcement signal $\hat{r}(t+1)$ [see Eq. (17)]. Moreover, the LSAE uses voice

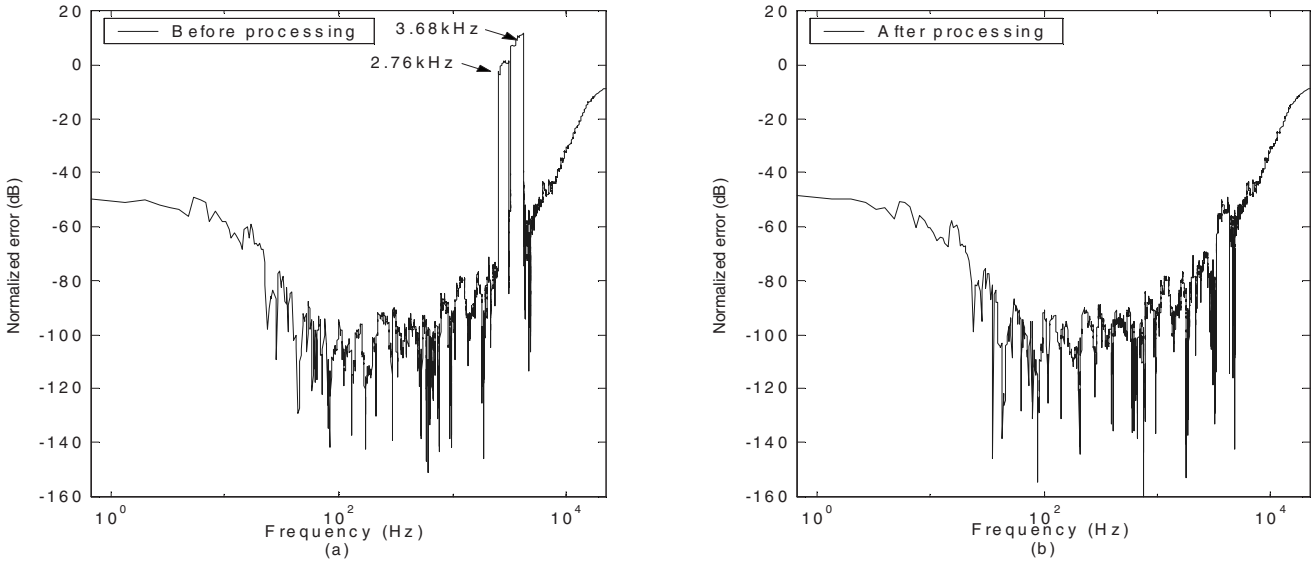


FIG. 13. Comparison based on 1/3 octave smoothing. Normalized error between (a) colored and original signals, and (b) output and original signals.

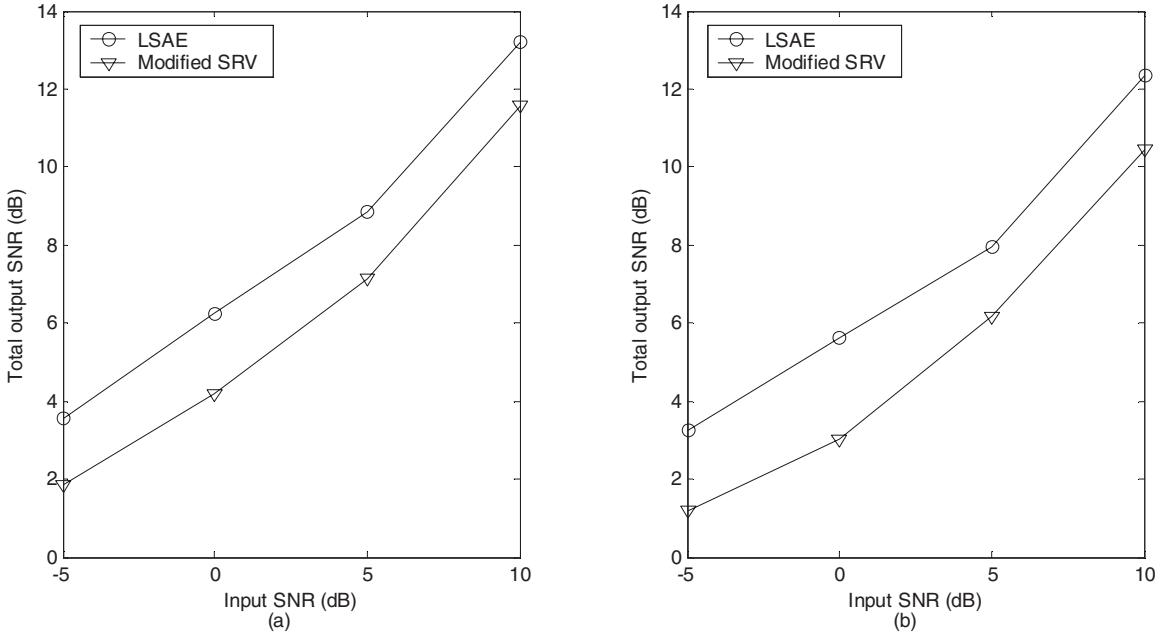


FIG. 14. Comparison of SNR for various noise types and levels by using LSAE and modified SRV. (a) Car interior noise and (b) white Gaussian noise.

activity detector (VAD) to control the update of the noise estimation, this design is similar to our approach that uses the internal reinforcement signal $\hat{r}(t)$ to update weights in Eqs. (18) and (19). However, the VAD-based estimator is difficult to tune and its application to low SNR speech usually results in clipped speech.[29] As a result, the denoised signal tends to be deteriorated in the tone. Reinforcement learning does not require an *a priori* dynamic model, but learns on the basis of the experience obtained directly from the environment. If the future reinforcement prediction $p(t+1)$ were right, the weights of ASE-SRV and ACE would be adjusted in the correct direction. The primary disadvantage of the reinforcement learning introduced is that many repeated experiences, in general, are required to reach an optimal (or suboptimal) strategy, especially for the system that starts in a poor initial condition. In addition, unlike supervised learning, the optimization problem is much harder because only partial information is available. As the evaluation element is used for estimating how many credit an earlier action should have, there is no other useful action evaluation fedback to the processor regarding its transient situation. It tends to ignore a destructive action or signal that seldom happens. In this invalid prediction case, it is better to adopt FCMAC as the quantizer, since the neighboring cells of FCMAC can share the same memory when they have similar inputs. Therefore, not only the weights of the current training input, but also those in its neighborhood are adjusted. This

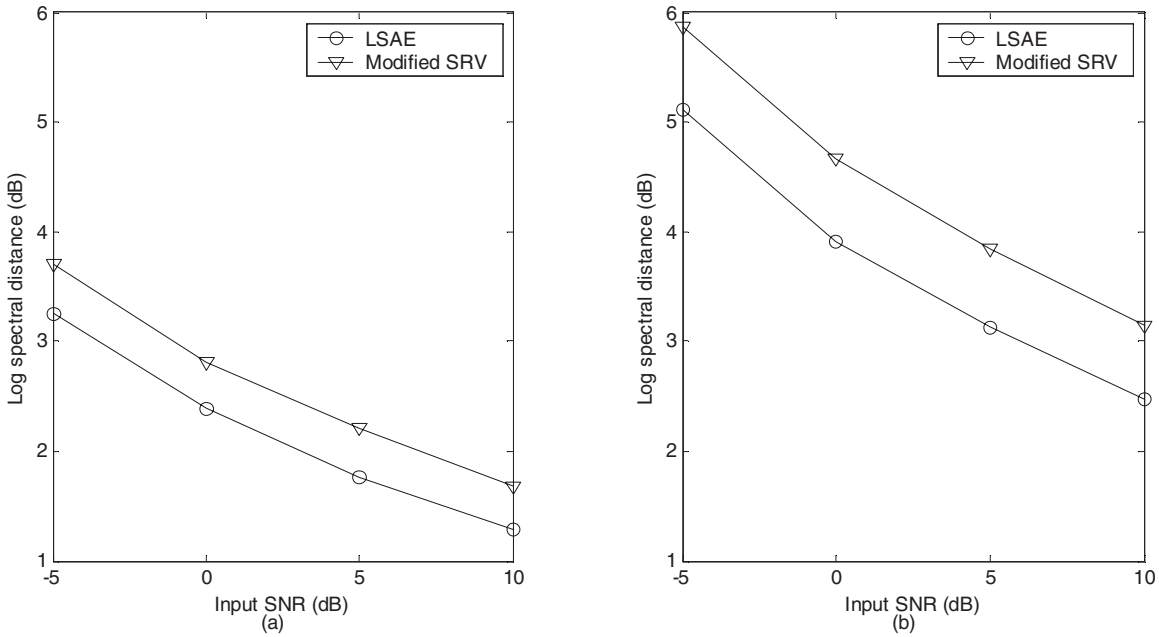


FIG. 15. Comparison of LSD for various noise types and levels by using LSAE and modified SRV. (a) Car interior noise and (b) white Gaussian noise.

sharing capability provides generalization, so the information can be interpolated even for the cells that have not yet been entered. Moreover, the performance of the modified SRV depends on the discount factor γ as well as $p(t+1)$. Notice that the closer γ is to 1, the longer the term of future external reinforcement signal is considered in the critic prediction. Owing to the intrinsically stochastic property, the modified SRV approach appears much more feasible than AHC method as a practical learning strategy in digital signal processing.

IV. CONCLUSION AND FUTURE RESEARCH

A primary difficulty for any reinforcement learning design is to identify an appropriate set of states, actions, and critics. The proposed signal-processing scheme covers all these issues and has a learning mechanism realized by a self-organizing FCMAC. The motivation to integrating these different strategies is to develop a more effective approach for systems with the characteristics of nonlinearity and uncertainty. In this experiment, a neural network is designed to remove the audio/speech signals colored by the additive noises. The learning mechanism plays a role of a compensator for the unwanted sound from the speakers to the listeners. The uncertainty and interaction between the sub-systems can also be handled in the learning process of the neuro-fuzzy system. We demonstrate that this FCMAC-SRV neural network has the capability of approximating complex, nonlinear functions with multidimensional inputs. In this research, the modified SRV algorithm is treated as the weight-update rule of FCMAC, which is used as a signal quantizer of the modified SRV from other viewpoints. The proposed system can be implemented on any nonlinear system by giving appropriate criteria as the performance measures. The simulation results show that a noise-removing task can be learnt even with little *a priori* knowledge. However, other applications are still limited to the problems of the highly demanding memory and expensive computation. We are currently investigating how to generalize this approach to cover broader applications such as signal separation and cross-talk cancellation.

ACKNOWLEDGMENTS

The work reported in the paper was supported in part by a grant from the Ministry of Education of Taiwan. The authors also thank the anonymous reviewers for their useful comments that were essential in improving the content of this paper.

[1]B. Gold and N. Morgan, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music* (Wiley, New York, 2000).
[2]S. Haykin, *Adaptive Filter Theory* (Prentice–Hall, Englewood Cliffs, NJ, 2001).
[3]S. V. Vaseghi and B. P. Milner, "Noise compensation methods for hidden Markov model speech recognition in adverse environments," IEEE Trans. Speech Audio Process. **5**, 11–21 (1997).
[4]L. Arslan, A. McCree, and V. Viswanathan, "New methods for adaptive noise suppression," IEEE Int. Conf. Acoustics, Speech, and Signal Processing **1**, 812–815 (1995).
[5]M. Karjalainen, "Immersion and content—A framework for audio research," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Oct. 1999, pp. 71–74.
[6]K. S. Hwang, C. S. Lin, and C. L. Chang, "Fuzzy cerebellar model articulation controller," *Proceedings of the Second Chinese World Congress on Intelligent Control and Intelligent Automation*, Xian, China, June 23–27, 1997, pp. 1536–1541.
[7]A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," IEEE Trans. Syst. Man Cybern. **13**, 834–846 (1983).
[8]V. Gullapalli, "A stochastic reinforcement learning algorithm for leaning real-valued functions," Neural Networks **3**, 671–692 (1990).
[9]V. Gullapalli, J. A. Franklin, and H. Benbrahim, "Acquiring robot skills via reinforcement learning," IEEE Control Syst. Mag. **14**, 13–24 (1994).
[10]Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **33**, 443–445 (1985).
[11]I. Cohen, "Speech enhancement using a noncausal *a priori* SNR estimator," IEEE Signal Process. Lett. **11**, 725–728 (2004).
[12]I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," IEEE Signal Process. Lett. **9**, 113–116 (2002).
[13]B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking," *Time Series Analysis*, M. Rosenblatt, editor, Wiley, New York, Jun. 1963, pp. 209–243.
[14]A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1989).
[15]I. Yamada and K. Sakaniwa, "An optimal design for a homomorphic deconvolution system," IEEE Trans. Signal Process. **40**, 2250–2260 (1992).
[16]A. Lo Schiavo and A. M. Luciano, "Powerful and flexible fuzzy algorithm for nonlinear dynamic system identification," IEEE Trans. Fuzzy Syst. **9**, 828–835 (2001).
[17]L. A. Zadeh, "Fuzzy logic," IEEE Comput. **21**, 83–93 (1988).
[18]C.-C. Jou, "A fuzzy cerebellar model articulation controller," *IEEE International Conference on Fuzzy Systems*, San Diego, CA, 1992, pp. 1171–1178.
[19]S. H. Lane, D. A. Handelman, and J. J. Gelfand, "Theory and development of higher-order CMAC neural networks," IEEE Control Syst. **12**, 23–30 (1992).
[20]C. S. Lin and C. Kyriakakis, "A fuzzy cerebellar model approach for synthesizing multichannel recordings," *113th AES Convention*, Los Angeles, CA, Oct. 2002, Preprint No. 5675.
[21]J. A. Franklin, "Refinement of robot motor skills through reinforcement learning," *Proceedings of the 27th IEEE Conference on Decision and Control*, Austin, TX, Dec. 1998, pp. 1096–1101.
[22]C. W. Anderson, "Learning to control an inverted pendulum using neural networks," IEEE Control Syst. Mag. **9**, 31–37 (1989).
[23]R. S. Sutton, "Learning to predict by the methods of temporal difference," Mach. Learn. **3**, 9–44 (1988).
[24]R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Mach. Learn. **8**, 229–256 (1992).
[25]C. S. Lin and C. Kyriakakis, "Multi-frequency noise removal based on reinforcement learning," *115th AES Convention*, New York, NY, Oct. 2003, Preprint No. 5966.
[26]J. Koring and A. Schmitz, "Simplifying cancellation of cross-talk for playback of head-related recordings in a two-speaker system," Acustica **79**, 221–227 (1993).
[27]B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 3rd ed. (Academic Press, San Diego, CA, 1989).
[28]J. S. Albus, "A new approach to manipulator control: The cerebellar model articulation controller (CMAC)," J. Dyn. Syst., Meas., Control **97**, 220–227 (1975).
[29]D. Burshtein and S. Gannot, "Speech enhancement using a mixture-maximum model," IEEE Trans. Speech Audio Process. **10**, 341–351 (2002).

Adaptive spatial combining for passive time-reversed communications[a)]

João Gomes[b)]
Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisboa, Portugal

António Silva[c)] and Sérgio Jesus[d)]
Institute for Systems and Robotics, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

(Received 5 April 2007; revised 27 May 2008; accepted 28 May 2008)

Passive time reversal has aroused considerable interest in underwater communications as a computationally inexpensive means of mitigating the intersymbol interference introduced by the channel using a receiver array. In this paper the basic technique is extended by adaptively weighting sensor contributions to partially compensate for degraded focusing due to mismatch between the assumed and actual medium impulse responses. Two algorithms are proposed, one of which restores constructive interference between sensors, and the other one minimizes the output residual as in widely used equalization schemes. These are compared with plain time reversal and variants that employ postequalization and channel tracking. They are shown to improve the residual error and temporal stability of basic time reversal with very little added complexity. Results are presented for data collected in a passive time-reversal experiment that was conducted during the MREA'04 sea trial. In that experiment a single acoustic projector generated a 2/4-PSK (phase-shift keyed) stream at 200/400 baud, modulated at 3.6 kHz, and received at a range of about 2 km on a sparse vertical array with eight hydrophones. The data were found to exhibit significant Doppler scaling, and a resampling-based preprocessing method is also proposed here to compensate for that scaling.
 [DOI: 10.1121/1.2946711]

PACS number(s): 43.60.Dh, 43.60.Tj, 43.60.Gk, 43.60.Fg [DRD] Pages: 1038–1053

I. INTRODUCTION

Time reversal is a wave backpropagation technique that cleverly exploits the reciprocity of linear wave propagation to concentrate signals at desired points in a waveguide with little knowledge about the medium.[1] The potential of time reversal for underwater communications was recognized and attracted much attention after the practical feasibility of this technique was demonstrated in the ocean.[2]

Active time-reversed (TR) focusing is achieved by transmitting a channel probe from the intended focal spot to an array of transducers that sample the incoming pressure field. These signals are then reversed in time and retransmitted, creating a replica field that converges on the original source location and approximately regenerates the initial wave form, undoing much of the delay dispersion caused by multipath. Due to its peculiar mode of operation, this type of source/receiver array is often referred to as a time-reversal mirror (TRM). When the principle of time reversal is applied to digital communications, the measured probe source ping is modulated with an information-bearing wave form, which can then be demodulated at the focus with relatively low algorithmic complexity. Passive time reversal, or passive phase conjugation[3] (PPC), is conceptually similar to the above technique, yet both the probe and message are sequentially sent from the source, so the array only operates in receive mode. Focusing is performed synthetically at the array by convolving the time-reversed distorted probes with received data packets.[3] This is, in fact, a multichannel combining (MC) strategy[4] whose parameters are directly measured from the data, not derived by optimizing a cost function.

Time reversal finds applications in diverse areas such as optics, materials testing, imaging, and medicine.[5] In underwater acoustics, digital communications have provided the backdrop for many published applications of this technique. In fact, the wave form regeneration property of TRM is highly relevant in underwater acoustic communications, where intersymbol interference (ISI) caused by multipath is usually the single most important distortion to be compensated.[4] Stojanovic[6] provides an overview of much of the research work in this area.

Both active[7,8] and passive[9–11] TR communications have been demonstrated in the ocean, the latter being more popular due to a simpler hardware setup. These experimental results, and other theoretical analyses,[6,12,13] suggest that time reversal by itself will not ensure reliable detection of the transmitted symbols, and should be complemented by adaptive equalization at the receiver to remove the residual ISI and compensate for channel variations. Arguably, the overall reduction in computational complexity at the receiver af-

[a)]Portions of this work were presented at the MTS/IEEE Oceans'06 Conference, Boston, MA, September 2006.
[b)]Electronic mail: jpg@isr.ist.utl.pt
[c)]Electronic mail: asilva@ualg.pt
[d)]Electronic mail: sjesus@ualg.pt

forded by the integration of time reversal into an acoustic link more than makes up for the moderate degradation in performance.

Most of the experiments reported to date employ single-carrier coherent signaling, although time reversal can easily be adapted to other modulations as well.[14] In fact, in line with a popular trend in underwater communications, several techniques first developed in wireless terrestrial communications have been investigated and proposed for acoustic links based on time reversal. This trend continues with multiple-input/multiple-output communications, which have provided large performance improvements in wireless radio and show great promise in underwater communications.[15]

In the methods that have been proposed so far for simultaneous equalization and time reversal the two systems are operated in tandem, i.e., a TRM creates a single-channel signal which is then independently processed by an equalizer.[6,12,16] In PPC, however, the signals received at an array of hydrophones are synthetically combined after convolving them with estimates of the (reversed) channel impulse responses. This provides increased flexibility relative to active TR, as these signals may be individually postprocessed prior to generating a single-channel wave form.

This paper examines low-complexity PPC approaches where a single combining coefficient is used per array sensor to improve upon the performance of basic PPC under channel variations. These coefficients are adjusted at each symbol interval by either iteratively minimizing the output mean-square error (MSE) or maximizing the output magnitude. This approach is motivated by the observation that poor signal-to-interference (ISI+noise) ratio that occurs due to environment mismatch between the probe and packet transmissions can often be attributed to partially destructive interference between contributions from different hydrophone signals, in spite of appropriate temporal alignment. Note that the proposed structures are actually very short multichannel equalizers, and one could envisage using more elaborate filters as in conventional equalizers.

Environment mismatch is also addressed in decision-directed PPC[17] (DDPPC) by tracking the channel impulse response continuously throughout a data packet to virtually eliminate the delay between probe capture and filtering. It should be emphasized that the MC methods proposed here are simpler, as the probe is only captured during the packet preamble, and subsequently only one coefficient per sensor is tracked. By contrast, DDPPC must propagate a fully adaptive model of the channel response measured in each sensor (or at least the portion of it with higher energy and greater temporal stability), which is not necessarily less computationally demanding than conventional equalization schemes.

Results from a passive time-reversal experiment conducted off the west coast of Portugal during the MREA'04 sea trial are presented. A single acoustic projector generated a 2/4-PSK stream at 200 and 400 baud around a carrier frequency of 3.6 kHz, and the signals were received at a range of about 2 km on a vertical array with eight unevenly spaced hydrophones. The channel end points were in motion during this experiment, inducing significant Doppler scaling in the observed wave forms that could degrade the performance of TR focusing if left uncompensated.[11] A broadband Doppler compensation method is proposed here, theoretically analyzed, and shown to perform very effectively in practice. The method itself is simple and similar ideas have been proposed in the past, but to the best of our knowledge no prior analysis of Doppler compensation on TRM performance has been published. In addition to documenting the performance of proposed MC algorithms, this work also aims to provide experimental results on various aspects of channel characterization, conventional equalization, and plain time reversal.

The paper is organized as follows. Section II introduces the signal model used for time reversal and describes the Doppler compensation method. Section III presents the multichannel combining algorithms, illustrates their performance in a simulated scenario, and discusses the synchronization and normalization postprocessing steps that are required to estimate data symbols from the TRM output. Section IV describes the MREA'04 sea trial and presents experimental results. The proposed MC algorithms are characterized and compared with plain multichannel equalization, plain TRM, simultaneous TRM and single-channel equalization, and DDPPC. Channel/probe estimation issues are also considered. Finally, Sec. V summarizes the main results, draws some conclusions, and suggests future research.

II. MODELING OF TIME REVERSAL

This section presents the notation used in the sequel for coherent communication using time reversal. The reader is referred to Refs. 1 and 2 for an overview of (narrowband) TR theory, and to several references on (broadband) TR communications[8,9,11] for lengthier discussions on specific aspects of data transmission. Throughout the paper convolution is denoted by the binary operator $*$ and complex conjugation by the superscript $(\cdot)^*$.

As is usually done in the context of bandwidth-efficient coherent communications, a complex representation in terms of baseband equivalent signals (i.e., complex envelopes) will henceforth be adopted for the real passband wave forms that are actually transmitted and received.[18] Time reversal of bandpass signals must then be replaced by time reversal and conjugation of complex envelopes, but other than that all equations describing the self-focusing property remain unchanged.

Let $p(t)$ represent the ideal basic pulse shape of the modulated wave forms that are exchanged between the source and the TRM, which can also act as a convenient channel probe. Other wave forms used for channel identification, such as linear frequency modulated pulses, are also appropriate. Denoting by $g_m(t)$ the impulse response between the focal point and the mth sensor of an M-element array, the distorted received probe is $h_m(t)=p(t)*g_m(t)$. In active time reversal the TRM then transmits an arbitrary data packet with pulse shape $h^*(-t)$ which, by virtue of TR focusing, will be approximately regenerated at the focal spot with the original pulse shape $p^*(-t)$. In passive phase conjugation the probe is followed by a data packet transmitted by the same source after a guard interval. Coherent single-carrier modu-

lation is assumed throughout this work, such that the received signal component at the mth sensor is given by

$$y_m(t) = \sum_k a(k) h_m(t - kT_b). \tag{1}$$

In the complex baseband representation underlying Eq. (1) the information symbols $\{a(k)\}$, transmitted with interval T_b, belong to a discrete *signal constellation*.[18] This is defined as a finite set of points in the complex plane that represent groups of bits from a digital message. Physically, the real and imaginary components of $a(k)$ are used for amplitude modulation of the in-phase and quadrature carriers when generating real bandpass wave forms. The symbols $\{a(k)\}$ are assumed to be uncorrelated random variables with zero mean and unit variance. The noise component will be ignored in the characterization of time reversal given below.

A plain passive mirror emulates active time reversal synthetically in a receive-only array. Its output, $z(t)$, is obtained by convolving the received packet (1) with the TR probe replica to generate a match-filtered signal in each sensor, $z_m(t)$, and then adding all contributions

$$z_m(t) = h_m^*(-t) * y_m(t), \tag{2}$$

$$z(t) = \sum_{m=1}^{M} z_m(t) = \sum_k a(k) q(t - kT_b). \tag{3}$$

In Eq. (3) $q(t)$ denotes the sum of temporal autocorrelations of received pulse shapes, sometimes referred to as the q-function[19] (QF). In a static ocean environment it is related to the medium impulse responses as

$$q(t) = \sum_{m=1}^{M} h_m^*(-t) * h_m(t) = r(t) * \gamma(t), \tag{4}$$

where

$$r(t) = p^*(-t) * p(t), \quad \gamma(t) = \sum_{m=1}^{M} g_m^*(-t) * g_m(t). \tag{5}$$

The multipath self-compensation property of time reversal implies that the spectrum of $\gamma(t)$ should be approximately constant across the bandwidth of $p(t)$ [and $r(t)$], so that $q(t) \propto r(t)$ and an undistorted modulated wave form is regenerated. In practice a delay is introduced to ensure causality of the time-reversed probe in Eq. (2), and all operations are performed in L-oversampled discrete-time signals $y_m(n) \triangleq y_m(nT_b/L)$ and $h_m(n)$.

Decoding is particularly simple when $p(t)$ has a root raised-cosine shape because then $r(t)$ in Eqs. (3) and (4) is a Nyquist pulse.[18] Out-of-band noise removal can be accomplished by actually transmitting fourth-root raised-cosine signaling pulses[10] $s(t)$ such that $p(t) = s^*(-t) * s(t)$, and then prefiltering all received wave forms (probes and packets) by $s^*(-t)$ to reject noise and attain the desired equivalent pulse shape $p(t)$. To avoid unnecessarily complicating our notation we assume that h_m and y_m in Eqs. (1), (2), and (4) have already undergone filtering by $s^*(-t)$. The spectra of $p(t)$, $r(t)$, and $s(t)$ are then related by

$$P(\omega) = R^{1/2}(\omega), \quad S(\omega) = R^{1/4}(\omega), \quad R(\omega) \geq 0. \tag{6}$$

Coherence issues. When channel variations occur between the probe and packet transmissions, autocorrelations in Eq. (4) are replaced by crosscorrelations between received pulses at different instants. This decreases the TRM's focusing power by degrading the impulselike behavior of $q(t)$.

To reduce the latency and mismatch between probe measuring and focusing, it is possible to discard the actual probe transmission and estimate it directly from a known preamble in the data packet.[16] This has the added benefit of reducing the additive noise component in the probe estimate $h_m(t)$, which generates undesirable convolutional noise during focusing. In a nonstatic environment it may also filter out the contributions from paths with poor temporal coherence, thus reducing the jitter in match-filtered outputs. This idea is taken further in DDPPC,[17] where the channel is tracked throughout data packets. A sharp QF is thus preserved even with very long packets because a low-latency channel estimate is always available.

The total number of parameters to be estimated by DDPPC in typical (multiple-hydrophone) discrete tap-delay-line models of underwater channels may be quite large and impose a significant computational burden. Alternative strategies that perform simpler adaptation and require fewer parameters, such as the ones addressed in this paper, may therefore be of interest. While Flynn *et al.*[17] argue in favor of iterative block least-squares estimation, in this work we adopt the exponentially windowed recursive least-squares (RLS) algorithm for channel estimation and tracking, which has been extensively used in underwater channel equalization and identification.

A. Doppler distortion and compensation

By decomposing a source with arbitrary space-time dependence as a superposition of monochromatic point sources, TR focusing may be shown to hold even for moving sources.[1] This work addresses a restricted case where the source is assumed to be moving slowly enough over a sufficiently short period so that the medium impulse responses linking it to the array transducers remain approximately constant. TR experiments suggest that this hypothesis is more plausible for predominantly horizontal motion, as the size of the focal spot in the horizontal plane is larger than along the depth axis.

Given a nominal transmitted passband wave form with carrier frequency ω_c, $\tilde{x}(t) = \mathrm{Re}\{x(t)e^{j\omega_c t}\}$, the equivalent Doppler-distorted transmission over a single path is

$$\tilde{x}((1+\beta)t) = \mathrm{Re}\{x((1+\beta)t)e^{j\omega_c \beta t} e^{j\omega_c t}\}, \tag{7}$$

where β is the time compression/dilation factor. For a moving transmitter with velocity v heading toward a static receiver in a medium with sound speed $c \gg v$, β is given by[20]

$$\beta = \frac{1}{1 - v/c} - 1 \approx \frac{v}{c}. \tag{8}$$

In terms of baseband signals, Eq. (7) amounts to time scaling of the original $x(t)$ and multiplication by a complex exponential with angular frequency $\omega_c \beta$. In a multipath environment

several delayed contributions of the above type are observed at the receiver, but if the propagation geometry and motion are predominantly horizontal, all scaling factors will be similar and compensating for the average Doppler usually suffices. Scattering by suspended particles may complicate the observed Doppler and multipath profiles, and is beyond the scope of this work.

Given an estimate of β the Doppler-compensated received signal is obtained from $y_m(t)$ as

$$y'_m(t) = y_m\left(\frac{t}{1+\beta}\right)e^{-j\omega_c[\beta t/(1+\beta)]}, \tag{9}$$

and used in all subsequent TR processing. The correctness of Eq. (9) can be asserted for the Doppler-distorted complex envelope of Eq. (7), $y_m(t)=x((1+\beta)t)e^{j\omega_c\beta t}$, in which case $y'_m(t)=x(t)$ as intended. The same Doppler correction is applied to received channel probes whenever they are available. This operation, which is shown to preserve the sharpness of the QF in Sec. A 1, leads to equivalent pulse shapes and impulse responses that are much more convenient to process and visualize due to their comparatively low rate of variation in time. A similar resampling approach has been proposed by Song *et al.*,[11] but justified only heuristically. Alternative Doppler compensation methods are needed when β cannot be assumed constant over a packet duration, but this hypothesis proved to be fully satisfactory for MREA'04 data.

The type of Doppler processing proposed here bears some resemblance to methods that have been proposed for varying the focal range of an active TRM through frequency shifting of a single captured probe.[21,22] The approach relies on the fact that, for a given environment, nominal frequency ω and range r, the ratio between (small) relative changes $\Delta\omega/\omega$ and $\Delta r/r$ is a constant known as the waveguide invariant. In a broadband communications context this property implies that the channel impulse response between a static source and a receiver at range $\Delta r+r$ approximately equals a time-scaled version of the nominal one.[23] The Doppler processing method described above can then be interpreted as follows: The packet transmission originating at range r is time scaled at the TRM to compensate for Doppler compression due to source motion, but this also rescales the underlying impulse response and makes it appear as though the source is positioned at a different range. Direct matched filtering using measured probes would then result in loss of sharpness of the QF due to range mismatch. Time scaling of channel probes prior to filtering counters this effect by replicating exactly the same impulse response distortion introduced in data packets, so that the range mismatch vanishes and a sharply focused signal is again obtained.

III. MULTICHANNEL COMBINING

A sum of matched filters such as the one used in PPC is a known generic front end for optimal multichannel data receivers under several criteria, including minimum MSE (Ref. 24) and minimum probability of sequence error under additive white Gaussian noise.[25] It should ideally be followed by a single-channel receiver to deal with residual ISI. Alternative strategies for multichannel equalization have been proposed for reducing the overall computational complexity or providing more flexibility when the channel responses are imperfectly known at the receiver.[1]

Formal justification for the multipath compensation property can be found elsewhere,[1,2,26] but intuitively it may be understood as follows. Each term $h_m^*(-t)*h_m(t)$ in Eq. (3) has a main lobe at time $t=0$ and (conjugate symmetrical) secondary lobes at other delays due to multipath. Main lobe contributions are all real, positive, and hence add up in phase. Secondary lobes, however, are not aligned in delay and phase for different sensors, and are not expected to reinforce each other the way that the main lobes do. As a result, an impulselike QF with dominating main lobe emerges as more and more terms are added. Perfect ISI removal through time reversal is only theoretically attained in the limit as the number of individual multipaths and/or uncorrelated receive elements gets very large.

The proposed multichannel combining algorithms are based on the assumption that, for moderate mismatch, the constructive interference of pulse contributions in $q(0)$ is partially lost even though the *shapes* of individual terms in the summation are not severely affected with respect to the static case. It should then be possible to mitigate this effect by multiplying each term by a single complex coefficient w_m to restore the phase alignment at $t=0$. Denoting by $h'_m(t)$ the actual pulse shape during focusing that differs from the channel probe $h_m(t)$, the modified TRM output in the presence of mismatch is given by

$$z(t) = \sum_{m=1}^{M} w_m z_m(t) = \sum_k a(k)q(t-kT_b), \tag{10}$$

$$q(t) = \sum_{m=1}^{M} w_m q_m(t), \quad q_m(t) = h_m^*(-t) * h'_m(t). \tag{11}$$

In terms of symbol-rate-sampled variables in Eqs. (10) and (11), $z(n) \triangleq z(nT_b)$, $z_m(n)$, $q(n)$, $q_m(n)$, one seeks to choose the coefficients w_m so that $q(n)$ approximates a discrete impulse and hence $z(n) \propto a(n)$. Note that the same notation is used for matched and mismatched q-functions in Eqs. (4) and (11); unless otherwise stated (e.g., in DDPPC), the mismatched case (11) will be assumed henceforth.

A. Numerical simulation

The main goal of this section is to provide motivation for the practical multichannel combining cost functions to be presented in Sec. III B by examining several approaches for merging QF contributions. To this end, the impact of Eq. (10) on TR focusing is illustrated in a simulated scenario that resembles the conditions of the MREA'04 sea trial described in Sec. IV A. Note, however, that this is an idealized simulation with no noise and in which a clairvoyant receiver precisely knows the individual contributions $q_m(t)$ to the overall QF as defined in Eq. (11).

The simulated environment is a range-independent ocean cross section with 130 m depth. The sound-speed profile, which was chosen as representative of MREA'04 mea-

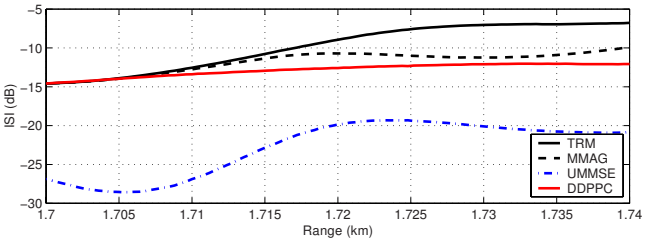


FIG. 1. (Color online) Evolution of the ISI metric, quantifying the similarity between the symbol-rate-sampled QF and a discrete impulse, in a simulated environment resembling the conditions of the MREA'04 sea trial (see Fig. 2). Results are shown for plain TRM, DDPPC, and two simplified criteria for multichannel combining (MMAG and UMMSE).

surements, is downward refracting with a thermocline at a depth of 20 m. The source is located at 70 m depth and 1.7 km nominal range. Surface reflection was modeled as a deterministic angle-dependent coefficient equal to the average (coherent) specular component $\alpha_S = e^{-2(k\sigma \sin\theta)^2}$, where k is the wave number at the carrier frequency of 3.6 kHz, $\sigma = 0.4$ m is the root mean-square surface roughness, and θ is the grazing angle. A constant bottom reflection coefficient $\alpha_B = 0.6$ was used. Attenuation/delay arrival data were generated with the Bellhop Gaussian beam ray tracer,[27] and used to compute received pulse shapes (200 baud, 100% rolloff) across the eight array sensors, where the maximum delay spread is about 30 ms. Delays were normalized so that the first arrival at the array always occurs at time 0 regardless of range.

Residual ISI. The source range was varied between 1.7 (nominal) and 1.74 km, the mismatched QFs calculated according to Eq. (11) and sampled at symbol rate. Residual ISI at each range is quantified by[28]

$$\mathrm{ISI}(q) = \frac{\Sigma_{n\neq 0}|q(n)|^2}{|q(0)|^2} = \sum_n \left| \frac{q(n)}{q(0)} - \delta(n) \right|^2, \tag{12}$$

which measures the similarity of $q(n)$, normalized to have unit magnitude at $n=0$, to an ideal discrete impulse $\delta(n)$. When $q(0)=1$ this coincides with the MSE $E\{|a(n)-z(n)|^2\}$ assuming that there is no noise and the symbols $a(n)$ are uncorrelated with unit power (Sec. II). Figure 1 depicts the behavior of the ISI metric for the following choices of combining coefficients w_m in Eq. (11):

(1) Plain time reversal using $w_m = 1$.
(2) Fully constructive interference of QF contributions at time 0 using $w_m = e^{-j \arg q_m(0)}$. As the maximum magnitude (MMAG) criterion of Sec. III B is an approximation of this, the same acronym is used in Fig. 1.
(3) The $q_m(n)$ are regarded as vectors over a finite time span around 0, $\mathbf{q}_m = [q_m(n)]_{n=-D}^{D}$ and linearly combined by the w_m to yield the best approximation to a discrete impulse $\boldsymbol{\delta} = [\delta(n)]_{n=-D}^{D}$ in the least-squares sense

$$\arg\min_{w_m} \left\| \boldsymbol{\delta} - \sum_{m=1}^{M} w_m \mathbf{q}_m \right\|^2. \tag{13}$$

This is approximated by the unconstrained minimum MSE (UMMSE) criterion of Sec. III B.

(4) DDPPC (Ref. 17) using $w_m = 1$ but fully updated channel estimates at each range, as in Eq. (4).

Figure 1 shows that restoring the constructive interference in QF contributions (MMAG) can indeed improve the performance of TRM under moderate mismatch, although there is clearly more residual ISI than with full channel tracking (DDPPC). Unlike the other algorithms in Fig. 1, UMMSE directly optimizes a MSE criterion that is related to the ISI metric, which partially explains its large performance benefit for the nominal range. In fact, mismatched QF contributions under these idealized conditions are such that very accurate approximations to $\delta(n)$ can be found for other ranges as well, and interestingly these do not seem to involve fully constructive interference at time 0. Note that the improvements afforded by MMAG, UMMSE, and DDPPC relative to plain TRM may vary significantly for other ranges and simulation setups.

Arguably, Fig. 1 should be interpreted with caution because UMMSE involves a form of channel inversion rather than pure channel matching, and therefore it is unfair to compare it to the other methods. Only in Sec. III B will practical approximate methods based only on TRM outputs be introduced, whose performance may legitimally be compared under a common MSE metric. Although channel inversion methods may potentially lead to smaller dispersion at the TRM output under ideal conditions, these are known to be much more sensitive to slight variations in the channel characteristics. This and the impact of noise help to explain the fact that performance gains of UMMSE using real data are much more modest than predicted here. In particular, the experimental results of Sec. IV B show that plain constructive interference of matched QF contributions using DDPPC provides lower ISI metrics.

B. Cost functions

Practical MC algorithms are formulated in terms of branch output sequences $z_m(n)$, rather than the underlying correlations $q_m(n)$ used in Sec. III A. A number of blind and nonblind (reference-driven) algorithms were developed in the course of this work,[29] two of which are examined here. Regardless of design criteria, the primary metric for performance assessment in Sec. IV is MSE defined as $E\{|a(n)-z(n)|^2\}$, where $a(n)$ denotes a transmitted symbol and $z(n)$ is the corresponding soft output of the receiver.

1. Maximum magnitude

Each coefficient in Eq. (10) performs a pure phase rotation, $w_m = e^{-j\theta_m}$, and the phases θ_m are chosen to maximize the expected squared magnitude of the mirror output. The first sensor is arbitrarily chosen as a reference by setting $w_1 = 1$. The cost function is

$$J_{\mathrm{mag}} = E\left\{ \left| z_1(n) + \sum_{m=2}^{M} z_m(n) e^{-j\theta_m} \right|^2 \right\}. \tag{14}$$

From $z_m(n) = q_m(n) * a(n)$ and the assumption of unit-power uncorrelated transmitted symbols introduced in Sec. II, Eq. (14) yields

$$J_{\mathrm{mag}} = \sum_n \left| q_1(n) + \sum_{m=2}^{M} q_m(n) e^{-j\theta_m} \right|^2 . \quad (15)$$

If the system operates with low mismatch, such that $q(n) \approx C\delta(n)$ in Eq. (11), then Eq. (15) will be largely dominated by the contribution for $n=0$. Ignoring the remaining terms, an optimal solution for θ_m is then readily given by $\theta_m = \arg q_m(0) - \arg q_1(0)$.

In Appendix B a simple adaptation rule for the angles based on gradient ascent is derived by differentiating Eq. (14) with respect to θ_m, obtaining a stochastic approximation to the gradient, and then using it as an error signal driving a phase-locked loop (PLL)-type filter.[4] This yields

$$\theta_m(n+1) = \theta_m(n) + K\Phi_m(n), \quad (16)$$

$$\Phi_m(n) = \mathrm{Im}\{(z(n) - z_m(n)e^{-j\theta_m})^* z_m(n) e^{-j\theta_m}\}, \quad (17)$$

where the loop gain K is adjusted empirically. This adaptation rule does not require a reference signal and, as in most blind filtering algorithms, a residual phase ambiguity exists that manifests itself as a rotation of the signal constellation. Similar to plain TR, postprocessing is therefore needed to properly align and scale the output constellation.

2. Unconstrained minimum MSE

Rather than aligning the z_m with unit-magnitude rotations, arbitrary coefficients w_m can be used to minimize the output error. In this work an exponentially weighted least-squares cost function is used,

$$J_{\mathrm{umse}}(n) = \sum_{k=0}^{n} \lambda^{n-k} \left| a(k) - \sum_{m=1}^{M} w_m(n) z_m(k) \right|^2 , \quad (18)$$

so that time adaptation of the w_m is actually carried out by the RLS algorithm (Ref. 30). Such a system effectively constitutes a very simple multichannel equalizer with one tap per sensor, which exploits probe preprocessing to significantly reduce the number of parameters to track. This approach will be termed UMMSE as in Ref. 29 although, strictly speaking, Eq. (18) is not a statistical criterion but rather a deterministic one. In reference-driven filtering schemes the packet symbols $a(k)$ to be used in Eq. (18) are assumed known during an initial training period, and afterward decisions based on the receiver output are used (decision-directed mode). Not only does this method handle phase synchronization, it also eliminates the need for output normalization.

C. TRM postprocessing

Symbol synchronization. In a practical TRM the output should undergo symbol synchronization to determine the time offset that maximizes a performance metric such as detection signal to noise ratio (SNR). Because the Doppler compensation technique of Sec. II A virtually eliminates any discrepancies in symbol rate, we simply calculate the L polyphase components of the oversampled discrete-time output, $z^{(l)}(n) \triangleq z((l+nL)T_b/L)$, $l=0,\ldots,L-1$, and choose the one with strongest average power. This is unnecessary when probes are estimated from data packets, as the best time offset is known to be zero beforehand because channel identification removes delay ambiguities in pulse shapes.

Phase synchronization. Doppler compensation was found to be effective at eliminating carrier frequency mismatches that result in sustained rotation of the TRM output over time. Still, a popular PLL approach[4] for phase synchronization was used as a postprocessor to track slow phase variations and hence properly align the output constellation. Specifically, a loop filter similar to Eq. (16) is used to update the estimated phase θ, driven by the error signal $\Phi(n) = \mathrm{Im}\{a(n)^* z(n) e^{-j\theta}\}$. Similar to Eq. (18), local symbol decisions should be used for computing $\Phi(n)$ upon entering decision-directed mode after the packet preamble. To simplify the comparison between different algorithms for ISI compensation (equalization, TRM, and multichannel combining), the same reference-driven phase synchronization method is used throughout this work. More appropriate choices are available for carrier synchronization in practical systems when ISI mitigation does not rely on an external reference.[18]

Output normalization. The final operation to be performed after symbol synchronization and constellation alignment is to account for an unknown scaling introduced by the channel and amplifiers at the transmitter and TRM. This gain varies throughout data packets as the channel and QF change, and should therefore be tracked by an automatic gain control (AGC)-like system. A simple possibility is to recursively compute an exponentially weighted average of the unnormalized TRM magnitude

$$\kappa(n) = \mu\kappa(n-1) + (1-\mu)|z(n)|, \quad 0 < \mu \leq 1, \quad (19)$$

and for a unit-magnitude constellation generate the normalized output as $z'(n) = z(n)\kappa^{-1}(n)$. Strictly speaking, normalization is not required to slice M-PSK constellations, but it is useful for estimating output MSE values.

IV. EXPERIMENTAL RESULTS

A. The MREA'04 experiment

The Maritime Rapid Environmental Assessment (MREA'04) sea trial[31] was conducted on the continental shelf off the west coast of Portugal in April 2004, in an area to the north of the Setúbal Canyon shown in Fig. 2. The weather was calm during the acoustic trials, with sea state between 1 and 2, low wind of less than 10–15 knot, generally from the North quadrant, and wave height less than 2 m. Extensive ground truth measurements of environmental parameters were performed before, during, and after the trial.[31]

The acoustic source was suspended from the NRV Alliance at depths ranging from 60 to 70 m, depending on vessel speed (up to about 1.6 m/s). The TR experiment started at a close range of 0.6 km to the south of the receiver array (38.36° N, 9.00° W) and the source progressively opened range to the southeast, up to 2 km, along an approximately range-independent path with 110 m water depth and a 1.5-m-thick silt and gravel sediment layer over a hard uniform sub-bottom. From Julian time 100.375 onward the Al-

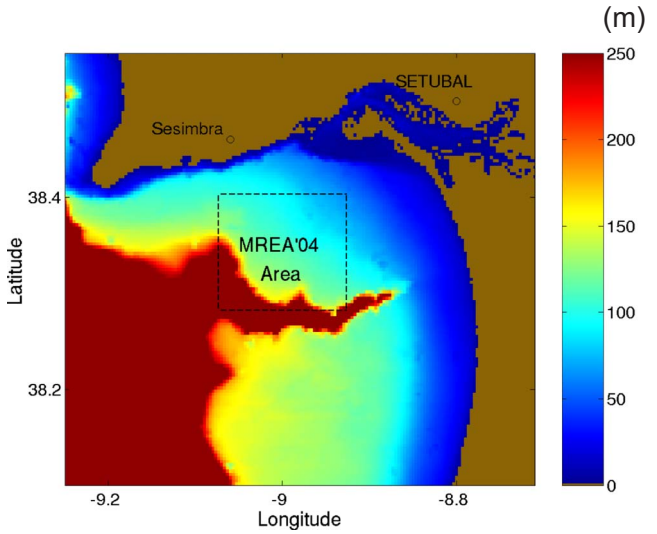


FIG. 2. (Color online) Site map for the MREA'04 sea trial, conducted off the west coast of Portugal in April 2004. The TR experiment took place in an approximately range-independent area (38.36°N, 9.00°W) with 110 m depth and downward-refracting sound-speed profile. The drifting receiver array had eight hydrophones at depths 10, 15, 55, 60, 65, 70, 75, and 80 m. The acoustic source was suspended from the NRV Alliance at depths of 60–70 m and towed at up to 1.6 m/s. Throughout the experiment the source-array range varied from 0.6 to 2 km.

liance maneuvered around a fixed position. The sound-speed profile was downward refracting, with a thermocline at a depth of about 20 m.

The drifting receiver is an acoustic oceanographic buoy (AOB) developed at the University of Algarve. The AOB can digitize and record eight hydrophone signals in the frequency band $[0.1,16]$ kHz, sampled at up to 60 kHz. A wireless link (WLAN) provides remote access to status information and data snapshots at ranges up to 10–20 km. In addition to eight hydrophones, vertically placed at depths 10, 15, 55, 60, 65, 70, 75, and 80 m, the AOB also has a 16-sensor thermistor chain spanning 80 m for water column temperature monitoring.

During a period of approximately 90 min modulated data were transmitted at a carrier frequency of 3.6 kHz, using symbol rates of 200 or 400 baud, and both 2-PSK and 4-PSK constellations. As discussed in Sec. II, fourth-root raised-cosine signaling pulses with 100% rolloff were used to simplify out-of-band noise removal at the receiver by matched filtering to the transmitted pulse shape. The signal bandwidth is therefore 400 Hz at 200 baud and is 800 Hz at 400 baud. Each individual transmission comprises a single truncated signaling pulse (ping) acting as a channel probe with symmetrical guard intervals for a total duration of 1 s, followed by a 20 s data packet. To enhance the SNR when directly measuring channel responses, probe pulses were sent with double the amplitude of signaling pulses in data packets. The source sequentially transmitted four packets for each of the following modulation formats: 2-PSK/ 200 baud, 2-PSK/ 400 baud, 4-PSK/ 200 baud, and 4-PSK/ 400 baud. The whole activity cycle, lasting for 336 s, was repeated every 360 s. The data set analyzed here comprises 200 probe/packet pairs.

B. Performance analysis

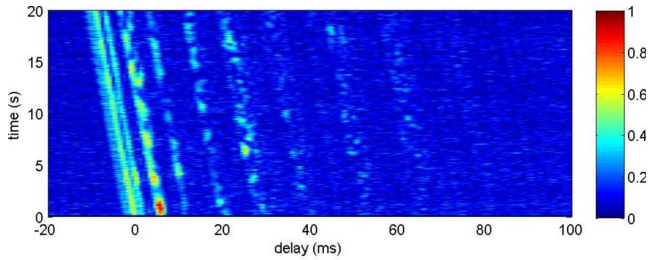
Received signals were passband filtered, sampled at 20 080 Hz, and converted to baseband. Packets were classified and frame synchronized by crosscorrelation with the first 2 kilosamples of all known modulated wave forms, then match filtered by the appropriate fourth-root raised-cosine pulse and resampled (oversampled) at $L=4$ times the symbol rate. No attempt was made to detect the channel probes; they were segmented based on their known position relative to the beginning of data packets, then match filtered and resampled as above.

1. Notes on filtering

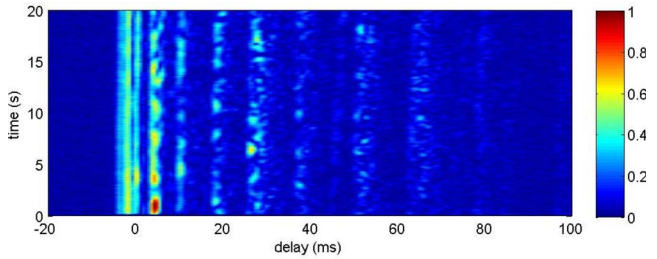
Throughout Sec. IV B (m,n) will denote the length of a single-channel filter with m causal coefficients and n anticausal ones. A multiple-input/single-output filter comprising p single-channel parallel filters of length (m,n) whose outputs are added to create a scalar output will be denoted by $(m,n)\times p$. Such p-channel filters are used when processing fractionally sampled communications wave forms and/or multisensor data, such that p equals the oversampling factor (relative to the symbol rate) times the number of sensors. In practical reference-driven filtering schemes, n anticausal coefficients are effectively obtained when the reference signal is delayed by n samples with respect to the received signals. In non-reference-driven (blind) schemes the distinction between causal and anticausal coefficients is meaningless and (m,n) should simply be interpreted as a filter with $m+n$ coefficients.

Regarding equalization, choosing a good combination of filter lengths from channel estimates under fractional sampling is known to be unreliable and often done offline by trial and error. For decision-feedback equalizers (DFEs) popular design guidelines[32] recommend using the feedback filter to cancel causal (postcursor) ISI, whereas feedforward filters will be much shorter to capture multipath energy and cancel anticausal (precursor) ISI. In this work appropriate equalizer lengths were set empirically for each packet in each experiment by searching over a plausible range of candidate lengths and selecting the one yielding the best performance. Somewhat unexpectedly, the best lengths reported below were found to be consistent across a clear majority of both 200 and 400 baud packets.

As in other references,[33] the impact of symbol errors on the performance of reference-driven channel estimation, equalization, and phase tracking algorithms is not addressed in this work. These subsystems are always operated in training mode, where the correct symbols are known, and the calculated performance metrics should then be interpreted as optimistic estimates of what could actually be achieved. For values of output MSE higher than about −5 dB the number of symbol errors is sufficiently large to have a significant impact on performance in decision-directed mode. This might cause divergence of adaptive algorithms if corrective measures are not implemented, such as freezing the updating of coefficients when unreliable decisions are detected.



(a)



(b)

FIG. 3. (Color online) Evolution of amplitude-normalized estimated channel responses at depth 75 m (hydrophone 7) for packet 149 (400 baud). A horizontal slice through any of the plots represents a snapshot of the time-varying response. The coherence time for this channel was estimated to be about 1 s. Estimates are based on RLS transversal filtering, four-oversampling, filter order (41, 10) per polyphase component (for order notation see Sec. IV B 1), and $\lambda=0.95$. (a) Before Doppler compensation, $f_d=1.65$ Hz at the carrier frequency of 3.6 kHz. (b) After broadband Doppler compensation as described in Sec. II A.

2. Channel responses

Figure 3(a) shows the evolution of the estimated impulse response at depth 75 m in one of the received 400 baud packets (PKT 149) with significant Doppler distortion. Channel identification was performed independently for each sensor using the exponentially windowed RLS algorithm.[30] For computational efficiency, each L-oversampled hydrophone signal was split into L polyphase components, $y_m^{(l)}(n)=y_m(l+nL)$, and these were used as references to a bank of L parallel RLS transversal filters fed by the known packet symbols. Each filter operates with 41 causal and 10 anticausal coefficients [abbreviated as (41,10) using the notation introduced in Sec. IV B 1] and forgetting factor $\lambda=0.95$ empirically adjusted to minimize the residual error variance. This technique decreases the overall computational complexity by a factor of L relative to direct identification of $y_m(n)$ from the zero-interpolated symbol sequence. Snapshots of the RLS coefficient vectors (estimated polyphase components of impulse responses) were taken every 20 symbol intervals and rearranged in the correct temporal order to produce the plot. The multipath arrival structure, spanning about 50 ms, is reasonably sparse and clearly visible in Fig. 3(a), as well as a time compression due to Doppler that causes the arrivals to slip by 14 samples (3.5 symbols) in the course of a 20 s packet. Figure 3(b) shows the impulse response estimate for the same packet after Doppler compensation as described in Sec. II A and Appendix A, where the multipath structure is seen to remain essentially unaltered. The coherence time for the channel of Fig. 3(b) was estimated to be about 1 s (Doppler bandwidth of 1 Hz). Once the average Doppler scaling

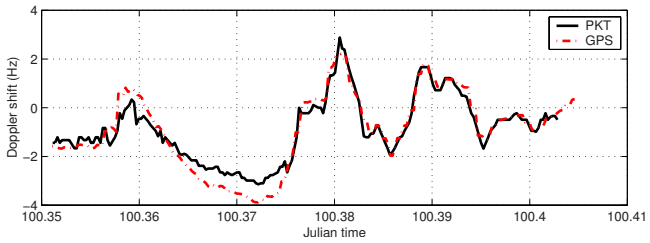


FIG. 4. (Color online) Estimated Doppler shift at the carrier frequency $f_c=3.6$ kHz from packet data and GPS navigation data. GPS-based estimates were obtained by determining the source velocity along the source-receiver direction, calculating the time scaling factor β according to Eq. (8), and plotting $f_c\beta$. In packet-based estimates β was directly obtained from the ratio between received and transmitted packet durations, averaged across all receiver hydrophones.

in received signals was compensated, it was found that including other dedicated symbol synchronization subsystems at the receiver was unnecessary. The various structures described below use fractional sampling, which can automatically perform fine adjustments to the sampling instants when needed.

The causal/anticausal filter lengths used in Fig. 3 were empirically chosen to capture most of the multipath energy in all 400 baud packets of the data set. In 200 baud packets the filter lengths used for channel identification could be reduced to (21, 7) without significantly affecting the residual error, i.e., while still capturing all relevant multipaths.

Figure 4 shows estimates of Doppler shift at the carrier frequency, f_c, obtained from (i) measurements of packet time scaling ($f_d=f_c\beta$, see Appendix A) and (ii) navigation data from the global positioning system (GPS) obtained at the source and receiver. The agreement between both curves is very good, suggesting that the physically motivated Doppler resampling procedure (9) is indeed plausible. As will be seen presently, best results are obtained in the low Doppler region to the right in Fig. 4, where the Alliance was relatively stationary.

Regarding input power levels, Fig. 5 shows the evolution of the signal to interference plus noise ratio (SINR) estimate defined in Eq. (C3) of Appendix C. It makes sense to

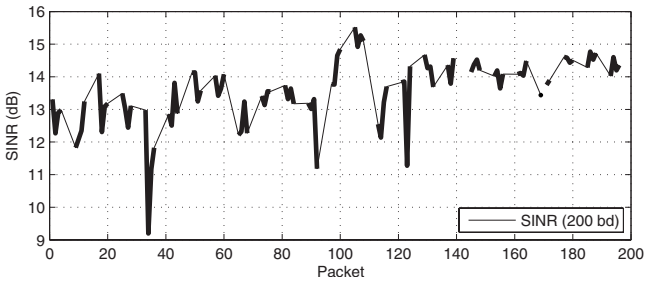


FIG. 5. Estimated input SINR (see Appendix C) based on channel identification [RLS transversal filtering, four-oversampling, order (21, 7) per polyphase component]. Forgetting factors were varied in the range $0.93\leqslant\lambda\leqslant0.98$, and observed MSE values were fitted to theoretical expressions accounting for excess adaptation MSE in RLS to estimate the actual power of interferences (ambient noise and reverberation). Signal power is directly given by the norm of the RLS coefficient vector for the best forgetting factor. SINR estimates are averaged across all receiver hydrophones. Values for 400 baud packets are omitted, as the required filter lengths are outside the valid range of theoretical expressions for excess MSE in RLS.

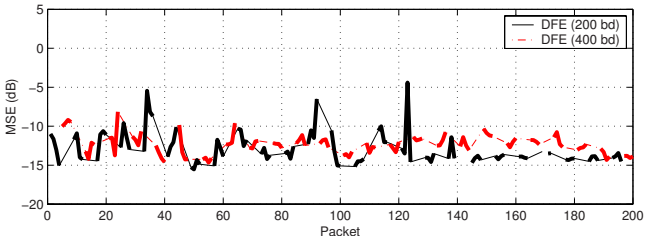


FIG. 6. (Color online) Performance of multichannel decision-feedback equalization using RLS, two-oversampling, eight sensors, $\lambda=0.95$. For each packet the lowest MSE obtained in a set of candidate equalizer lengths is shown. In most packets the best equalizer has $(2,1)\times 16$ feedforward and three feedback coefficients (for notation see Sec. IV B 1). The short feedback filter length suggests that the equalizer only exploits the arrivals shown in Fig. 3(b) up to 11 ms.

adopt this metric as the definition of input SNR ($\mathrm{SNR}_{\mathrm{in}}$), as it attempts to account for the useful input signal energy that can be captured by RLS-based methods used for equalization and probe estimation, casting as interference the remaining received energy. To compute SINR as outlined in Appendix C the forgetting factor λ was varied in the range $[0.93; 0.98]$. As in Fig. 3, lowest RLS residual errors were consistently obtained for $\lambda\approx 0.95$. The SINR estimates for 400 baud packets are not shown in Fig. 5, as the (41,10) filter order that is needed for reasonable modeling of multipath is too large and leads to an invalid denominator in Eq. (C2).

Similar to SINR, a SNR measure at the output of a digital receiver is given by[6,11,17] $\mathrm{SNR}_{\mathrm{out}}\approx \mathrm{MSE}^{-1}$ (see also Appendix C). For suitably defined SNR metrics, one expects to find $\mathrm{SNR}_{\mathrm{out}}\leqslant \mathrm{SNR}_{\mathrm{in}}$, where equality is attained when the receiver perfectly eliminates ISI without noise enhancement.[6,11] As discussed below, optimal MSE values presented in Fig. 6 turn out to be in reasonably good agreement with SINR in Fig. 5, particularly for packets 120,…,200 where Doppler scaling is low and $\mathrm{MSE}^{-1}\approx \mathrm{SINR}\approx 14$ dB. This suggests that (i) the effective input SNR is well captured by SINR and (ii) equalization essentially achieves the practical lower bound for output MSE, so these results can be used to assess the performance degradation incurred by alternative TR-based receivers.

3. Equalization

To benchmark the performance of TRM demodulation algorithms, data packets were processed by a conventional RLS-based ($\lambda=0.95$) multichannel decision-feedback equalizer.[25] Given the effectiveness of Doppler compensation, the equalizer was able to cope with residual phase fluctuations without the need for a carrier recovery subsystem. Presumably, such a system would be required with higher carrier frequencies. The full set of eight sensor wave forms was used, fractionally sampled by $L=2$. Fractional sampling of band limited single-carrier amplitude/phase-modulated signals eliminates aliasing and avoids having to estimate and track a single optimal sampling instant per symbol interval, using a timing recovery loop,[18] prior to equalization. The technique is used throughout this work either for stand-alone equalization or cascaded time reversal/equalization.

Figure 6 shows the best MSE values that were obtained on each packet by cycling over a set of candidate equalizer lengths. Most of these correspond to (2, 1) feedforward coefficients per polyphase component in each sensor [abbreviated as $(2,1)\times 16$ using the notation introduced in Sec. IV B 1] and three feedback coefficients. The equalizer time span is considerably shorter than the ISI duration shown in Fig. 3, but it agrees with the empirical determination of optimal TRM probe lengths reported in Sec. IV B 4. According to the DFE design guidelines outlined at the start of Sec. IV, the short feedback filter length suggests that the equalizer essentially exploits the arrivals shown in Fig. 3(b) up to the one at 11 ms. This could be due to fluctuations in the remaining contributions that preclude the coherent combination of multipath energy. Distinct MSE curves are given for 200 and 400 baud packets, the former showing better performance in packets where low Doppler shifts indicate more stationary channel responses (e.g., packets 120,…,200) because ISI spans fewer symbols. Under stronger Doppler (e.g., packets 40,…,80), the higher equalizer update rate in 400 baud packets seems to enable more effective tracking of channel variations and closes the MSE gap.

For completeness, the performance of multichannel fractionally spaced equalizers (FSE) was also evaluated. Results have been omitted here in the interest of space, but the MSE curves are similar to those of Fig. 6, shifted upward by about 0.5 dB, and these are attained in most packets using $(2,1)\times 16$ coefficients with the full set of eight sensors. Both the DFE and FSE use fractional sampling in the feedforward filter, but the former has an additional feedback filter that is fed by previous symbol decisions. If correct, these decisions enable the elimination of postcursor ISI without noise enhancement, whereas a FSE must compensate for both precursor and postcursor ISI using only the linear feedforward filter. Usually, this means that FSE filters will be longer than DFE feedforward filters, and the best output MSE will be higher due to noise enhancement.

The similarity of results between DFE and FSE (small MSE differences and similar optimal feedforward filter lengths) strengthens the previous argument that late arrivals in the MREA'04 data set lack coherence and hence cannot be effectively combined by these equalizers. The disparity between DFE and FSE will increase when equalization follows time reversal (see Fig. 8), as imperfect focusing results in longer-range, albeit mild, ISI.

As discussed in Sec. IV B 1, the above results assume that equalizers are only operated in training mode. However, comparable results are attained in most of the packets using a 400-symbol training sequence (which is used for probe estimation in TR-based algorithms below) followed by decision-directed adaptation if the RLS forgetting factor is increased to around $\lambda=0.99$ to achieve a larger effective averaging window.

4. Plain time reversal

Figure 7 shows the plain TRM output for a 200 baud/2-PSK packet, postprocessed as described below regarding Fig. 8. Pulse shapes for time reversal were obtained by directly observing the response to the single pulse that precedes each packet (see Sec. IV A). The plot shows quite stable behavior of the real part of the TRM output, $\mathrm{Re}\{z(n)\}$, over the packet

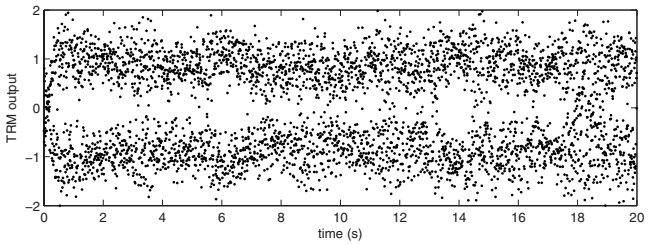


FIG. 7. Time evolution of the plain TRM output (real part) in PKT 155 (200 baud/2-PSK) using eight hydrophones. Pulse shapes for time reversal were obtained by directly observing the responses to the single pulse (ping) that precedes each packet. Postprocessing for symbol/phase synchronization and AGC is described in Sec. III C.

duration. TRM performance was found to be similarly stable in several other processed packets, which is somewhat unexpected given that the source was moving throughout much of the experiment, albeit at speeds not exceeding about 1.5 m/s. This contrasts with results reported by Flynn *et al.*,[17] where the collapse of plain TRM within 1 or 2 s of probe transmission motivated the development of DDPPC. The discrepancy may be attributed to the higher symbol rates (more than 2 kbaud) and denser multipath profile in that experiment, which probably translate into longer matched filters (probes) with less stable coefficients.

Figure 8(a) summarizes the MSE performance of plain TRM using observed probes (in short-term subintervals, see below). Results for 200 and 400 baud packets are not discriminated, as no significant differences in performance were found. Regarding the postprocessing steps of Sec. III C, phase alignment of the symbol-rate-synchronized TRM output is accomplished by a PLL whose generic loop filter update recursion (16) uses $K=5\times10^{-2}$. The AGC for output normalization (19) uses $\mu=0.995$, for an effective power averaging window length of about $(1-\mu)^{-1}=200$ samples. Whenever applicable, the same parameters are used in other TRM variants as well. When compared with the equalization results of Fig. 6 the MSE is seen to be higher by about 5 dB, with stronger interpacket variations. Notice the reduction in MSE and improved consistency across packets in the low-Doppler region to the right of Fig. 8 (see also Fig. 4), which agrees with the equalization results of Fig. 6.

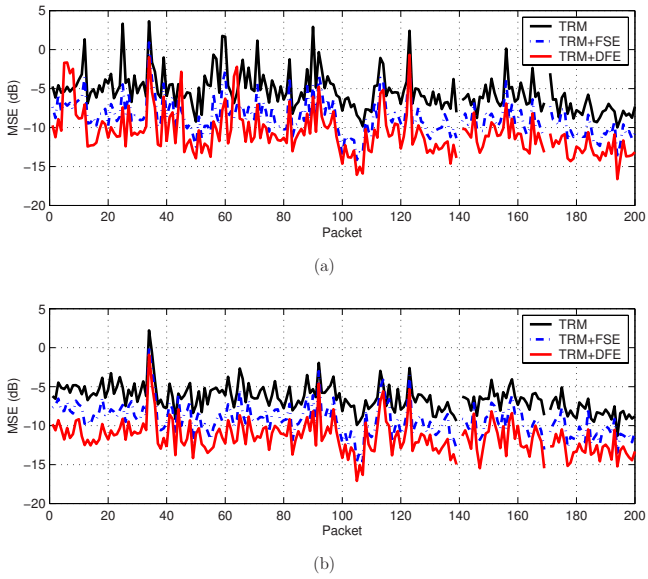


FIG. 8. (Color online) Performance of plain TRM using eight hydrophones and TRM with postequalization by DFE or FSE. MSE performance is evaluated on a short-term interval $0.5\leqslant t\leqslant2.5$ s. The equalizers use RLS, two-oversampling, $\lambda=0.95$. For each packet the lowest MSE obtained in a set of candidate equalizer lengths is shown. In most packets the best DFE has $(3,4)\times2$ feedforward and 20 feedback coefficients, whereas the best FSE uses mostly $(9,4)\times2$ coefficients. (a) Observed probes from a single transmitted ping before each packet. (b) Estimated probes by channel identification on a 400-symbol packet preamble. RLS parameters for identification were set as described in Fig. 3.

Figure 8(a) additionally shows MSE values when the TRM output, decimated to $L=2$ samples per symbol, is postprocessed by a DFE or a FSE. Within the set of considered equalizer lengths, best results for DFE were obtained with $(3,4)\times2$ feedforward and 20 feedback coefficients. This short feedforward filter/long feedback filter combination is consistent with the DFE design guidelines described at the start of Sec. IV. Imperfect TR focusing using a sparse receiver array results in moderate, but longer-term, residual ISI (both precursor and postcursor), which explains the longer filter lengths needed to cope with it relative to plain multichannel equalization. For FSE $(9,4)\times2$ coefficients yielded the best performance in most packets.

On average the MSE for TRM+DFE exceeds that of the multichannel DFE of Fig. 6 by 1.1 dB. Figure 8(b) repeats the above results for probes estimated from the packet's initial 400 symbols, with identification parameters set as previously described in Sec. IV B 2. Lower MSE values are usually obtained with this method than with probe observation, and performance is more consistent across successive packets, but otherwise similar comments apply. Unless explicitly noted, probe estimation is used in the remainder of this section.

To quantify the degradation in TRM output over time, MSE values were averaged over short-(S), medium-(M) and long-term (L) subintervals in each packet. These were defined as S: $0.5\leqslant t\leqslant2.5$ s, M: $4\leqslant t\leqslant6$ s, and L: $10\leqslant t\leqslant12$ s. The reference $t=0$ was set to the beginning of packets for observed probes and the end of the identification preamble for estimated probes. The following results were obtained.

Plain TRM. Figures 8(a) and 8(b) pertain to S intervals. The average increases in MSE over M and L intervals relative to S for all 200 baud (respectively, 400 baud) packets are 0.6 and 1.2 dB (respectively, 1.1 and 2.0 dB). Increased sensitivity to mismatch at 400 baud was expected, as oversampled pulse shapes have finer temporal structure.

TRM+DFE. The average increases in MSE over M and L intervals relative to S for 200 baud (respectively, 400 baud) packets are 0.2 and 0.4 dB (respectively, 1.1 and 1.7 dB). Thus, while TRM+DFE provides significantly lower *absolute* MSE than plain TRM (Fig. 8), it was found that MSE *fluctuations* over time are not effectively attenuated by postequalization in 400 baud packets. This is consistent with Fig. 8, where the MSE of TRM+DFE tracks that of TRM at an approximately constant offset.

Actual MSE fluctuations from S to M or L vary widely across packets, with standard deviations exceeding 1 dB.

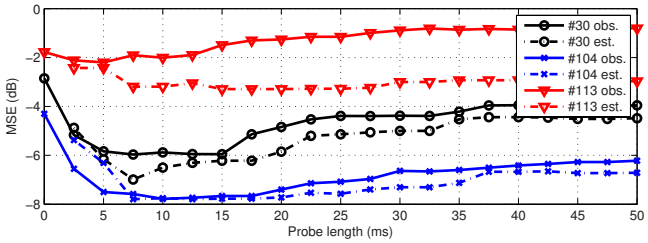


FIG. 9. (Color online) Impact of probe length on short-term plain TRM performance (eight hydrophones) for three individual packets (PKT 30, 104, 113) using direct probe measurements (obs.) or channel estimation from packet preambles (est.). Lowest MSE values are obtained for truncated probes of about 7–15 ms duration, which discard much of the multipath structure shown in Fig. 3. The trend for other packets (not shown) is similar, and in agreement with the short feedback filter lengths that were found for the equalization results shown in Fig. 6.

Contrary to what was expected, no obvious correlation was found between the extent of MSE degradation in M, L and the severity of Doppler distortion depicted in Fig. 4.

The above results were obtained with probe lengths optimized for each individual packet. Figure 9 depicts the evolution of short-term MSE in three representative packets as a function of probe length using either direct measurements (observed) or channel estimation from packet preambles (estimated). Interestingly, the lowest MSE values are obtained for truncated probes of about 7–15 ms duration, which discard much of the multipath structure shown in Fig. 3. This may be due either to the low energy of these discarded arrivals relative to the noise background or, more likely, to fluctuations that preclude the effective coherent combination of energy from those paths. Note that all reported MSE/ISI values for time reversal based on optimal probe lengths should be interpreted as lower performance bounds in practical systems where probes are truncated before focusing and packet decoding.

Finally, Fig. 10 shows the evolution of short-term MSE and bit error rate in TRM and TRM+DFE as a function of the number of sensors M, averaged over all packets. Array elements are sequentially selected from top, such that by increasing M the size of the aperture increases as well. Reported experiments[11] indicate that the performance of a long and dense TRM tends to saturate with growing M after an initial rapid improvement, as the device already captures much of the energy in the water column and its ISI mitigation ability improves very slowly. By contrast, Fig. 10 shows a more linearlike trend (in decibel scale) as a function of M, with MSE gains dropping from -1.5 dB for $M=2$ until -0.7 dB for $M=8$. This suggests that the TRM still operates in a nonsaturated regime where each additional sensor introduces significant spatial diversity. Equivalently, impulse responses are sufficiently different between sensors spaced 5 m apart to ensure that QF contributions do not interfere constructively in the sidelobe region. The reduction in MSE by about 6 dB as M varies between 1 and 8 is consistent with simulation results based on the parameters of Sec. III A. Postprocessing by the same DFE of Fig. 8 yields similar MSE improvements, in agreement with other TR experiments.[11]

The existence of a thermocline at 20 m induces rapid delay variations in the arrivals with low grazing angles as a function of depth, thus enhancing differences in the pattern of early arrivals (direct path and surface reflection) between the upper and lower sections of the array (located above and below the thermocline, respectively). When adding sensors to the TRM from top one would therefore expect to see

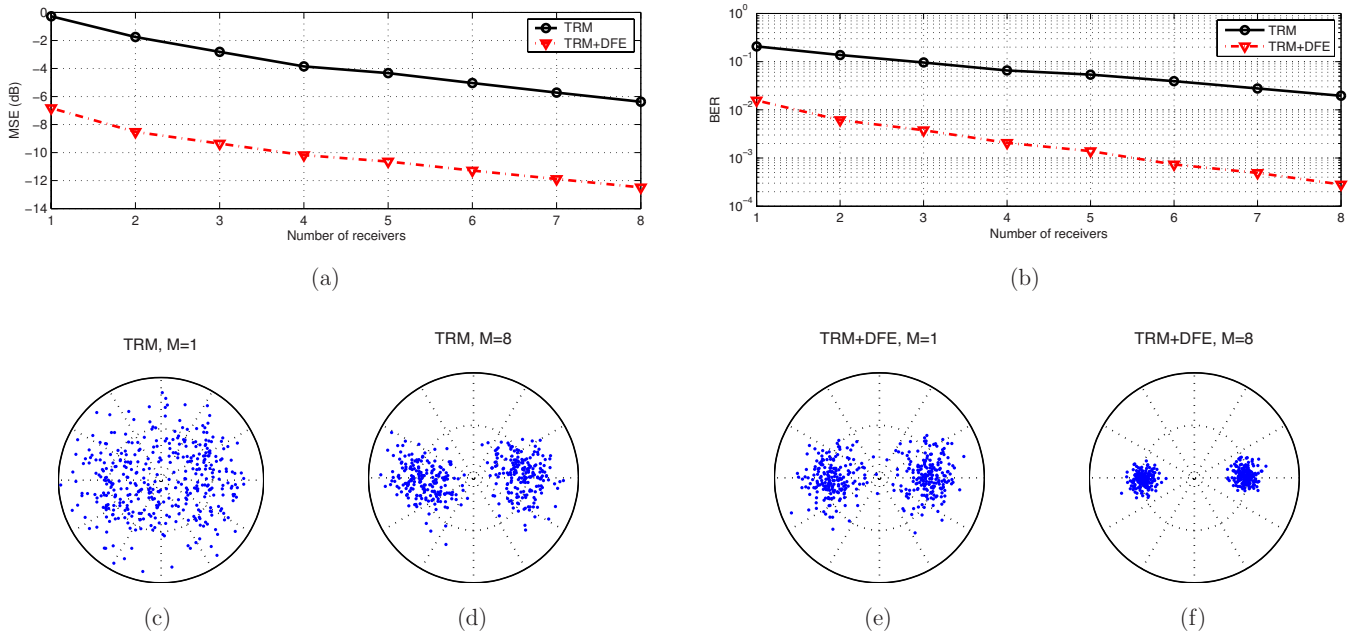


FIG. 10. (Color online) Impact of the number of sensors, M, selected from top, on the short-term performance of plain TRM and TRM with DFE postequalization [RLS, two-oversampling, (mostly) $(3,4)\times 2$ feedforward and 20 feedback coefficients, $\lambda=0.95$]. Probes estimated from packet preambles. (a) Output MSE, averaged over 200 packets. In this sparse TRM, saturation of MSE for sufficiently large M is not yet visible. (b) Average bit error rate. [(c) and (d)] Scatter plots for PKT 155, plain TRM, $M=1$ and 8. [(e) and (f)] Scatter plots for TRM+DFE.

lower performance gains once the thermocline is crossed, from $M=3$ onward, as received pulse shapes become more similar. Although not very clearly visible, the slope of the upper curve in Fig. 10(a) does decrease for $M \geqslant 4$. The fact that the inflection point is not located at $M=3$ could be due to the observed partial confinement of acoustic energy in the region below the thermocline, so that adding sensors for $M \geqslant 3$ provides additional signal power that improves performance beyond what would be expected by considering delay disparities alone.

5. Multichannel combining

The performance gains of proposed MC algorithms (MMAG and UMMSE) and DDPPC were quantified by comparing their MSEs with that of plain TRM in S, M, and L packet subintervals. The following results were obtained.

MMAG. The average improvements in MSE relative to plain TRM in the S, M, and L intervals for 200 baud (respectively, 400 baud) packets are 0.3, 0.5, and 0.8 dB (respectively 0.3, 0.5, and 1.0 dB). MSE gains increase as one progresses from S to M, and L, indicating that MMAG can partially compensate for the degradation in performance of plain TRM due to mismatch. Improvements of less than 1 dB are modest, but nonetheless these represent 10%-20% of the average MSE of plain TRM in the M and L regions.

UMMSE. MSE gains over TRM in the S, M, and L intervals for 200 baud (respectively, 400 baud) packets are 3.5, 4.0, and 4.3 dB (respectively 2.4, 2.9, and 3.5 dB). UMMSE yields larger MSE gains than MMAG and DDPPC, which is not surprising as this is an equalization-based scheme that minimizes a MSE-like performance metric.

DDPPC. S, M, and L MSE gains for 200 baud (respectively, 400 baud) packets are 1.0, 1.6, and 2.3 dB (respectively, −0.5, 0.5, and 1.4 dB). Channel tracking is seen to improve upon plain TRM under mismatch, but the MSE gains are actually smaller than those afforded by UMMSE and the number of parameters to track is substantially higher.

In MMAG $K=5\times10^{-2}$ is used in the loop filter update recursion (16) for each weight $w_m=e^{-j\theta_m}$. The same parameters for output phase synchronization and AGC used in plain TRM are also adopted in MMAG and DDPPC (the RLS-based UMMSE algorithm requires neither of them, and uses $\lambda=0.95$).

Similar to what was done for TRM and TRM+DFE in Sec. IV B 4, the self-degradation of the above algorithms was assessed by comparing their MSE in M/L intervals with S.

MMAG. The average MSE in M and L in 200 baud (respectively, 400 baud) packets increases by 0.4 and 0.7 dB (respectively, 0.8 and 1.3 dB) relative to S. The values are lower than those found for TRM, indicating that the algorithm is more robust to channel variations.

UMMSE. Similar comments apply to this algorithm, where MSE increases by 0.2 and 0.4 dB (200 baud) or 0.7 and 0.9 dB (400 baud).

DDPPC. Not surprisingly, DDPPC provides essentially constant performance throughout packets. The average degradations in MSE are −0.03 and −0.2 dB (200 baud) or 0.1 and 0.03 dB (400 baud).

Although MMAG and UMMSE have been proposed as low-complexity alternatives to DDPPC, note that DDPPC and UMMSE (or MMAG, for that matter) can be readily combined into a composite algorithm that will attain the best overall performance.

To conclude the analysis, QFs and ISI metrics were computed as in Sec. III A and plotted in Figs. 11(a), 11(f), and 11(k). Actually, these show the difference in ISI metric relative to plain TRM in the S, M, and L regions of individual packets, in line with what was done previously for MSE. Again, it can be concluded that these algorithms partially compensate for mismatch, yielding increasing gains over TRM as time progresses. This, too, can be seen in the scatter plots included in Fig. 11. Note also that, while UMMSE achieves lower MSE than DDPPC, its residual ISI is higher. This might be attributed to the nonequivalence of both performance metrics in the presence of noise (which was absent in the simulation reported in Sec. III A), in which case MMSE solutions are known to retain some residual ISI.[18] The observed inversion in relative performance of UMMSE and DDPPC in terms of MSE and ISI could also be due to inaccuracies in the method used to estimate the QFs. In fact, probe shapes were first estimated throughout each packet as described previously, snapshots taken every 100 symbol intervals, QF contributions computed according to Eq. (11), and then linearly combined by a set of weighting coefficients that reflect the intended behavior of the various algorithms, as described in Sec. III A. In the case of MMAG and UMMSE, discrepancies with respect to the actual coefficients computed according to Sec. III B may exist.

V. CONCLUSION

Experimental results were presented demonstrating demodulation of 200/400 baud PSK data collected during the MREA'04 sea trial in an eight-sensor receiver array. Several receiver architectures were examined and compared, namely, multichannel equalization, passive TRM with and without postequalization, DDPPC, and two multichannel combining methods (MMAG and UMMSE). The analysis included issues such as the characterization of time variability in channel responses and the impact on TRM performance of probe length, probe observation/estimation, and the number of array sensors.

MREA'04 data were collected with a moving source and drifting receiver, resulting in Doppler scaling of wave forms that was compensated by a simple resampling method. The technique was found to be very effective, generating nearly time-invariant equivalent channels where TR focusing often lasts for the full packet duration (20 s) with moderate degradation due to mismatch that does not strongly depend on the original Doppler distortion. Possibly, focusing would have been less stable at higher data rates, as reported in other TR experiments. Best TRM performance was obtained for truncated probes that retain multipath energy only on a few arrivals, suggesting that the remaining ones were less stable and could not be coherently combined to enhance the signal energy at the mirror output.

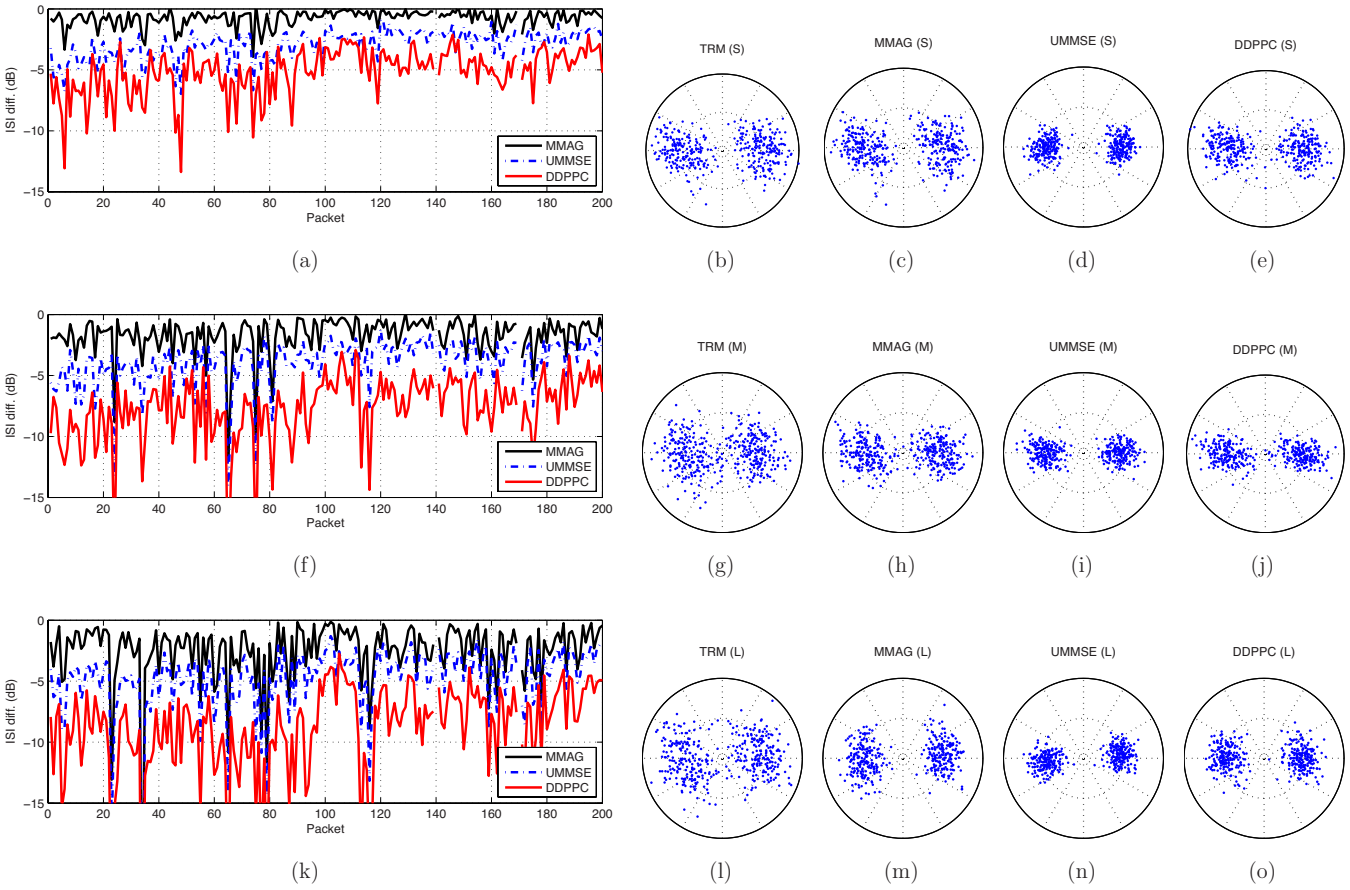


FIG. 11. (Color online) ISI reduction, on individual packets, of MC algorithms relative to plain TRM using eight hydrophones and estimated probes. Results are shown for MMAG, UMMSE, and DDPPC on short-term intervals $0.5 \leqslant t \leqslant 2.5$ s, medium-term intervals $4 \leqslant t \leqslant 6$ s, and long-term intervals $10 \leqslant t \leqslant 12$ s. Larger ISI gains as t increases indicate that the algorithms can compensate for some of the degradation in plain TRM due to environment mismatch. (a) Short-term ISI reduction. [(b)–(e)] Scatter plots for PKT 155 in short-term intervals for plain TRM, MMAG, UMMSE, and DDPPC. (f) Medium-term ISI reduction. [(g)–(j)] Scatter plots. (k) Long-term ISI reduction. [(l)–(o)] Scatter plots.

Plain TRM incurs a significant performance penalty relative to multichannel equalization, but postequalization was found to narrow the gap to about 1 dB. The latter presents a lower-complexity alternative to multichannel equalization, but the computational savings were small in MREA'04 data, where very short multichannel filters achieved the best performance. Again, the conclusion might be different at higher data rates.

The MMAG and UMMSE algorithms were proposed to compensate for (moderately) degraded focusing due to mismatch by adaptively weighting, rather than simply adding, sensor contributions. Unlike DDPPC, they do not require tracking of received pulse shapes, which makes them less computationally intensive. Temporal stability results for the various algorithms showed that MMAG and UMMSE can indeed improve upon plain TRM, although the performance is not as stable as in DDPPC. The average MSE improvement of MMAG was modest, partly due to the good stability of plain TRM itself in many of the packets. UMMSE is actually a multichannel equalization approach with a single coefficient per sensor that outperformed DDPPC in terms of output MSE and provided improvements of 3–4 dB over TRM.

This work considered MMAG and UMMSE as alternatives to DDPPC, but these can actually be combined to create multichannel receivers with adjustable pulse shapes and sensor weighting coefficients. Developing adaptation criteria for such structures, accounting for the effect of channel uncertainties, and establishing links with channel-estimation-based equalization are topics for future research.

ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the PIDDAC Program and project NUACE POSI/CPS/47824/2002. The authors would like to thank the NATO Undersea Research Centre for the organization of the MREA'04 sea trial, NRV Alliance master and personnel, and the scientist in charge Emanuel F. Coelho. Several helpful comments and suggestions by the anonymous reviewers are also gratefully acknowledged.

APPENDIX A: DOPPLER COMPENSATION

Assume first that a single propagation path with impulse response $g_m(t)$ exists between the source and the mth mirror sensor. Using Eq. (7), if the original baseband transmitted signal is denoted by $x(t)$, then from the receiver viewpoint the path is excited by $x((1+\beta)t)e^{j\omega_c\beta t}$ and the received signal is given by the convolution

$$y_m(t) = \int x((1+\beta)(t-\tau))e^{j\omega_c\beta(t-\tau)}g_m(\tau)d\tau$$
$$= \int x(t-\tau')\frac{e^{j[\omega_c\beta/(1+\beta)](t-\tau')}}{1+\beta}g_m\left(\frac{\tau'+\beta t}{1+\beta}\right)d\tau'. \quad \text{(A1)}$$

This is recognized as a linear convolution between $x(t)$ and a time-varying impulse response whose magnitude equals $|g_m(\tau)|$ for $|\beta| \ll 1$, with the origin shifted to $-\beta t$. In other words, the shape of this impulse response is essentially obtained by gradually sliding $g_m(\tau)$ along the τ axis as t progresses, such that the position of any feature in g_m traces a line with slope $-\beta$ in the (t,τ) plane. This property can be shown to remain valid even for an impulsive response $g_m(t)=g_m\delta(t-\tau_m)$.

In a multipath channel $h_m(t,\tau)$ is the sum of individual path contributions $h_{mp}(t,\tau)$, each having a different Doppler shift β_p. The analysis of experimental data reveals that most often all the β_p are approximately equal, and therefore all paths trace parallel trajectories (straight lines) in the (t,τ) plane [see Fig. 3(a)]. This happens when the propagation geometry and motion are predominantly horizontal, and suggests Doppler estimation algorithms based on identification of the time-variant impulse response and the common slope β of multipath arrivals. While an offline method of this sort using the Radon transform was indeed used to compute the Doppler scaling in MREA'04 packets, it should be emphasized that the topic of Doppler *estimation* is not central to this work and other more practical options are available. For example, a simple method has been proposed whereby the compression is estimated by detecting and measuring the delay between known prolog and epilog sequences in each packet.[34]

1. Resampling

Knowing the Doppler factor β and its theoretical effect on the complex envelope of the transmitted signal (7), we compensate it according to Eq. (9) by canceling the term $e^{j\omega_c\beta t}$ and then resampling to eliminate the time scaling. Resampling in discrete time can be performed in a number of ways, e.g., using low-complexity parabolic interpolation. In this work an efficient polyphase implementation (MATLAB resample function) was used for block resampling of full packets.

Naturally, the question arises as to whether Doppler compensation disrupts TR focusing by disturbing the medium transfer function $g_m(t)$. To address that issue we proceed as previously, expressing the resampled signal as a convolution between the ideal transmitted signal $x(t)$ and a time-varying impulse response. Using Eqs. (9) and (A1) and performing a change of variables yields

$$y'_m(t) = y_m\left(\frac{t}{1+\beta}\right)e^{-j\omega_c[\beta t/(1+\beta)]}$$
$$= \int x(t-\tau')\frac{g_m\left(\frac{\tau'}{1+\beta}\right)}{1+\beta}e^{-j\omega_c[\beta\tau'/(1+\beta)]}d\tau'. \quad \text{(A2)}$$

This is a convolution between $x(t)$ and a time-invariant impulse response that may be easily related to the original medium response in the time and frequency domains

$$g'_m(t) = \frac{g_m\left(\frac{t}{1+\beta}\right)}{1+\beta}e^{-j\omega_c[\beta t/(1+\beta)]}, \quad \text{(A3)}$$

$$G'_m(\omega) = G_m\left((1+\beta)\left(\omega+\frac{\omega_c\beta}{1+\beta}\right)\right). \quad \text{(A4)}$$

The spectrum (A4) is a frequency-shifted and (slightly) rescaled version of the original $G_m(\omega)$, so it seems reasonable to expect that focusing will be preserved. Because both the packet and probe undergo the same Doppler compensation procedure, the residual ISI is determined by the new medium autocorrelation function which, similar to Eq. (5), is given by

$$\gamma'(t) = \sum_{m=1}^{M} g'^{*}_m(-t) * g'_m(t)$$
$$= \sum_m \int g'^{*}_m(\tau-t)g'_m(\tau)d\tau$$
$$= \frac{\gamma\left(\frac{t}{1+\beta}\right)}{1+\beta}e^{-j\omega_c[\beta t/(1+\beta)]}. \quad \text{(A5)}$$

As in Eq. (A4), the Fourier transform of this function is a frequency-shifted and rescaled version of the original one,

$$\Gamma'(\omega) = \Gamma\left((1+\beta)\left(\omega+\frac{\omega_c\beta}{1+\beta}\right)\right). \quad \text{(A6)}$$

The (static) multipath compensation property of time reversal implies that $R(\omega)\Gamma(\omega) \propto R(\omega)$ in Eq. (5), so $\Gamma(\omega)$ is approximately flat in the signal band. In the MREA'04 experiment this band is at most $2r_{b\,\max}=800$ Hz for $r_b=400$ baud packets with 100% pulse rolloff. According to Eq. (8) $\beta_{\max}$ is about 10^{-3} for $v_{\max}\approx 1.5$ m/s and $c\approx 1.5\times 10^3$ m/s, hence the maximum Doppler shift is about $f_{d\,\max}=\omega_c\beta_{\max}/2\pi < 4$ Hz at the carrier frequency. Over the frequency band of interest $[(-r_{b\,\max};r_{b\,\max})=(-400;400)$ Hz in baseband] the behavior of $\Gamma'(\omega)$ is defined by the values of the original $\Gamma(\omega)$ in the interval $(-r_{b\,\max}(1-\beta_{\max})-f_{d\,\max};r_{b\,\max}(1+\beta_{\max})+f_{d\,\max})=(-403.2;404.0)$ Hz. It may be concluded that if $\Gamma(\omega)$ is flat in the band of $R(\omega)$, then the same will be true for $\Gamma'(\omega)$, with the possible exception of very narrow intervals at either the upper or lower edges of the signal band. With high probability time-reversed focusing will therefore be preserved by the proposed resampling method for Doppler compensation.

APPENDIX B: MMAG ADAPTATION RULE

The cost function (14) is to be iteratively maximized over the set of real angles θ_m, $m=1,\ldots,M$. Actually, only the gradient of Eq. (14) is needed to obtain an ascent iteration. To streamline the notation, the explicit dependence on the time instant n of the various sequences appearing in this section will be dropped. The gradient $\partial J_{\text{mag}}/\partial\theta_i$ is calculated by lumping together all terms that are independent of θ_i, viz.,

$$\begin{aligned} J_{\text{mag}} &= E\{|a_i - z_i e^{-j\theta_i}|^2\} \\ &= \sigma_{a_i}^2 + \sigma_{z_i}^2 - 2\,\text{Re}\{x_i e^{-j\theta_i}\} \\ &= \sigma_{a_i}^2 + \sigma_{z_i}^2 - 2\,\text{Re}\{x_i\}\cos\,\theta_i - 2\,\text{Im}\{x_i\}\sin\,\theta_i, \end{aligned} \tag{B1}$$

where

$$a_i = -z_1 - \sum_{m\neq i} z_m e^{-j\theta_m} = -z + z_i e^{-j\theta_i}, \tag{B2}$$

$$x_i = E\{a_i^* z_i\}, \quad \sigma_{a_i}^2 = E\{|a_i|^2\}, \quad \sigma_{z_i}^2 = E\{|z_i|^2\}. \tag{B3}$$

The gradient with respect to θ_i is now readily obtained as

$$\begin{aligned} \frac{\partial J_{\text{mag}}}{\partial\theta_i} &= 2\,\text{Re}\{x_i\}\sin\,\theta_i - 2\,\text{Im}\{x_i\}\cos\,\theta_i \\ &= 2\,\text{Im}\{E\{(z - z_i e^{-j\theta_i})^* z_i e^{-j\theta_i}\}\}. \end{aligned} \tag{B4}$$

A simple gradient ascent iteration is given by

$$\theta_i(n+1) = \theta_i(n) + K\frac{\hat{\partial} J_{\text{mag}}}{\partial\theta_i}, \tag{B5}$$

where a common stochastic approximation to the gradient is used, which simply amounts to ignoring the statistical expectation in Eq. (B4). The stochastic gradient in Eq. (B5) may be viewed as an error signal driving a simple PLL, whose loop filter may be refined to obtain more robust tracking behavior.[4] In this work, however, the performance of a first-order filter was found to be satisfactory.

APPENDIX C: SNR ESTIMATION

SNR estimation is a nonconsensual topic, and several different definitions and approaches are commonly used. Our method for estimating SNR with recursive channel identification parallels that of Flynn *et al.*[17] for block-based estimation. Specifically, we concentrate on an approach that tries to include reverberation in the total noise present during packet reception, as it cannot be coherently combined at the receiver to improve symbol estimates when using identification-based algorithms (equalization and TRM with probe estimation).

Let the samples of the mth estimated pulse shape $h_m(t)$ in Eq. (1), oversampled by a factor of L, be collected in vector $\mathbf{h}_m$. As described in Sec. IV B, the L polyphase components of h_m can be separately identified from a common training sequence and the polyphase components of the received signal y_m using a parallel bank of RLS filters, whose residual MSE is theoretically identical and will be denoted by $\sigma_{m\ \text{id}}^2$. The variance $\sigma_{m\ \text{id}}^2$ includes not only power from physical disturbances, $\sigma_{m\ d}^2$ (ambient noise and reverberation), but also excess MSE, ζ_m, due to misadjustment between actual channel coefficients and those estimated by the adaptive identification algorithm

$$\sigma_{m\ \text{id}}^2 = \sigma_{m\ d}^2 + \zeta_m, \quad \sigma_{m\ d}^2 \triangleq \sigma_{m\ \text{amb}}^2 + \sigma_{m\ \text{rev}}^2. \tag{C1}$$

Similarly, the expected value of the coefficient vector norm $\|\mathbf{h}_m\|^2$ (useful signal energy) exceeds the true norm by the trace of its covariance matrix. Approximate closed-form expressions are available for the variance of these estimation errors in several adaptive algorithms.[30] For RLS operating with forgetting factor λ and N coefficients that provide enough degrees of freedom to model the underlying system, the output MSE is[30]

$$\sigma_{m\ d}^2 + \zeta_m = \frac{2\sigma_{m\ d}^2 + \dfrac{\gamma_m^2}{1-\lambda}}{2-(1-\lambda)N}, \tag{C2}$$

where the constant γ_m is related to the degree of nonstationarity of the system. One possible approach for estimating σ_{md}^2 is to empirically evaluate the residual MSE $\sigma_{m\ \text{id}}^2$ for a range of values of λ and fit that curve to the right-hand side of Eq. (C2) to obtain the unknown σ_{md}^2 (and γ_m^2 as a by-product). Note that the theoretical derivation of Eq. (C2) assumes $\lambda\approx 1$ and $(1-\lambda)N\ll 1$, neither of which are very accurate for the values of λ and N where the lowest empirical σ_{md}^2 is obtained in MREA'04 data.

The average SINR is defined as[17]

$$\text{SINR} = \frac{1}{M}\sum_{m=1}^{M}\frac{\|\mathbf{h}_m\|^2}{\sigma_{m\ d}^2 L}. \tag{C3}$$

Technically, $\|\mathbf{h}_m\|^2$ in Eq. (C3) should be fine-tuned by subtracting the excess norm due to RLS adaptation. This was deemed unnecessary given the coarseness of approximations in the estimation of σ_{md}^2. In summary, the following steps were carried out to estimate SINR for each packet.

(1) Estimate the channel response in each array sensor as described in Sec. IV B for a range of RLS forgetting factors $\lambda_{\min}\leqslant\lambda_i\leqslant\lambda_{\max}$, $i=1,\ldots,F$.
(2) Compute the steady-state MSE $\sigma_{m\ \text{id}}^2$ at the filter output for each λ_i, averaging across polyphase components. Collect these in an $F\times 1$ vector $\mathbf{b}_m$.
(3) Build an $F\times 2$ matrix $\mathbf{A}$ whose ith row $\mathbf{a}_i$ equals

$$\mathbf{a}_i = \frac{1}{2-(1-\lambda_i)N}\begin{bmatrix} 2 & \dfrac{1}{1-\lambda_i}\end{bmatrix}, \tag{C4}$$

where N is the order of the identification filter per polyphase component.
(4) Solve $\mathbf{A}\mathbf{x}_m=\mathbf{b}_m$ in the least-squares sense to get $\mathbf{x}_m=[\sigma_{md}^2\ \gamma_m^2]^T$.
(5) Compute the SINR according to Eq. (C3), where $\mathbf{h}_m$ is the L-oversampled identified channel response for λ_i where the smallest identification residual $\sigma_{m\ \text{id}}^2$ was obtained.

Finally, note that throughout this work TRM performance is

expressed in terms of output MSE. Most authors who prefer to use a SNR_{out} metric[6,11,17] essentially adopt a simple sign change in decibel scale, $SNR_{out} \approx MSE^{-1}$.

[1]D. R. Jackson and D. R. Dowling, "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181 (1991).
[2]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).
[3]D. R. Dowling, "Acoustic pulse compression using passive phase-conjugate processing," J. Acoust. Soc. Am. **95**, 1450–1458 (1994).
[4]M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Reduced-complexity spatial and temporal processing of underwater acoustic communication signals," J. Acoust. Soc. Am. **98**, 961–972 (1995).
[5]M. Fink, "Time-reversed acoustics," Sci. Am. **281**(5), 91–97 (1999).
[6]M. Stojanovic, "Retrofocusing techniques for high rate acoustic communications," J. Acoust. Soc. Am. **117**, 1173–1185 (2005).
[7]G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[8]G. F. Edelmann, H. C. Song, S. Kim, W. S. Hodgkiss, W. A. Kuperman, and T. Akal, "Underwater acoustic communications using time reversal," IEEE J. Ocean. Eng. **30**, 852–864 (2005).
[9]D. Rouseff, D. R. Jackson, W. L. J. Fox, C. D. Jones, J. A. Ritcey, and D. R. Dowling, "Underwater acoustic communication by passive-phase conjugation: Theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[10]A. Silva, S. Jesus, J. Gomes, and V. Barroso, "Underwater acoustic communication using a 'virtual' electronic time-reversal mirror approach," Proceedings of the V European Conference on Underwater Acoustics (ECUA'00), edited by P. Chevret and M. E. Zakharia, Lyon, France, 2000.
[11]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, W. J. Higley, K. Raghukumar, T. Akal, and M. Stevenson, "Spatial diversity in passive time reversal communications," J. Acoust. Soc. Am. **120**, 2067–2076 (2006).
[12]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, M. Stevenson, and T. Akal, "Improvement of time-reversal communications using adaptive channel equalizers," IEEE J. Ocean. Eng. **31**, 487–496 (2006).
[13]D. Rouseff, "Intersymbol interference in underwater acoustic communications using time-reversal signal processing," J. Acoust. Soc. Am. **117**, 780–788 (2005).
[14]P. Hursky, M. B. Porter, M. Siderius, and V. K. McDonald, "Point-to-point underwater acoustic communications using spread-spectrum passive phase conjugation," J. Acoust. Soc. Am. **120**, 247–257 (2006).
[15]H. C. Song, P. Roux, W. S. Hodgkiss, W. A. Kuperman, T. Akal, and M. Stevenson, "Multiple-input-multiple-output coherent time reversal communications in a shallow-water acoustic channel," IEEE J. Ocean. Eng. **31**, 170–178 (2006).
[16]J. Gomes and V. Barroso, "Asymmetric underwater acoustic communication using a time-reversal mirror," Proceedings of MTS/IEEE OCEANS'00, Providence, RI, 2000, Vol. **3**, 1847–1851.
[17]J. A. Flynn, J. A. Ritcey, D. Rouseff, and W. L. J. Fox, "Multichannel equalization by decision-directed passive phase conjugation: Experimental results," IEEE J. Ocean. Eng. **29**, 824–836 (2004).
[18]J. G. Proakis, *Digital Communications*, 4th ed. (McGraw-Hill, New York, 2000).
[19]T. C. Yang, "Temporal resolution of time-reversal and passive phase-conjugation processing," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[20]L. J. Ziomek, *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing* (CRC, Boca Raton, FL, 1995).
[21]H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[22]S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal, "Robust time reversal focusing in the ocean," J. Acoust. Soc. Am. **114**, 145–157 (2003).
[23]H. Cox, "Navy applications of high frequency acoustics," Proceedings of the High-Frequency Ocean Acoustics Conference (HFOAC'04), La Jolla, CA, 2004, Vol. **728**, 449–455.
[24]P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio—Part I: Theoretical considerations," IEEE Trans. Commun. **40**, 885–894 (1992).
[25]M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," J. Acoust. Soc. Am. **94**, 1621–1631 (1993).
[26]A. Parvulescu, "Matched-signal ("MESS") processing by the ocean," J. Acoust. Soc. Am. **98**, 943–960 (1995).
[27]M. B. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**, 1349–1359 (1987).
[28]O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," IEEE Trans. Inf. Theory **36**, 312–321 (1990).
[29]J. Gomes, A. Silva, and S. Jesus, "Joint passive time reversal and multichannel equalization for underwater communications," Proceedings of MTS/IEEE OCEANS'06, Boston, MA, 2006, 1161–1166.
[30]A. H. Sayed, *Fundamentals of Adaptive Filtering* (Wiley-IEEE, New York, 2003).
[31]S. Jesus, C. Soares, P. Felisberto, A. Silva, L. Farinha, and C. Martins, "Acoustic maritime rapid environmental assessment during the MREA'04 sea trial," Technical Report No. 02/05, Centro de Investigacao Tecnologica do Algarve, Universidade do Algarve, 2005, URL: ftp://ftp.ualg.pt/users/sjesus/pubs/B21.pdf, accessed on 6/28/2008.
[32]S. Ariyavisitakul and L. J. Greenstein, "Reduced-complexity equalization techniques for broadband wireless channels," IEEE J. Sel. Areas Commun. **15**, 5–15 (1997).
[33]J. C. Preisig, "Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment," J. Acoust. Soc. Am. **118**, 263–278 (2002).
[34]B. S. Sharif, J. Neasham, O. R. Hinton, and A. E. Adams, "A computationally efficient doppler compensation system for underwater acoustic communications," IEEE J. Ocean. Eng. **25**, 52–61 (2000).

Sources of variability in distortion product otoacoustic emissions

Cassie A. Garner[a)]
The Department of Special Education and Communication Disorders, The University of Nebraska, 301 Barkley, Lincoln, Nebraska 68583

Stephen T. Neely and Michael P. Gorga
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 1 February 2008; revised 6 May 2008; accepted 6 May 2008)

The goal of this study was to determine the extent to which the variability seen in distortion product otoacoustic emissions (DPOAEs), among ears with normal hearing, could be accounted for. Several factors were selected for investigation, including behavioral threshold, differences in middle-ear transmission characteristics either in the forward or the reverse direction, and differences in contributions from the distortion and reflection sources. These variables were assessed after optimizing stimulus parameters for individual ears at each frequency. A multiple-linear regression was performed to identify whether the selected variables, either individually or in combination, explained significant portions of variability in DPOAE responses. Behavioral threshold at the f_2 frequency and behavioral threshold squared at that same frequency explained the largest amount of variability in DPOAE level, compared to the other variables. The combined model explained a small, but significant, amount of variance in DPOAE level at five frequencies. A large amount of residual variability remained, even at frequencies where the model accounted for significant amounts of variance. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2939126]

PACS number(s): 43.64.Jb, 43.64.Kc [BLM] Pages: 1054–1067

I. INTRODUCTION

It is not uncommon in audiological research and clinical-decision making that ears with thresholds of 20 dB hearing level (HL) or less are considered normal and, therefore, do not require services. Despite the fact that normal-hearing ears are often grouped into one category, recent research has indicated that there is variability in the measured level of distortion product otoacoustic emissions (DPOAEs) among normal-hearing ears (Dorn *et al.*, 1998) and that perhaps a more stringent definition of normal hearing should be used (Mills *et al.*, 2007). In this study, we sought to determine whether the following variables are possible reasons for the observed variability in DPOAE level among normal-hearing ears: (1) differences in behavioral thresholds within the generally accepted normal range of hearing, (2) differences in middle-ear (ME) transmission characteristics either in the forward or reverse direction, and (3) varying contributions from the two sources thought to contribute to the DPOAE response measured in the ear canal. These variables were studied for stimulus conditions that were individually optimized for each subject.

Outer hair cells are necessary for both normal behavioral thresholds and the production of DPOAEs. Even within the normal range of hearing thresholds, there is a trend for decreasing DPOAE level with increasing threshold. Dorn *et al.* (1998) compared audiometric threshold within the traditionally defined range of normal hearing and DPOAE level in response to a single set of primary levels (L_1/L_2 = 65/55 dB sound pressure level, SPL), and found that as thresholds increased from −5 to 20 dB HL, the DPOAE level decreased by about 8 dB, which is about one-fourth of the measurable range of DPOAE levels in humans. In other words, DPOAE level decreases gradually with increasing hearing threshold, even when the hearing threshold is within the normal range.

Typical stimulus parameters used to elicit DPOAEs include a primary-frequency ratio of about 1.22 and a fixed relationship between L_1 and L_2 that is applied, regardless of frequency, with all subjects. Research has shown that these stimulus parameters may not be optimal either for individual ears or individual frequencies (Whitehead *et al.*, 1995a; Whitehead *et al.*, 1995b; Neely *et al.*, 2005; Johnson *et al.*, 2006a). When comparing the DPOAE levels observed with the optimal path developed by Neely *et al.* to the results obtained with other stimulus paradigms (such as the scissor paradigm described by Kummer *et al.*, 1998), it was found that the DPOAE level was greater and the average standard deviation was less in the DPOAEs for a group of subjects with normal hearing when the individually optimized primary-level paths were used (Neely *et al.*, 2005).

Some variability in DPOAE level among normal-hearing ears may be due to differences in ME transmission characteristics. Differences in forward ME transmission could affect DPOAE levels by altering effective stimulus levels at the place of DPOAE generation within the cochlea. Differences in reverse ME transmission can cause differences in the measured otoacoustic emission (OAE) level in

[a)]Author to whom correspondence should be addressed. Electronic mail: cassie.garner@ttuhsc.edu

 0001-4966/2008/124(2)/1054/14/$23.00

the ear canal, independent of behavioral threshold (which depends only on forward energy transmission).

Shera and Guinan (1999) provided a descriptive framework for OAEs that is based on OAE generation rather than on the stimulus used to elicit the response. In their classification scheme, there are two backward-traveling waves produced during DPOAE measurements, one originating by nonlinear distortion and one by coherent reflection. The constructive and destructive interference of the two sources of DPOAEs can create fine structure, or minima and maxima, in the DPOAE response. Differences in fine structure (either in magnitude or in frequency location) may cause variability in normal DPOAE responses. Suppression of the reflection source contribution to the DPOAE would presumably reduce variations in DPOAE level across frequency (Mauermann and Kollmeier, 2004) and possibly the DPOAE level variation seen among normal-hearing ears.

Johnson *et al.* (2006b) varied suppressor levels in an attempt to determine which suppressor levels effectively suppress the reflection source without also affecting the overall DPOAE level. They found that there were several suppressor levels at each L_2 that could effectively eliminate the reflection-source contribution without reducing the overall level of the DPOAE; however, there was variability both in the effective suppressor level and in the evidence of fine structure across subjects and frequency. In fact, Johnson *et al.* (2006b) found that almost all subjects showed fine structure at 2000 Hz while very few demonstrated fine structure at 4000 Hz (Müller *et al.*, 2005). When combined with variations in the extent to which suppressor tones affected fine structure, it is not surprising that even under conditions in which efforts are made to suppress the reflection source, variability among normal ears persists.

The purpose of the present study was to determine whether a better understanding of variability in DPOAE levels among ears with normal hearing could be obtained. This was accomplished by studying the effects of behavioral threshold, ME transmission, and source contribution, alone and in combination, on the DPOAE level elicited with individually optimized stimulus conditions.

II. METHODS

Forty normal-hearing subjects participated in this study, 14 males and 26 females. Subjects ranged in age from 16 to 62 years. The better ear in each subject, based on screening audiometric and tympanometric results, was selected for data collection. Subjects were recruited primarily through the Human Research Subjects Core at Boys Town National Research Hospital and were paid for their participation. Each data-collection session lasted approximately 2 h and each subject returned for three to five sessions. To expand variability in the behavioral-threshold dimension, normal hearing was defined with a wider than usual range of thresholds, going from −10 to 25 dB HL, as determined audiometrically for pure tones ranging in frequency from 250 to 8000 Hz in $\frac{1}{2}$ octave steps. A 226-Hz tympanogram was measured with a clinical ME screening tympanometer (Grason-Stadler, Inc., GSI-37). A normal tympanometric assessment, defined as a peak pressure between −100 and +50 daPa and peak static acoustic admittance greater than or equal to 0.30 mmhos, was required prior to each data-collection session. This is a more liberal definition of normal ME function than the one used by Mills *et al.* (2007) in which they included subjects whose tympanometric peaks fell within ±25 daPa. Subjects were excluded if they reported any known history of ear disease.

All data were collected using custom-designed software that controlled a 24 bit sound card (CARDDELUXE, Digital Audio Labs) housed in a PC. An ER-10C (Etymotic Research) probe-microphone system was used to present stimuli and record responses from the ear. The ear tip was placed in the ear canal as deeply as possible in an effort to avoid or reduce standing-wave problems at frequencies of interest (Siegel and Hirohata, 1994; Whitehead *et al.*, 1995b).

Stimuli were calibrated *in situ* based on sound pressure level at the plane of the probe, using a chirp stimulus. A spectrum of the ratio of the input to the output was used for calibration. If a smooth spectrum was observed through the highest frequency at which experimental measurements were to be made, then testing was initiated on the experimental measures described below. If there were any notches present in the spectrum, the insertion depth of the ear tip was adjusted and the calibration was repeated. Experimental data were only collected if a smooth spectrum was observed or the notch was moved above the range of frequencies of interest in the present study. Even though an effort was made to remove notches from the spectrum, this cannot assure that all standing waves were eliminated. This calibration procedure was completed prior to collection of each data set. An additional probe-calibration procedure was required for the ME acoustic transfer-function (ATF) measurements. Probe calibration was performed once a day, prior to testing. The Thévenin acoustic-source characteristics of the probe microphone were estimated by a two-tube calibration procedure based on methods described by Keefe and Simmons (2003). Briefly, Thévenin's theorem is used to calculate the impedance across the ER-10C probe with two known load impedances. The values obtained can then be used to convert ear-canal measurements into ME transfer-function values in many different ears, which represent many different acoustic loads. The first tube is sufficiently long (295 cm) that the incident waveform (a click consisting of energy from 250 to 8000 Hz) can be measured without significant tube reflections. The second tube is sufficiently short (8.4 cm) that its response waveform contains multiple reflections and that the reflectance characteristics of the probe can be calculated. Each tube has an inner diameter of 0.794 cm, which is the approximate diameter of a typical adult ear canal, and is closed at one end with a 2-cm steel rod. The long and short tubes serve as the two known loads in the Thévenin calculation.

Custom-designed software was used to collect DPOAE data (EMAV, Neely and Liu, 1994). During DPOAE measurements, DPOAE levels were estimated at the $2f_1$-f_2 frequency. Noise-level estimates included $2f_1$-f_2 and several adjacent frequency bins to reduce variability in the estimate.

Data collection continued until noise estimates were below −25 dB SPL or 32 s of artifact-free averaging time was included in the measurement, whichever occurred first.

Data collection at each frequency was organized into sets, with behavioral threshold, optimized L_1, and controlled reflection source representing one set. An effort was made to collect a complete set of data in one session, with a single probe placement, in order to reduce variability in the measurements.

A. Auditory thresholds

After tympanometric and audiometric screening measures demonstrated that inclusion criteria were met, auditory thresholds were measured again using a more rigorous psychophysical procedure because of the presumed importance of behavioral thresholds in determining DPOAE level even within the normal range (Dorn *et al.*, 1998). The subject was seated in a comfortable recliner facing a computer monitor in a sound booth. Three buttons would light up consecutively on the monitor, representing two temporal intervals and the time after which a response from the subject was required. A tone was present in one of the two intervals and once the response button was lit, the subject was required to choose the interval in which the tone was present by clicking on the appropriate button with a mouse. The buttons were lit for 1000 ms each, with a 500 ms interval between presentations. One training session was completed to ensure that the subject understood the task.

Data were collected using custom-designed software that controls the sound card described above. The ER-10C probe-microphone system was used to present stimuli that were calibrated in the ear (in SPL at the plane of the probe, just as it was done prior to DPOAE measurements) prior to each test. A two-down, one-up method, targeting 71% correct performance, was used to define threshold. In implementing this procedure in the present study, the first four reversals used an 8-dB step size, after which the step size changed to 2 dB. Stimuli consisted of pure tones at nine f_2 frequencies (ranging from 500 to 8000 Hz in $\frac{1}{2}$ octave steps) and at the corresponding nine $2f_1$-f_2 frequencies (ranging from 340 to 5120 Hz) when $f_2/f_1 \approx 1.22$. Stimulus duration was 250 ms with 10 ms rise/fall time. The threshold at each frequency was measured three times, and an average of the two thresholds with the better standard error was taken as the final threshold value at that frequency.

B. Optimizing stimulus levels for individual ears

In an effort to reduce variability in DPOAE levels due to the relative level of the two primary tones, L_1 was optimized for each ear individually to produce the largest DPOAE response at each L_2. A Lissajous-path optimization method was used for f_2 frequencies of 1500–8000 Hz (Neely *et al.*, 2005). The L_1 and L_2 levels were each amplitude modulated such that the levels varied sinusoidally in decibels. L_2 varied from 17 to 69 dB SPL and L_1 varied from 43 to 76 dB SPL. All Lissajous-path stimuli had 23 cycles of L_1 modulation and 17 cycles of L_2 modulation in each half of a stimulus buffer.

For f_2 frequencies lower than 1500 Hz (500–1000 Hz), the signal-to-noise ratio (SNR) was often insufficient for the Lissajous-path method. Instead, optimization was performed with discrete primary levels distributed over a nonrectangular region of the L_1, L_2 space. L_2 levels were set to 40, 50, 60, and 70 dB SPL while L_1 varied from 10 dB below to 20 dB above L_2 (in 5 dB steps) for L_2=40, 50, and 60 and from 10 dB below to 10 dB above L_2 (in 5 dB steps) for L_2=70 dB SPL. The L_1 that produced the largest DPOAE level was determined for each L_2. Linear regression was then used to fit a line to these data points. This was considered the optimal linear path for the low f_2 frequencies; we used extrapolation to find the optimal L_1 for L_2 levels below 40 dB SPL.

The calculated optimal linear paths, individualized for each subject, were then used to set stimulus levels during DPOAE measurements. Following the exploration of the $L_1 \times L_2$ space, an optimal path input/output (I/O) function was measured with EMAV (Neely and Liu, 1994). L_2 was varied from −15 to 80 dB SPL, with L_1 varying according to each individual's optimal linear path. EMAV was chosen for the main data-collection efforts because it has several special features, including measurement-based stopping rules, ability to measure signal and noise in the same frequency bin, and capability of presenting a suppressor to reduce reflection-source contributions (see below).

C. Controlling the contributions from the reflection source

The influence of source contribution was investigated at three of the nine f_2 frequencies: 1000, 2000, and 4000 Hz. At these three f_2 frequencies, a second I/O function was collected with an additional suppressor tone. Two of these three frequencies were selected because optimal suppressor conditions were described at those frequencies (2000 and 4000 Hz, Johnson *et al.*, 2006b). In addition, previous data have shown that DPOAE behavior, at least in terms of fine structure, may be different at 2000 and 4000 Hz, with fine structure being evident in essentially all normal-hearing subjects at 2000 Hz and seldom present at 4000 Hz (e.g., Müller *et al.*, 2005; Johnson *et al.*, 2006b; Wagner *et al.*, 2007). An f_2 of 1000 Hz was also used, since it appears that fine structure is more likely to be observed at lower f_2's such as 1000 Hz (Dhar and Shaffer, 2004).

The suppressor tone (f_3) was played 16 Hz below $2f_1$-f_2 and tracked L_2 in level according to the equation $y=49.5+0.285x$ at 1000 Hz, $y=47+0.33x$ at 2000 Hz, and $y=42+0.42x$ at 4000 Hz (Johnson *et al.*, 2006b), where y =suppressor tone level (L_3) and $x=L_2$. L_2 varied from −15 to 80 dB SPL with L_1 varying according to each ear's optimal path. Suppressor levels were chosen according to line-fit equations at each f_2 that optimized the suppressor effect of f_3 while minimizing the effect on overall DPOAE level (Johnson *et al.*, 2006b). Using the above equations, L_3 varied from 45 to 72 dB SPL at 1000 Hz, from 42 to 73 dB SPL at 2000 Hz, and from 27 to 75 dB SPL at 4000 Hz.

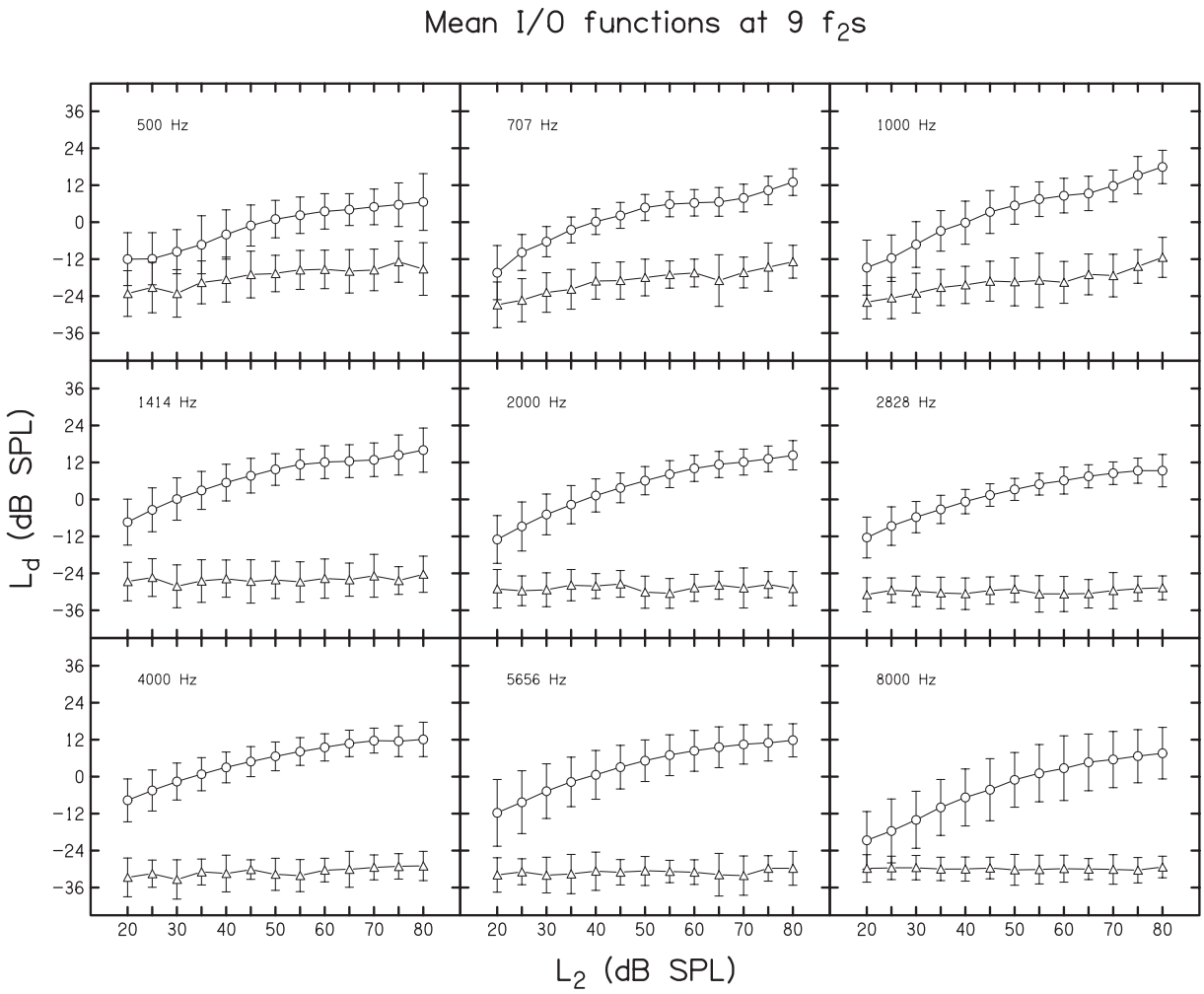


FIG. 1. Mean DPOAE (circles) and noise (triangles) levels as a function of L_2 for each of the nine f_2 frequencies. Error bars represent ±1 SD.

D. Middle-ear energy transfer

In order to assess DPOAE variability due to ME transfer, a wideband ATF was measured at ambient static pressure (Keefe and Simmons, 2003). The ATF measurement was used to provide an estimate of the forward- and reverse-transmission characteristics of the ME and was related to DPOAE level and behavioral threshold.

The ME ATFs were collected using custom-designed software (REFLWIN, Keefe, 2000) that controls the same sound card that was used for the other measurements. The input wideband stimulus used in the ATF test was an electrical signal designed to produce a short-duration acoustic "click" with energy from 250 to 8000 Hz. Thirty-two buffers were collected and averaged to give the final response, which corresponds to a test time of approximately 1.5 s. Analysis of the ATF was completed with 1/24 octave precision, providing a sufficient number of data points necessary for comparison with data from the other measurements.

E. Data analysis

In order to determine whether a combination of the above-mentioned factors was responsible for the variability seen in DPOAE responses, multiple-linear regression (MLR) analyses were performed. Select variables were used in the MLR, taking into account all of the independent variables (behavioral threshold, ME transmission, and contribution from the reflection source). The MLR attempted to model the relationship between the independent variables and the dependent variable by fitting a linear equation to the observed data.

III. RESULTS

Figure 1 displays mean DPOAE level as a function of L_2 with error bars representing ±1 standard deviation (SD) for each of the test frequencies. Also shown in this figure is the mean noise level ±1 SD. When performing a linear regression between DPOAE level and L_2 at f_2=4000 Hz, a significant model was observed ($f_{1,1038}$=1312, $p<0.0005$). The adjusted R^2=0.56, indicating that over half of the variance in DPOAE level at 4000 Hz could be accounted for by L_2. This trend was observed at all other test frequencies as well.

While L_2 accounts for approximately half of the variability, there remains variability across subjects at a fixed L_2 or the variability independent of L_2, which is the focus of this study. We needed to remove the L_2 dependence in order to examine the effect of behavioral threshold, ME transmission, and reflection source on DPOAE level, without the effect of L_2. Within the range of L_2 levels from 20 to 80 dB SPL, the

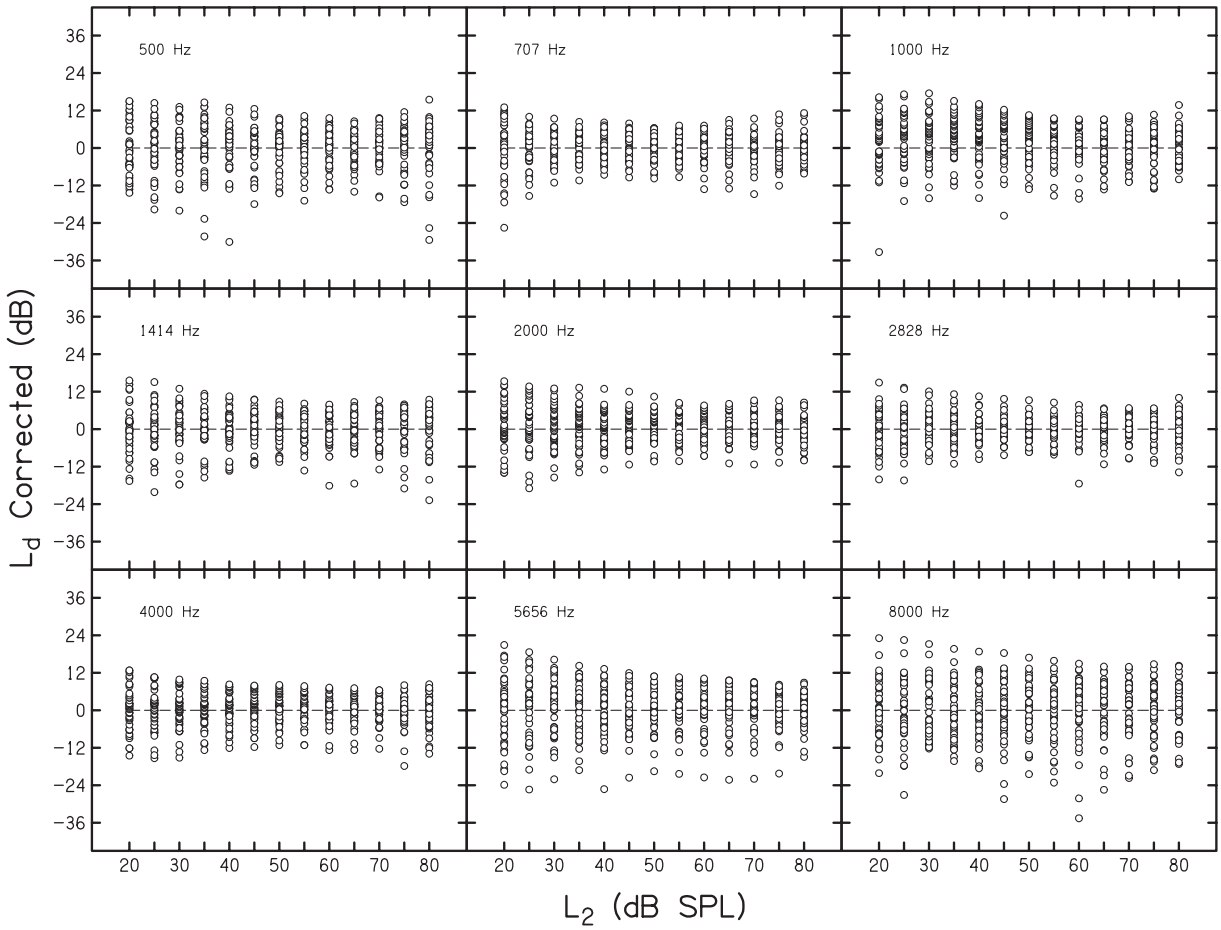


FIG. 2. L_d corrected plotted as a function of L_2 at each of nine f_2 frequencies. Each panel includes estimates of L_d from 40 subjects at each of 13 L_2 levels, for a total of 520 points per panel. The dashed line at L_d corrected=0 is the mean residual variance and provides a point of reference.

mean DPOAE levels for each L_2, based on data from all 40 subjects, were subtracted from each subject's individual DPOAE level at the corresponding L_2 for each of the nine f_2 frequencies. The new values of DPOAE level for each subject are referred to as L_d corrected. Figure 2 depicts L_d corrected as a function of L_2 with the mean represented as a dashed line at zero. It is the residual variability, surrounding L_d corrected equal to 0, that was used in further analyses to assess the effect of behavioral threshold, ME transmission, and source contribution on normal DPOAE variability.

In an effort to reduce the number of independent variables in the final analysis, a separate MLR was completed at each of the nine f_2 frequencies. Two variables representing forward transmission and two representing reverse transmission through the ME were chosen to be included in the final analysis because they accounted for the most variability and were significant at a majority of the nine f_2 frequencies. The variables that were chosen as estimates of forward transmission were equivalent reflectance and admittance magnitude at the f_2 frequency. Equivalent reflectance and equivalent volume at the f_d frequency were chosen as estimates related to ME transmission in the reverse direction.

One additional variable was included in the regression at 1000, 2000, and 4000 Hz; DPOAE level was measured with or without a suppressor slightly lower in frequency than $2f_1$-f_2. These six explanatory variables (seven at 1000, 2000, and 4000 Hz) were included in the final MLR. The coefficients used to describe the relationship between the individual variables and L_d corrected are taken from the combined model from the final multiple-linear regression. While including the behavioral threshold and middle-ear variables in the final analysis together may introduce colinearity issues, the method was chosen in order to provide a complete model that accounts for the most variance possible while simplifying interpretation of the results by maintaining consistency across frequency.

A. Behavioral threshold

In Fig. 3, L_d corrected is plotted as a function of behavioral threshold at all nine f_2 frequencies, and a quadratic function was fitted to the data. A quadratic function was chosen due to the fact that curvature has been observed in plots of DPOAE level as a function of audiometric threshold (Dorn *et al.*, 1998). Behavioral threshold accounted for a significant amount of the variability at only three of nine f_2 frequencies, 500, 2000, and 8000 Hz. Similarly, behavioral threshold squared accounted for significant amounts of variability only at 500, 4000, and 5656 Hz. Behavioral threshold squared was included in the analysis in order to fit the data with a quadratic function.

At 500 Hz, where both behavioral threshold and behavioral threshold squared are significant, there is a trend for DPOAE level to decrease as behavioral threshold increases.

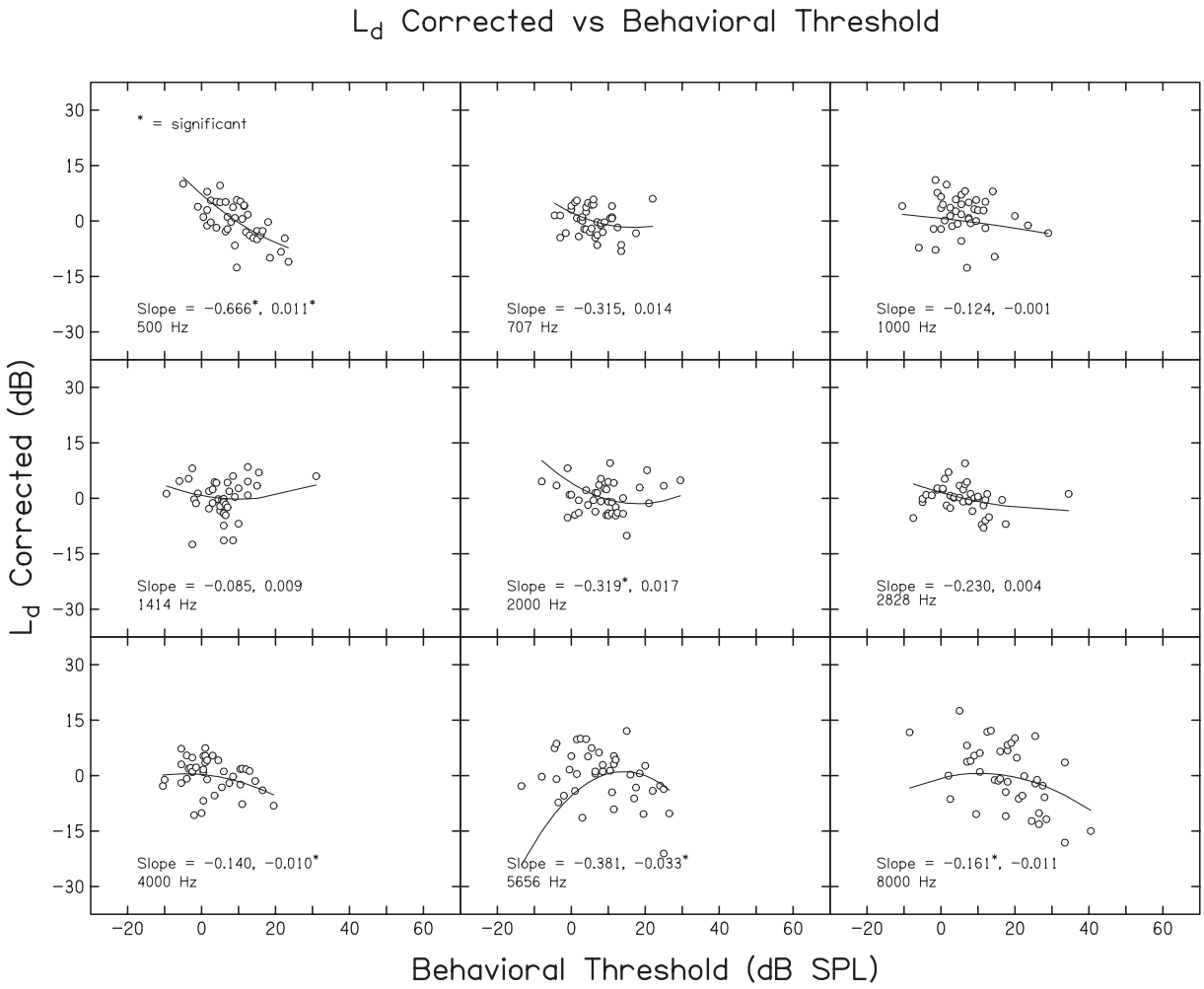


FIG. 3. L_d corrected as a function of behavioral threshold, with data at nine f_2 frequencies shown separately in each panel. The line in each panel represents the quadratic function that was fitted to the data. The slope values represent the unstandardized coefficients for behavioral threshold and behavioral threshold squared, respectively. An asterisk indicates a significant relationship.

A similar trend is observed at 2000, 4000, and 8000 Hz, but not as strongly as at 500 Hz. The scatter plots (see Fig. 3) and accompanying indications of significance (asterisk in panels for which significant effects were observed) show a lack of correlation between DPOAE level and behavioral threshold at other frequencies (707, 1000, 1414, 2828, and 5656 Hz). These results indicate that differences in the status of the cochlea, as reflected by behavioral-threshold measures, do not account for the variability seen in DPOAE levels among normal-hearing ears, at least among the present sample of subjects. In fact, in some cases, the slope of the function is positive, although not significant, which is opposite to what we predicted at the outset of this study.

B. Middle-ear energy transfer

Figure 4 displays L_d corrected as a function of energy reflectance at the f_2 frequency. Energy reflectance accounted for a significant amount of the variability in DPOAE level at only one of nine f_2 frequencies, 4000 Hz. That is, this measure of ME transmission did not account for any of the remaining variability in DPOAE level once the effects of L_2 were removed. Prior to data collection, we hypothesized that DPOAE level would decrease as energy reflectance increased. This trend was seen slightly at 500, 2000, 4000, and 8000 Hz, but the unexpected pattern of positive slopes was noted at the remaining frequencies.

Admittance magnitude at the f_2 frequency, another measure used to describe forward ME transmission, accounted for a significant amount of variability in DPOAEs at only three of nine f_2 frequencies, 4000, 5656, and 8000 Hz. These trends are shown in Fig. 5, which displays L_d corrected as a function of admittance magnitude at the f_2 frequency. The scatter plots and accompanying indications of significance (asterisk in panels for which significant effects were observed) do not reveal a correlation between DPOAE level and admittance magnitude at the low- and midfrequencies, with significant effects only at the three highest frequencies. Even at these frequencies, results were inconsistent. The L_d corrected decreased as admittance magnitude increased at 4000 and 8000 Hz (an effect opposite to our prediction), whereas the reverse was true at 5656 Hz. To assess the effect of reverse ME transmission on the variability in DPOAE level, the resulting coefficients from the MLRs for energy reflectance and equivalent volume at the f_d frequency were examined. Energy reflectance at the f_d frequency did not account for a significant amount of variability in DPOAE level at any frequency. Figure 6 displays L_d corrected as a

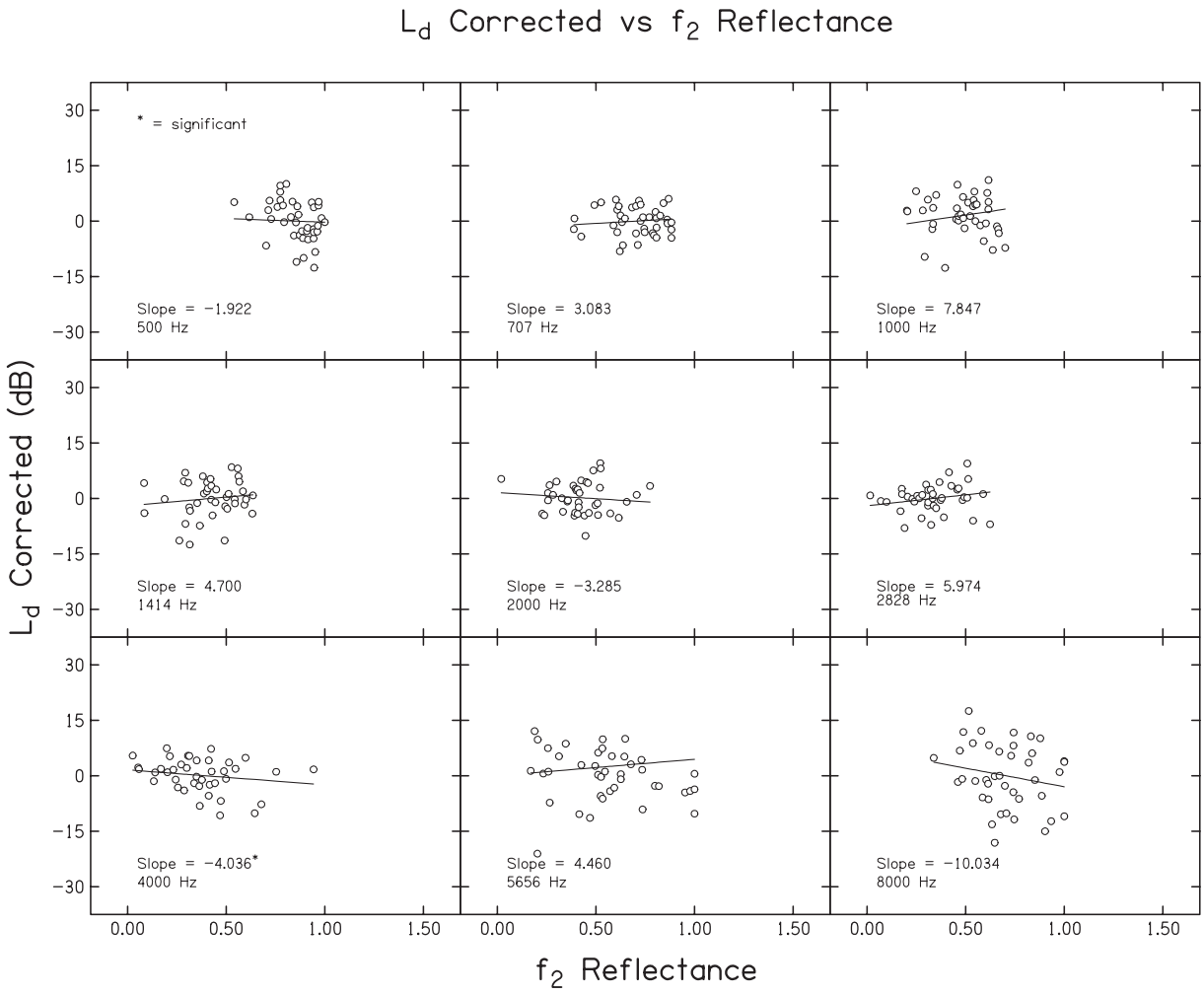


FIG. 4. L_d corrected as a function of energy reflectance at the f_2 frequency, with data at nine f_2 frequencies shown in each panel. The line in each panel represents a linear fit to the data. The slope values represent the unstandardized coefficients for energy reflectance at the f_2 frequency. The unstandardized coefficients were examined for all variables because not all of the independent variables share the same variance and they are based on the original, not the standardized scores of the predictor. An asterisk indicates a significant relationship.

function of energy reflectance at the f_d frequency. Eight of nine f_2 frequencies displayed a negative relationship. Equivalent volume at the f_d frequency, another factor that was used to estimate reverse ME transmission characteristics, versus L_d corrected is displayed in Fig. 7. Four of the nine test frequencies displayed the expected negative relationship between L_d corrected and equivalent volume, but the effect was significant at only two frequencies (1000 and 4000 Hz).

It appears that ME variables, presumably representing the forward or reverse transmission through the ME, account for only a small amount of the variability in DPOAE level and only at a few frequencies.

C. Source contribution

DPOAEs were measured with and without an "optimal" suppressor when f_2=1000, 2000, and 4000 Hz. The results are plotted in Fig. 8 in the form of DPOAE I/O functions. The solid and dashed lines represent the DPOAE level measured without and with a suppressor, respectively. DPOAE levels with and without a suppressor at 2000 and 4000 Hz show small differences at low levels, which is consistent with the view that relative contributions from distortion and reflection sources are more similar at low levels. Thus, one would expect that removal of the reflection source would be more evident for low-level stimulus conditions, compared to responses for higher level conditions. The results observed at 2000 and 4000 Hz are consistent with this expectation.

Larger differences exist between the suppressed and unsuppressed conditions, and across a broader range of levels, at 1000 Hz. The presentation of a suppressor tone whose frequency was slightly lower than $2f_1$-f_2 resulted in a reduction in mean DPOAE level essentially for all stimulus conditions. This result is attributed to the reduction in the reflection-source contribution, which, in the average data, had a positive impact on the DPOAE level measured in the ear canal when no suppressor was present. In fact, this is the only f_2, of three tested, where the suppressor variable accounted for a significant amount of the variability in DPOAE level.

D. Combined model

After examining the relationships between the individual variables and L_d corrected, we next evaluated the extent to

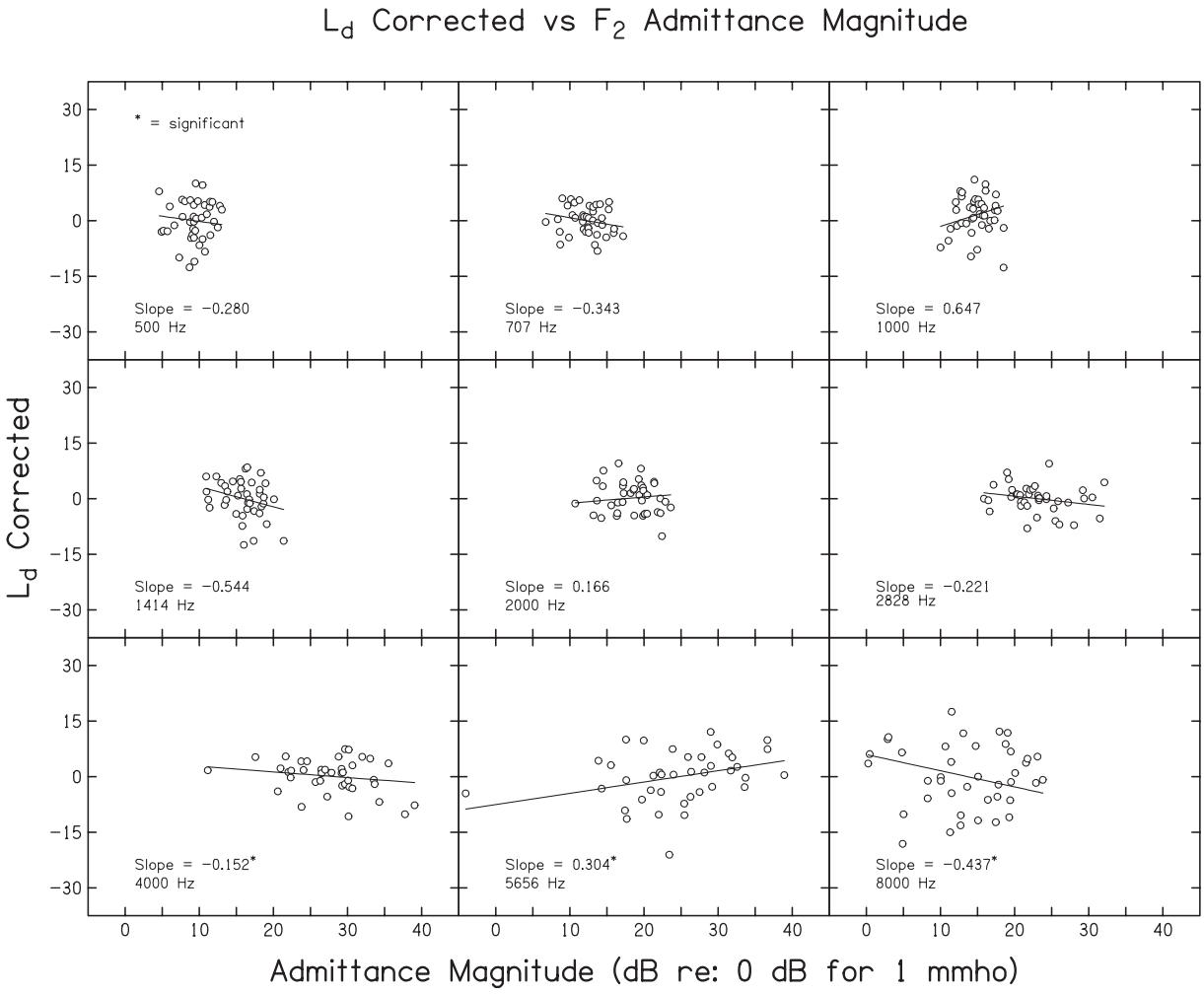


FIG. 5. L_d corrected as a function of admittance magnitude at the f_2 frequency, with data at nine f_2 frequencies shown separately in each panel. The line in each panel represents a linear fit to the data and the slope values represent the unstandardized coefficients for admittance magnitude at the f_2 frequency. An asterisk indicates a significant relationship.

which the combined model accounted for variance in DPOAE level. MLR was completed in an attempt to model the relationship between the independent variables which included behavioral threshold at the f_2 frequency and the same behavioral threshold squared, energy reflectance and admittance magnitude at the f_2 frequency, energy reflectance and equivalent volume at the f_d frequency, and DPOAE level measured with or without a suppressor slightly lower in frequency than $2f_1$-f_2. L_d corrected served as the dependent variable. Up to this point, the individual independent variables and their relationship to L_d corrected have been presented, even though the information was based on observations from the combined model. Here we combine these same variables and determine, in total, how much variance was accounted for when all factors were considered simultaneously.

Figure 9 plots normalized slope values for each of the independent variables as a function of f_2. After analyzing the data, it was noted that behavioral threshold at the f_2 frequency and the same behavioral threshold squared accounted for more of the variability in L_d corrected than the other independent variables. Thus, results for behavioral threshold (which provides an indication of the integrity of the cochlea) are plotted in the top panel of Fig. 9. Next, we added the ME variables to the analysis to determine if more of the variance could be accounted for. The second panel represents results for variables that presumably relate to forward ME transmission (energy reflectance and admittance magnitude at the f_2 frequency). The third panel represents results for variables that were used as estimates related to reverse ME transmission (energy reflectance and equivalent volume as a function of f_d). Finally, at 1000, 2000, and 4000 Hz, the contribution from the reflection source was taken into account to assess the extent to which the addition of that variable accounted for more variability in L_d corrected, which is represented in the fourth panel. Filled symbols indicate those variables that accounted for a significant amount of variability in L_d corrected.

An equation was derived at each f_2 frequency using the coefficients and constants from the MLR to predict DPOAE level. Table I provides a summary of the combined models at each frequency. The table provides a constant, R^2, p-values, and the residual or variance left unaccounted for after taking into account all of the independent variables examined in this study. The combined model accounted for a significant, although small, amount of variance in L_d corrected at five fre-

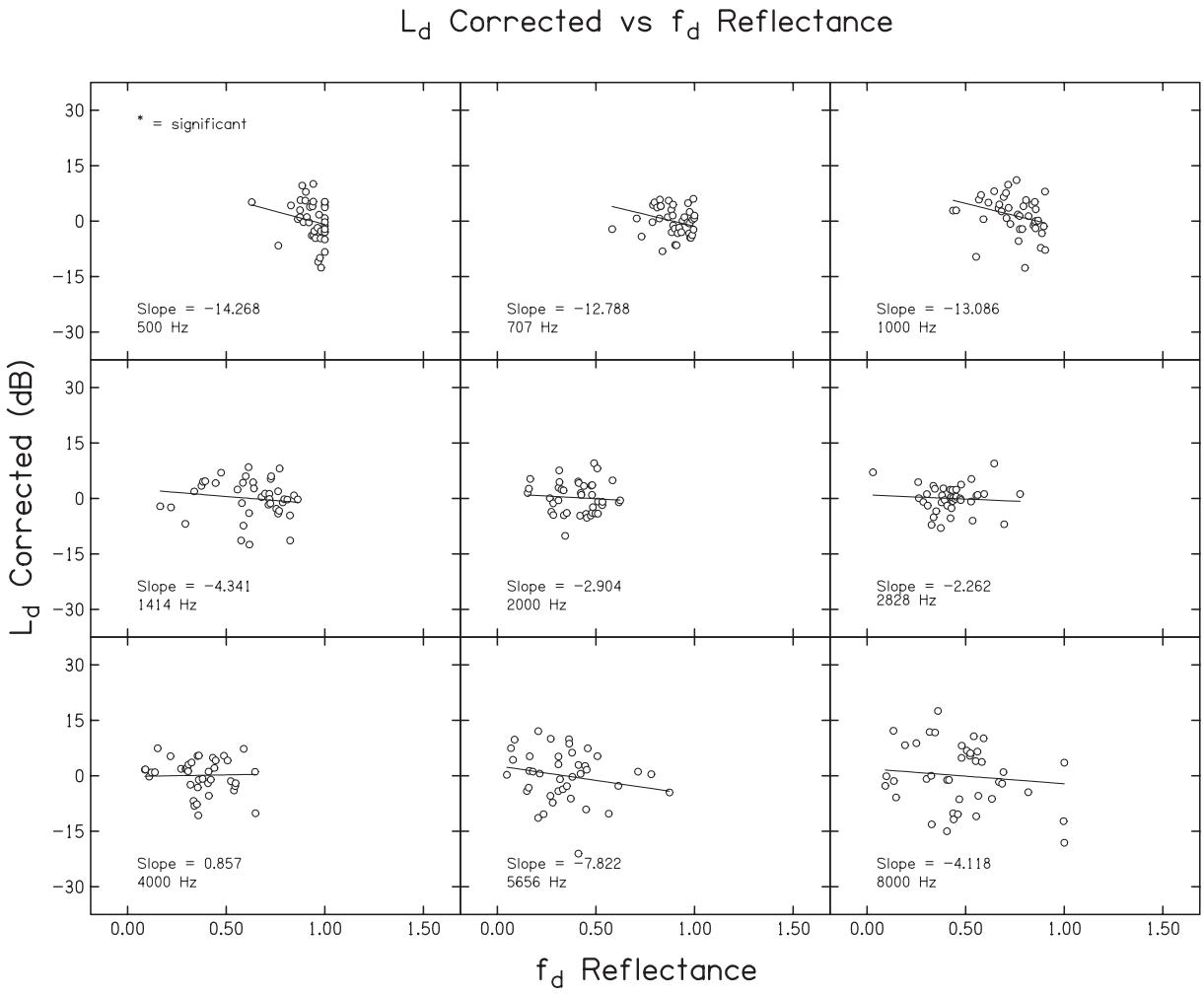


FIG. 6. L_d corrected as a function of energy reflectance at the f_d frequency, with data at nine f_2 frequencies shown separately in each panel. The line in each panel represents a linear fit to the data and the slope values represent the unstandardized coefficients for energy reflectance at the f_d frequency. An asterisk indicates a significant relationship.

quencies, or more than half of the frequencies studied. In all of the individual variable cases, significant amounts of variance were accounted for at fewer than five frequencies. The observations associated with the combined model, therefore, indicate that, in combination, the variables account for more of the variance. There still is a large amount of residual variability, or variability that is left unaccounted for, even for those frequencies for which the model was able to account for significant amounts of variance.

IV. DISCUSSION

The underlying aim of this study was to understand the sources of variability in DPOAE level among normal-hearing ears, which might lead to ways to reduce it. In turn, it was hoped that reducing variability among ears with normal hearing might result in greater separation between the distributions of responses from normal and impaired ears and thus improve the accuracy with which auditory status can be determined from DPOAE measurements.

In examining the mean I/O functions, the frequencies that displayed the most variability were 500, 1000, 5656, and 8000 Hz. The increased variability for low frequencies could be due to the fact that noise increases as frequency decreases while greater variability for both low and high frequencies could be a consequence of middle-ear transmission characteristics at these frequencies. The noise levels are higher at lower f_2 frequencies (especially 500 Hz), making it difficult to separate the signal from the noise. The ME transmission characteristics are such that less energy is passed through the ME at the lower and higher frequencies, potentially affecting measurements at 500, 1000, 5656, and 8000 Hz. Another reason for the variability seen at 8000 Hz could be due to increased system distortion at that frequency. Although system distortion measurements were not made as part of this study, they are routinely measured in the laboratory in which the current study was completed. The procedures used in measuring system distortion have been described in a previous report (Dorn *et al.*, 2001). While it is unclear how increased distortion would result in increased variability, it is the case that system distortion would reduce the reliability of the measurements at 8000 Hz.

Variability in the I/O functions was also seen across primary level with the greatest variability observed at the lowest and the highest primary levels. The increased variability at the lowest levels is due to the fact that DPOAE and noise levels are more similar when using low stimulus levels, with

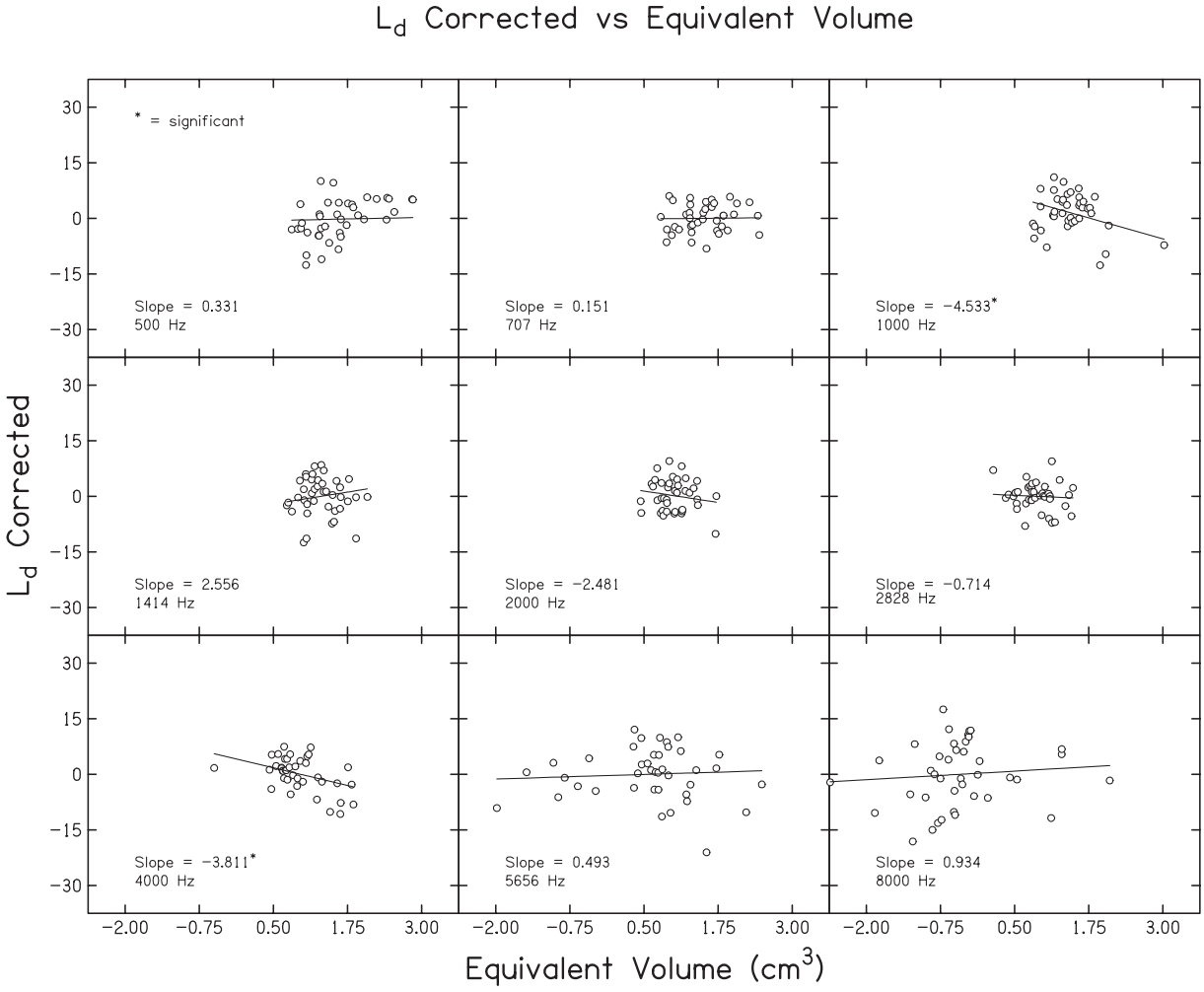


FIG. 7. L_d corrected as a function of equivalent volume at the f_d frequency, with data at nine f_2 frequencies shown separately in each panel. Following the same conventions as in Fig. 5, the line in each panel represents a linear fit to the data and the slope values represent the unstandardized coefficients for equivalent volume at the f_d frequency. An asterisk indicates a significant relationship.

the variable noise level exerting greater relative influence on measured response level. Measurement-based stopping rules were used in an effort to achieve similar response quality across stimulus conditions and across subjects, which were designed to reduce the noise to a constant level across conditions. For a constant noise level, however, DPOAE variability will increase as DPOAE level decreases, which is what happens because DPOAE level would be more similar to noise level. In contrast when the response levels are higher (which is typically the case at high stimulus levels), variability is less if the noise floor remains constant.

In any case, DPOAE level is highly dependent on stimulus level. Since the majority of variability seen in DPOAE level is due to primary level, it was necessary to factor out L_2 so that other sources of variability could be evaluated. The effects of L_2 were removed by subtracting the mean I/O function at each frequency from individual I/O functions. Subsequent analyses of these "corrected" DPOAE I/O functions revealed that no single variable accounted for large amounts of the remaining variability seen in DPOAE level. When the variables were evaluated in combination, the extent to which significant effects were observed depended on frequency. In the combined case, the most variability was accounted for at 500 and 8000 Hz, with 35% and 33% accounted for, respectively. The R^2 values were smaller at all other frequencies ranging from 0% to 32%.

A. Behavioral threshold

Behavioral threshold accounted for the most variability in DPOAE level, out of all of the individual variables evaluated, but significant relationships were observed at only five of nine frequencies. Although a negative relationship was expected between behavioral threshold and DPOAE level (a decrease in DPOAE level as behavioral threshold increased), positive relationships were also seen at some frequencies. Decreases in DPOAE level with increasing threshold can be related to the codependency of both threshold and DPOAE level on the status of the outer hair cells. An underlying mechanism for the inverse of this relationship, however, is not obvious.

Dorn *et al.* (1998) compared audiometric threshold (within the traditionally defined range of normal hearing) and DPOAE level, and found that as thresholds increased from −5 to 20 dB HL, the DPOAE level decreased by about 8 dB, which is about one-fourth of the measurable range of

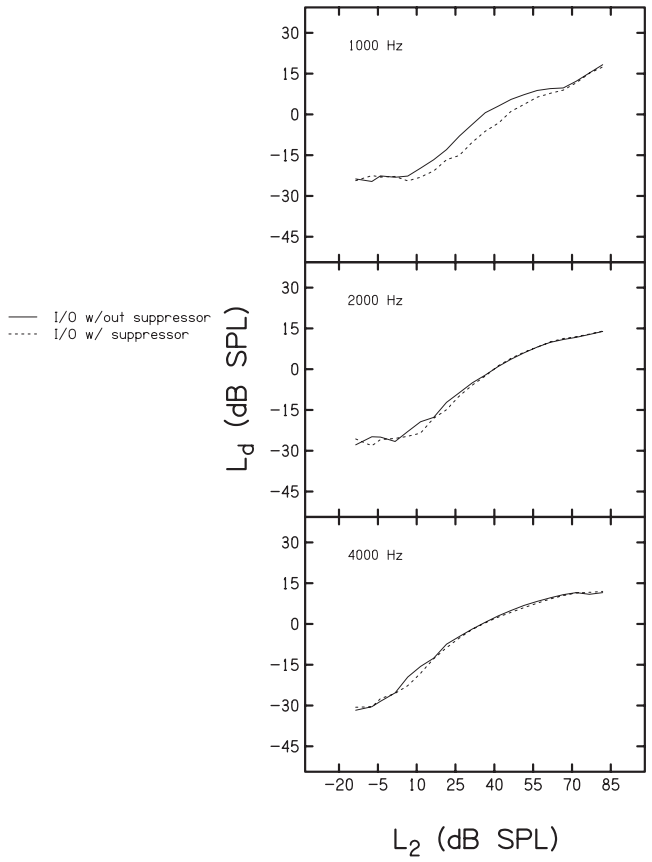


FIG. 8. DPOAE level as a function of L_2 for 1000, 2000, and 4000 Hz plotted separately in each panel. The solid line represents the DPOAE collected without a suppressor and the dashed line represents the DPOAE collected with a suppressor tone present.

DPOAE levels in humans. The range of normal thresholds for subjects included in the present study was −10 to 25 dB HL, leading us to expect that an even larger range of DPOAE levels would be observed. DPOAE level decreased by about 18 dB across that threshold range at 500 Hz, but that was the largest effect observed. At other frequencies, there was much less dependency of DPOAE level on threshold. Even though normal thresholds were broadly defined in the present study, it is possible that an even wider range may have resulted in significant findings. However, the results reported by Dorn *et al.* (1998), in which effects of behavioral threshold were seen over a smaller range of thresholds compared to the present case, are in conflict with the present findings and indicate that a wider range of thresholds is not necessary to see an effect. An alternative explanation for the lack of significance in the relation between behavioral threshold and DPOAE level may be the difference in the number of ears examined in the present study (40 ears), compared to the much larger number of ears included in the study of Dorn *et al.*. It is possible that the larger *N* resulted in more reliable estimates, leading to the observation of a significant relationship between threshold and DPOAE level in the normal-per-frequency group (audiometric status was treated on an individual frequency basis), and the smaller *N* in this study and the normal-across-frequency group (subjects with audiometric thresholds of 20 dB HL or better at all octave and half-octave frequencies from 250 to 8000 Hz) in the previous study limited the ability to see a significant relationship. The statistical results from Dorn *et al.*, therefore, are consistent with the hypothesis that the lack of significance in the present study may be due, in part, to the size of our sample.

Despite this, Mills *et al.* (2007) found that, even within a 10 dB HL or better threshold range (a smaller range than in the study of Dorn *et al.*), DPOAE responses showed a trend for increased emission thresholds with increased behavioral threshold. However, the relationship between emission threshold and behavioral threshold was significant only at 8000 Hz with an R^2 value of 0.27. The study of Mills *et al.* included 40 ears, the same number of ears in the present project.

Another possible reason that the effects of behavioral threshold were not more pronounced in the present study may relate to differences in stimulus conditions across studies. We optimized stimulus levels for each individual ear, and this optimization was performed for all subjects, including those with thresholds on the high end of the range included in the present study. By optimizing stimulus level on a subject-by-subject basis, it is possible that the present stimulus conditions mitigated the effects of threshold elevation. This individualized optimization was not utilized by either Dorn *et al.* (1998) or Mills *et al.* (2007), as they used fixed primary levels with a constant 10 dB difference in level between L_1 and L_2 for all ears, and did not compensate for the threshold elevations of individual subjects.

Behavioral threshold and ME acoustic transfer characteristics are not completely independent. The level of the pure tone used during measurements of behavioral thresholds will be affected by the transmission characteristics of the ME. Not only does the ME affect behavioral threshold by affecting stimulus level reaching the cochlea; in a similar fashion, it affects the DPOAE level by affecting the levels of the primary tones reaching the cochlea that elicit the DPOAE response. For example, Gorga *et al.* (2007) recently provided evidence that behavioral thresholds and ME reflectance covary (see their Fig. 1). The transmission characteristics of the ME may have affected the stimulus level reaching the cochlea in behavioral-threshold measurements, but did not have the same effect on the primary levels used in collecting the DPOAE data because we optimized stimulus levels for each individual ear, essentially factoring out the ME effects on the primary levels. This may have, in turn, decreased the observed effect of behavioral threshold on DPOAE level.

B. Middle-ear energy transfer

It appears that variables presumably representing forward or reverse transmission through the ME account for a small amount of the variability in DPOAE level and only at a few frequencies. These results may be due, in part, to the colinearity between behavioral threshold and ME measures. Since behavioral threshold accounted for the most variability in DPOAE level among the individual variables evaluated in the present study, and behavioral threshold and ME measures are not completely independent, the effects of the ME might be more pronounced if analyzed in the absence of behavioral threshold.

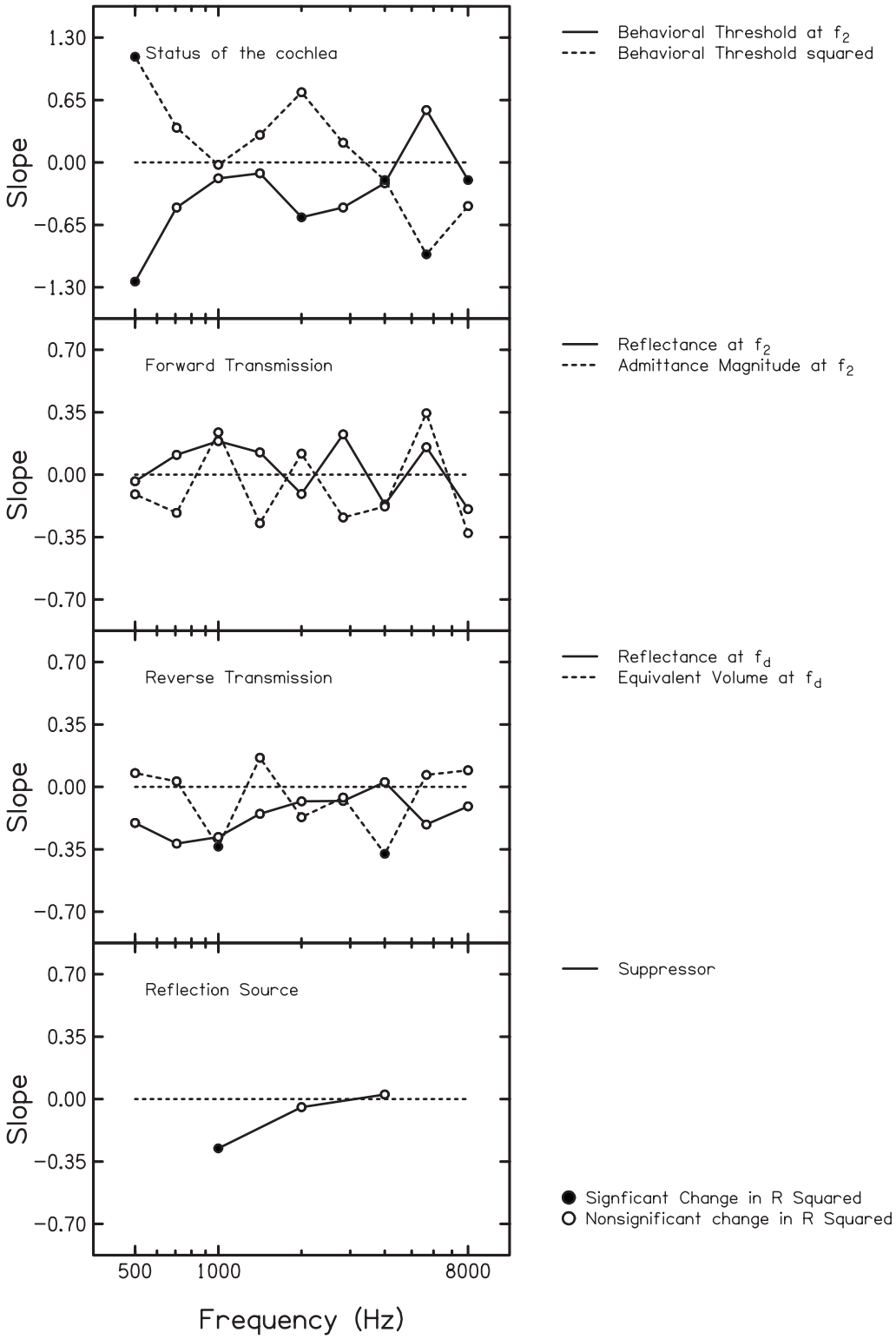


FIG. 9. Normalized slope plotted as a function of f_2 frequency. The top panel represents the status of the cochlea and plots behavioral threshold at the f_2 frequency and behavioral threshold squared at that same frequency. The second panel represents middle-ear transmission in the forward direction by plotting energy reflectance and admittance magnitude at the f_2 frequency. The third panel represents middle-ear transmission in the reverse direction by plotting energy reflectance and equivalent volume at f_d. The bottom panel represents the reflection-source contribution. Significant values are represented by filled symbols and nonsignificant values are represented by open symbols.

Furthermore, we may have reduced the effect of the ME in relation to the primary levels used in collecting the DPOAEs by optimizing the stimulus conditions for each individual ear. This effect is not unlike the effect optimization may have had on the relation between behavioral threshold and DPOAE variability. If we had not optimized stimulus conditions for each individual ear and instead used the same fixed primary levels to elicit the DPOAE in every subject, we might have seen a significant relationship between ME transmission characteristics and DPOAE level. Thus, optimal lev-

TABLE I. Constant, R^2, and p-values, and residuals for each of the nine f_2 frequencies.

Frequency	C	R^2	p-value	Residual
500	22.007	0.345	0.002	0.655
707	14.153	−0.057	0.688	1.0
1000	5.274	0.167	0.005	0.833
1414	5.824	0.016	0.380	0.984
2000	2.695	0.040	0.190	0.960
2828	5.734	0.008	0.411	0.992
4000	9.983	0.318	0.000	0.682
5656	−5.314	0.291	0.007	0.709
8000	22.756	0.326	0.003	0.674

els may have influenced our ability to observe significant ME effects (as well as behavioral-threshold effects), but was viewed as a good experimental design decision because our previous work has shown that the use of optimal stimulus conditions reduces response variability among normal-hearing ears (Neely *et al.*, 2005).

C. Source contribution

The suppressor variable accounted for a significant amount of variability in DPOAE level at only one of three f_2 frequencies for which it was evaluated (1000 Hz). This is consistent with past research in which there was more evidence for fine structure at lower frequencies than at higher frequencies (e.g., He and Schmiedt, 1993; Dhar and Shaffer, 2004; Johnson *et al.*, 2006b). The previously reported lack of fine structure at 2000 and 4000 Hz suggests that interactions between distortion and reflection sources are not as evident at these two frequencies. Given these observations for the higher frequencies, one would predict that there would be little or no difference in DPOAE level when the reflection source was suppressed, compared to the case when it was not. The results observed in the present study for $f_2=2000$ and 4000 Hz are consistent with this view.

Although the mean data indicate that the suppressor affected the DPOAE level, at least at 1000 Hz, we cannot be certain that the suppressor levels that we used suppressed only the reflection source and not the distortion source, which would also have the effect of decreasing the overall level of the DPOAE. Additional studies would be needed to assure that the reflection source was suppressed, without affecting the distortion source. For example, Mauermann and Kollmeier (2004) collected DPOAEs while varying frequency in fine steps and converting the responses to a time domain representation. A time windowing procedure was then used to separate the components from the two DPOAE sources. This method would provide evidence of whether or not the reflection source was suppressed.

Interestingly, it appears that the source-contribution variable was unaffected by the choice of stimulus conditions, in which stimulus level was optimized for each subject individually. This approach, which selects the L_1 resulting in the largest DPOAE for each L_2, might affect the L_2 at which source contributions were most evident, but it is unlikely that optimal stimulus conditions would minimize or eliminate contributions from the reflection source. Thus, it is perhaps not surprising that source contributions were evident in the frequency-dependent manner described in this study, although it is still the case that significant amounts of variance could not be assigned to this variable.

D. Combined model

In the combined model, behavioral threshold, ME ATFs, and a suppressor variable were combined to determine how much of the variance could be accounted for when the MLR included all three variables. The combined model accounted for a significant, although small, amount of variance in L_d corrected at five frequencies, slightly more than half of the frequencies studied. When the individual variables were considered separately, significant amounts of variance were accounted for at fewer frequencies, with the maximum being five for the behavioral threshold and behavioral-threshold squared variables. Thus, using the combined model accounted for more of the variance.

There still is residual variability, even for those frequencies for which the model was able to account for significant amounts of variance. Again, we may have reduced the effect of the individual variables by optimizing the stimulus conditions for each individual ear, because less variance remained to be explained by these variables. In particular, the use of the optimal stimuli might have reduced the influence of both the behavioral threshold and ME variables in the combined case, just as they may have had this effect when each variable was considered separately. If we had not optimized stimulus conditions for each individual ear and instead used the same fixed primary levels to elicit the DPOAE in all subjects, we may have produced models that accounted for significant amounts of the increased variance in DPOAE level at more frequencies. However, our goal was to identify and reduce variability in DPOAE levels, which is why optimal stimulus levels were used. It was not our goal to create stimulus conditions that maximized variability due to either behavioral threshold or ME transmission.

Another possible reason for the lack of variance accounted for may be due to the colinearity between behavioral threshold and the ME variables. Since the stimuli used in acquiring behavioral thresholds passed through the ME, there must be a relation between ME function and behavioral threshold, which would mean that these two measures are not independent.

Calibration may have been another reason for the lack of variance accounted for. Even though an attempt was made to maintain the same insertion depth throughout each testing session, this was not always possible. Different insertion depths were inevitable during the multiple sessions. The effects of changes in insertion depth are small, but could still have an effect on the calibration and measured SPL in the ear canal that could have contributed to the variability seen in DPOAE level.

Janssen *et al.* (2005) demonstrated that as SNR decreased, the DPOAE standard deviation increased. Although SNR was not a variable examined in this study, this could be another explanation for the increased variability at low f_2's.

In summary, behavioral threshold, ME transmission characteristics, and differences in source contributions were analyzed in order to determine which variables, either individually or in combination, account for the variability observed in DPOAEs among normal-hearing ears. The combined model accounted for a significant, although small, amount of variance in L_d corrected at five fresquencies, or more than half of the frequencies studied. There still is a large amount of residual variability that remains unaccounted for, even for those frequencies for which the model was able to account for significant amounts of variance. Although the use of procedures that optimized stimulus parameters for each individual ear may have reduced the effects of both behavioral threshold and ME energy transmission, the optimized stimulus is thought to have reduced the variance that the MLR was intended to explain. If we had used fixed primary levels in acquiring the DPOAE, we may have had a different outcome. Further study in which fixed stimulus conditions are used while the same variables are assessed would be needed to evaluate this hypothesis. However, our intention was not to observe significant effects of either behavioral threshold or ME transmission. Rather, we were interested in seeking ways to reduce variability in DPOAE levels; using optimal level ratios has been shown previously to have that effect.

ACKNOWLEDGMENTS

This work was supported by the NIH (NIDCD R01-DC02251 and P30-DC04662). The authors would like to thank Doug Keefe for providing the software that was used for measurements of middle-ear reflectance and for his helpful comments on the text on the middle ear. They also thank Sandy Estee for her help in subject recruitment and Judy Kopun for general assistance in the laboratory. The data summarized in this paper were taken from a dissertation submitted by the first author as part of the requirements for her Ph.D. from The University of Nebraska-Lincoln.

Dhar, S. and Shaffer, L. (**2004**). "Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams," Ear Hear. **25**, 573–585.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (**1998**). "On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions," J. Acoust. Soc. Am. **104**, 964–971.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2007**). "Low-frequency and high-frequency cochlear nonlinearity in humans," J. Acoust. Soc. Am. **122**, 1671–1680.

He, N. J., and Schmiedt, R. A. (**1993**). "Fine structure of the 2f1-f2 acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

Janssen, T., Boege, P., and von Mikusch-Buchberg, J. (**2005**). "Investigation of potential effects of cellular phones on human auditory function by means of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **117**, 1241–1247.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (**2006a**). "Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 418–428.

Johnson, T. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2006b**). "Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **119**, 3896–3907.

Keefe, D. H., and Simmons, J. L. (**2003**). "Energy transmittance predicts conductive hearing loss in older children and adults," J. Acoust. Soc. Am. **114**, 3217–3238.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the 2 f1-f2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.

Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.

Mills, D. M., Feeney, M. P., and Gates, G. A. (**2007**). "Evaluation of cochlear hearing disorders: normative distortion product otoacoustic emission measurements," Ear Hear. **28**, 778–792.

Müller, J., Janssen, T., Heppelmann, G., and Wagner, W. (**2005**). "Evidence for a bipolar change in distortion product otoacoustic emissions during contralateral acoustic stimulation in humans," J. Acoust. Soc. Am. **118**, 3747–3756.

Neely, S. T., Johnson, T. A., and Gorga, M. P. (**2005**). "Distortion-product otoacoustic emission measured with continuously varying stimulus level," J. Acoust. Soc. Am. **117**, 1248–1259.

Neely, S. T., and Liu, Z. (**1994**). "EMAV: Otoacoustic emission averager," Technical Memorandum No. 17, Boys Town National Research Hospital, Omaha, NE.

Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Siegel, J. H., and Hirohata, E. T. (**1994**). "Sound calibration and distortion product otoacoustic emissions at high frequencies," Hear. Res. **80**, 146–152.

Wagner, W., Heppelmann, G., Müller, J., Janssen, T., and Zenner, H. P. (**2007**). "Olivocochlear reflex effect on human distortion product otoacoustic emissions is largest at frequencies with distinct fine structure dips," Hear. Res. **223**, 83–92.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L2 below L1," J. Acoust. Soc. Am. **97**, 2346–2358.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L1,L2 space," J. Acoust. Soc. Am. **97**, 2359–2377.

Statistics of instabilities in a state space model of the human cochlea

Emery M. Ku,[a] Stephen J. Elliott, and Ben Lineton
Institute of Sound and Vibration Research, University of Southampton, Southampton, United Kingdom SO17 1BJ

(Received 16 January 2008; revised 8 May 2008; accepted 8 May 2008)

A state space model of the human cochlea is used to test Zweig and Shera's [(1995) "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**(4), 2018–2047] multiple-reflection theory of spontaneous otoacoustic emission (SOAE) generation. The state space formulation is especially well suited to this task as the unstable frequencies of an active model can be rapidly and unambiguously determined. The cochlear model includes a human middle ear boundary and matches human enhancement, tuning, and traveling wave characteristics. Linear instabilities can arise across a wide bandwidth of frequencies in the model when the smooth spatial variation of basilar membrane impedance is perturbed, though it is believed that only unstable frequencies near the middle ear's range of greatest transmissibility are detected as SOAEs in the ear canal. The salient features of Zweig and Shera's theory are observed in this active model given several classes of perturbations in the distribution of feedback gain along the cochlea. Spatially random gain variations are used to approximate what may exist in human cochleae. The statistics of the unstable frequencies for random, spatially dense variations in gain are presented; the average spacings of adjacent unstable frequencies agree with the preferred minimum distance observed in human SOAE data. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2939133]

PACS number(s): 43.64.Kc, 43.64.Jb, 43.40.Vn, 43.64.Bt [BLM] Pages: 1068–1079

I. INTRODUCTION

The existence of a cochlear amplifier (CA) was first postulated by Gold (1948), who argued an electromechanical action is necessary to counteract the heavy viscous damping in the fluid-filled cochlea. The discovery of spontaneous otoacoustic emissions (SOAEs) by Kemp (1979) has long served as indirect evidence supporting the presence of a CA. It is now widely accepted that the outer hair cells situated in the organ of Corti actively enhance the motion of the basilar membrane (BM) (e.g., Diependaal *et al.*, 1987), which gives rise to a mammal's sharply tuned sense of hearing. However, the precise mechanism underlying the generation of SOAEs is still in debate.

SOAEs are believed to be a feature of a normally functioning CA, as they are commonly detected in an estimated range of 33%–70% of all normally hearing ears (Talmadge *et al.*, 1993). Where SOAEs are detected, stimulus frequency-, distortion product- and transient evoked-otoacoustic emissions (SFOAEs, DPOAEs, and TEOAEs) are often present. There is evidence to suggest that all forms of OAEs are related and directly tied to the sensitivity of hearing (Zwicker and Schloth, 1984; McFadden and Mishra, 1993; Talmadge and Tubis, 1993; Shera and Guinan, 1999). Two primary classes of cochlea-based theories regarding the production of SOAEs are discussed below: a local-oscillator model and a distributed backscattering concept.

Gold (1948) first formed the basis of a local-oscillator model of SOAE generation when he proposed that a perturbation may "bring an [active] element into the region of self-oscillation, when it is normally so close to [instability]." Evidence in the literature suggests that SOAEs are associated with BM oscillations. For example, Nuttall *et al.* (2004) measured a SOAE that had a counterpart in spontaneous mechanical vibration of the BM at the same frequency. Further work performed by Martin and Hudspeth (2001) considered how locally unstable elements of the CA may be responsible for SOAEs. However, without careful tuning, a local-oscillator model fails to account for the regular spacings between unstable frequencies observed in mammalian SOAEs.

The strong peak in the distribution of spacings between adjacent SOAE frequencies, termed the preferred minimum distance (PMD), has been demonstrated by various studies (Dallmayr, 1985, 1986; Talmadge *et al.*, 1993; Braun, 1997). A similar value is found in the average frequency spacings between the spectral peaks of SFOAEs and TEOAEs when measured in the ear canal (Zwicker and Schloth, 1984; Shera, 2003). The PMD corresponds to a frequency spacing of approximately 0.4 bark, or a distance of about 0.4 mm along the human cochlea (Dallmayr, 1985, 1986). Most SOAEs occur in the range of 0.5–6 kHz (Probst *et al.*, 1990) and demonstrate the PMD, though Zweig and Shera (1995) and Shera (2003) showed that the average spacings of both SOAEs and the spectral peaks of SFOAEs measured in the ear canal vary somewhat with frequency.

Strube (1989) argued that a periodic variation or "corrugation" in the micromechanical parameters would also give rise to the observed PMD in SFOAE and TEOAE measurements in the ear canal. This was said to arise given distributed backscattering of the traveling wave (TW) similar to the phenomenon of Bragg reflection in a crystal. In this theory, the period of the corrugation must correspond to one-half of

[a]Electronic mail: ek@isvr.soton.ac.uk.

 0001-4966/2008/124(2)/1068/12/$23.00

the wavelength of the TW, thus generating constructive interference at particular frequencies. Kemp (1979) also proposed a theory of SOAE generation which assumed a distributed backscattering mechanism; his theory required multiple internal reflections of forward- and backward-traveling waves between the middle ear boundary and an inhomogeneous region of the cochlea.

Since Kemp (1979) first presented the idea, numerous authors have made contributions to the multiple-reflection theory (Zwicker and Peisl, 1990; Zweig, 1991; Shera and Zweig, 1993; Talmadge and Tubis, 1993; Zweig and Shera, 1995; Allen *et al.*, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 1999; Shera, 2003). Shera and Zweig (1993) proposed that a spatially dense and random array of reflection sites exists along the entire cochlea which acts in concert with the middle ear boundary to form standing waves, which Shera (2003) likens to a laser cavity. This concept was fully developed in Zweig and Shera (1995). Though energy is reflected at all frequencies by a perturbation along the cochlea, wavelets scattered from forward-traveling waves that peak in the region of the inhomogeneity dominate the response, since the amplitude is highest there.

For an active standing wave resonance to develop in this multiple-reflection theory, the spatial distribution of inhomogeneities in the given region must contain components at the wavenumber that creates constructive interference with the incoming wave, just as with Bragg scattering (Shera and Zweig, 1993; Zweig and Shera, 1995). Further requirements include an active region between the middle ear boundary and the reflection site to overcome the viscous damping in the cochlea, and a TW frequency that undergoes an integer number of cycles of round-trip phase change between the middle ear and the cochlear reflection site; this naturally gives rise to the PMD in SOAEs measured in the ear canal. However, the existence of a spontaneous oscillation in the cochlea does not guarantee its detection as a SOAE; it must also remain sufficiently powerful to be measurable in the ear canal after transmission through the middle ear.

An alternative theory suggests that irregular middle ear transmission characteristics may be a cause of some OAEs (Nobili *et al.*, 2003). However, the numerical accuracy of these simulation results has been contested elsewhere (Shera *et al.*, 2003), and such irregularities are not often reported. For the purposes of this investigation, a smooth middle ear boundary is implemented and only cochlea-based theories of SOAE generation are discussed.

It should be noted that this paper considers only the linear stability of the cochlear model. In a biological cochlea, the amplitude of an instability would eventually stabilize due to the natural saturation of the feedback force generated by the CA. Furthermore, it is possible that the number of SOAEs predicted by the linear model could change in a nonlinear model due to distortion or suppression, for example.

A. Aims and overview

The goal of this paper is to test whether the predictions formalized by Zweig and Shera's (1995) multiple-reflection theory of SOAE generation are observed in a mathematical model of linear cochlear mechanics. Previous work has relied upon phenomenological methods (Zweig and Shera, 1995; Shera, 2003), or multiple time-domain simulations (Talmadge *et al.*, 1998), to support this theory. In contrast, a state space formulation of the cochlea (Elliott *et al.*, 2007) is used here that is capable of rapidly and unambiguously calculating the unstable frequencies in a given linear model. This method is thus especially well suited to generating the large number of results from individual cochleae necessary to ensure statistically significant data.

Section II presents the revisions necessary to adapt the original model (Neely and Kim, 1986), on which the state space model of Elliott *et al.* (2007) was based, from representing a cat cochlea to representing a human cochlea. For instance, a boundary approximating the human middle ear is now included. The features of the model that are pertinent to the "cochlear laser" theory, such as the wavelength of a TW at its peak as a function of position, are examined. Sample frequency responses and the stability of a base line cochlear model are also briefly described.

In Sec. III, the smoothly varying BM impedance along the cochlea is perturbed with a variety of spatial inhomogeneities in the micromechanical feedback gain in order to introduce reflection sites. The following inhomogeneities are tested: a step change in gain; sinusoidal variations in gain; and band-limited spatially random variations in gain are applied in order to simulate what may exist in human cochleae. A large number of simulations from the last category are performed. The spacings of adjacent unstable frequencies in the randomly perturbed cochlear models are collected and statistically analyzed at the end of this section.

II. MODEL DESCRIPTION

Elliott *et al.* (2007) used a state space formulation to determine the stability of Neely and Kim's (1986) discrete, long wave model of the cat cochlea. The goal of the current work is to be able to compare numerical simulations to human measurements; thus, revisions to the model were necessary to account for the pertinent features of the human cochlea. The changes are described in this section: starting at the middle ear boundary at the oval window, followed by the micromechanical elements of the cochlea, and ending at the helicotrema boundary at the apex. Illustrative simulations and the features of the model pertaining to stability are presented after the revisions.

Shera and Zweig (1990) pointed out the importance of the middle ear boundary as the dominant source of reflections for retrograde TWs in the cochlea. As such, careful attention was given to creating a boundary condition in the revised model that approximates the key features of physiological measurements. The data of Puria (2003) was used as a target when revising Neely and Kim's (1986) mass-spring-damper boundary.

Table I. lists the modified values of the micromechanical elements used in this model, and Fig. 1. shows Z_{out}, the impedance looking out of the cochlea into the middle ear[1] for both the state space model and Puria's (2003) measurements.

TABLE I. Revised parameters of the micromechanical model, as described in Elliott *et al.* (2007), in SI units, where x is the longitudinal distance along the cochlea.

Quantity	Formula (SI)
$k_1(x)$	$4.95\times10^{9}e^{-320(x+0.00375)}$ N m^{-3}
$c_1(x)$	$1+19700\ e^{-179(x+0.00375)}$ N s m^{-3}
m_1	1.35×10^{-2} kg m^{-2}
$k_2(x)$	$3.15\times10^{7}e^{-352(x+0.00375)}$ N m^{-3}
$c_2(x)$	$113\ e^{-176(x+0.00375)}$ N s m^{-3}
m_2	2.3×10^{-3} kg m^{-2}
$k_3(x)$	$4.5\times10^{7}e^{-320(x+0.00375)}$ N m^{-3}
$c_3(x)$	$22.5\ e^{-64(x+0.00375)}$ N s m^{-3}
$k_4(x)$	$2.82\times10^{9}e^{-320(x+0.00375)}$ N m^{-3}
$c_4(x)$	$9650\ e^{-164(x+0.00375)}$ N s m^{-3}
γ	1
H	0.001 m
L	0.035 m
A_s	3.2×10^{-6} m^2
k_{ME}	2.63×10^{8} N m^{-3}
c_{ME}	2.8×10^{4} N s m^{-3}
m_{ME}	2.96×10^{-2} kg m^{-2}
c_H	210 N s m^{-3}
m_H	1.35×10^{-2} kg m^{-2}
N	500

The micromechanical model and the significance of all the quantities are described in Elliott *et al.* (2007). The values of the Neely and Kim's (1986) parameters have been scaled in order to obtain a distribution of characteristic frequencies that matches those of Greenwood (1990) over the range of interest. Whereas Elliott *et al.* (2007) left the boundary at the helicotrema as a pressure release, it is now revised to include a small amount of damping. In order to incorporate the damped boundary into the state space model, it was necessary to make a minor modification to the macromechanical fluid-coupling matrix. The details of the new boundary condition and the revised matrix are explained in the Appendix. This change only affects simulations at low frequencies by reducing the reflectivity of the helicotrema, thus simplifying the interpretation of results.

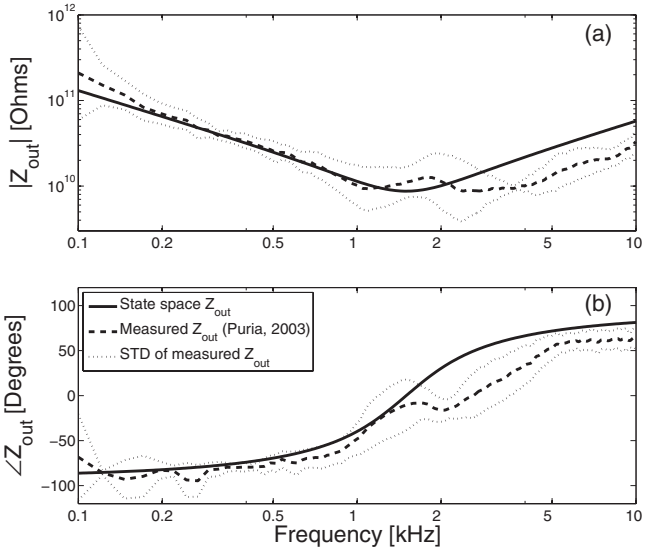


FIG. 1. Magnitude (a) and phase (b) of the impedance of the state space middle ear boundary and measured impedance looking out of the cochlea, Z_{out} (Puria, 2003).

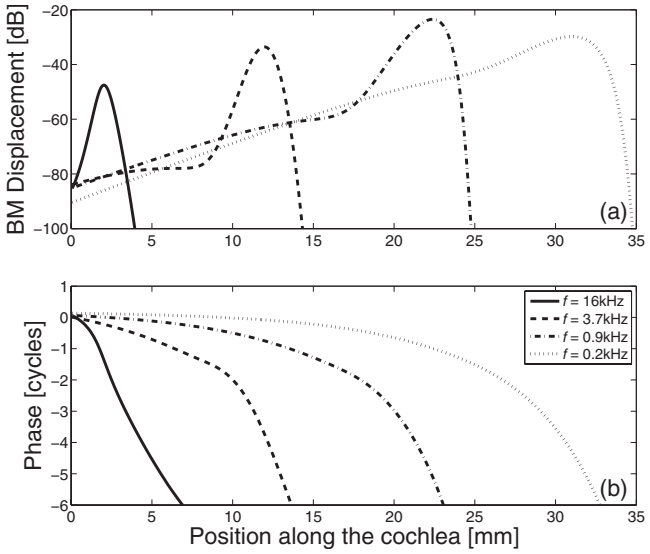


FIG. 2. BM displacement magnitude (a) and phase (b) given the four stimulus tones at $f=16$, 3.7, 0.9, and 0.2 kHz in the base line model [$\gamma(x)=1$].

The macromechanical formulation of the state space model (Elliott *et al.*, 2007) was based on work by Neely (1981) and Neely and Kim (1986). This uses a finite difference approximation to discretize the spatial derivatives in the wave equation and boundary conditions of the cochlea. The local activity of the cochlear partition segments is related to the fluid mechanics by

$$\mathbf{F}\mathbf{p}(t) - \ddot{\mathbf{w}}(t) = \mathbf{q}(t), \tag{1}$$

where $\mathbf{p}(t)$ and $\ddot{\mathbf{w}}(t)$ are the vectors of pressure differences and cochlear partition accelerations, $\mathbf{F}$ is the finite-difference matrix, and $\mathbf{q}(t)$ is the vector of source terms. The cochlear micromechanics of isolated partition segments are described by individual matrices. When Eq. (1) is substituted into an equation combining all the uncoupled elemental matrices, the coupled model of the cochlea can be described by the state space equations

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{2}$$

and

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \tag{3}$$

where $\mathbf{A}$ is the system matrix, $\mathbf{x}(t)$ is the vector of state variables, $\mathbf{B}$ is the input matrix, $\mathbf{u}(t)$ is a vector of inputs proportional to $\mathbf{q}(t)$, $\mathbf{y}(t)$ is the output variable (BM displacement in this case), $\mathbf{C}$ is the output matrix, and $\mathbf{D}$ is an empty feedthrough matrix.

Figure 2 illustrates typical BM displacement responses to tonal stimuli. The phase lag of these responses at CF is similar to measurements made in the middle of the squirrel monkey cochlea (Robles and Ruggero, 2001).

The magnitude of the impedance mismatch between the interface of the middle ear and the cochlea can now also be determined. The nominal value of the cochlear model's characteristic impedance, Z_c, has been determined to be 2

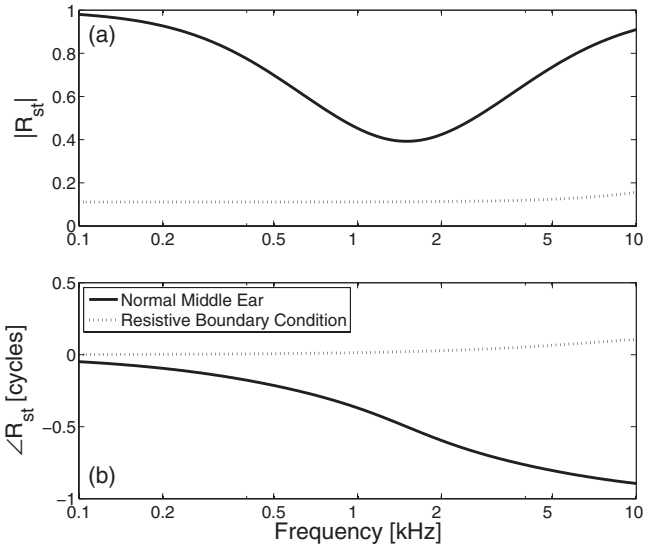


FIG. 3. Magnitude (a) and phase (b) of the basal reflection coefficient, R_{st}, given the base line middle ear (solid) and a largely resistive boundary (dotted).

$\times 10^{10}$ SI acoustic ohms. The reflection coefficient due to the middle ear as viewed from the cochlea, R_{st}, is given by Shera and Zweig (1990):

$$R_{\mathrm{st}} = \frac{Z_{\mathrm{out}} - Z_c}{Z_{\mathrm{out}} + Z_c}. \tag{4}$$

The magnitude and phase of the state space model's reflection coefficient are plotted in Fig. 3 for the base line middle ear boundary, and also a resistance-dominated boundary ($C_{\mathrm{ME}} = 8 \times 10^4$ N s m^{-3}).

Figure 4 shows the stability plot of the base line cochlear model given a nominal value of micromechanical feedback gain at all positions, $\gamma(x) = 1$. The stability plot shows the real (σ) and imaginary ($2\pi f$) parts of each of the poles of the coupled system, which are calculated from the eigenvalues of the system matrix, **A**, in Eq. (2). The imaginary components of the poles are converted from rad/s to kHz. For uniform values of feedback gain across the cochlea, the system becomes unstable at $\gamma(x) = 1.14$. This is indicated by the existence of at least one pole with a positive divergence rate, $\sigma(s) > 0$.

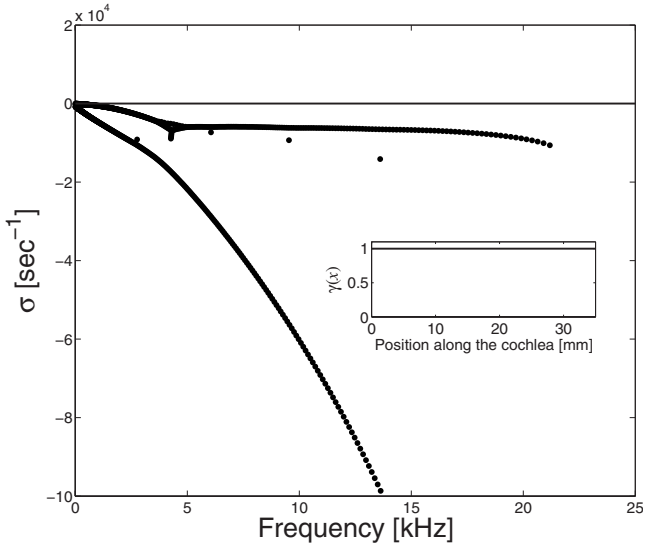


FIG. 4. A stability plot of the cochlear model given nominal gain, $\gamma(x) = 1$, and base line middle ear boundary.

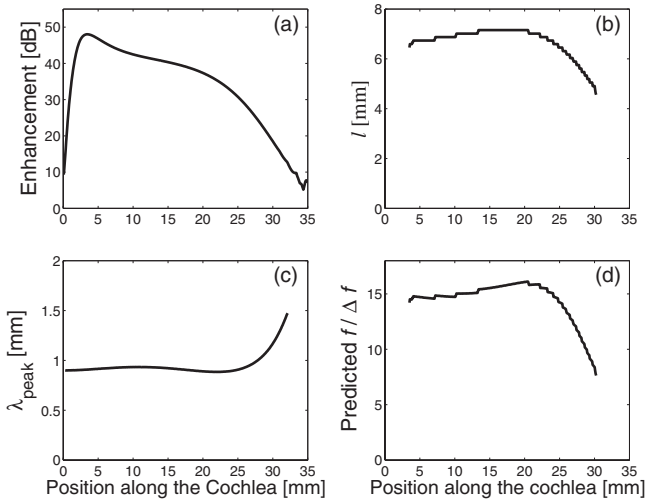


FIG. 5. Calculated characteristics of the model as a function of position along the cochlea: (a) active enhancement; (b) length scale, i.e., distance along the cochlea by which the characteristic frequency changes by a factor of e, as directly measured from the model; (c) λ_{peak}, wavelength of the TW in its peak region; (d) predicted spacing of SOAEs. Note that it was not possible to accurately calculate the length scale near the base and apex, hence the shortening of (b) and (d).

Shera (2003) argued that the CA is analogous to a laser's gain medium. One would expect a higher level of gain in the CA to result in greater system instability, given the same pattern of inhomogeneities in the cochlea. A higher value of feedback gain, $\gamma(x)$, results in greater active enhancement, which is defined here as the ratio of the cochlea's maximum active [$\gamma(x) = 1$] BM velocity to its maximum passive [$\gamma(x) = 0$] BM velocity across frequency at a given position, in dB. In the current model, the active enhancement is a function of position in the cochlea that is greatest (approximately 45 dB) near the base and gradually decreases toward the apex, shown in Fig. 5(a). This trend was demonstrated by Robles and Ruggero (2001), who made physiological measurements in animals.

According to Shera and Zweig (1993), the average distance between resonant positions of SOAEs along the cochlea is

$$\overline{\Delta x}_{\mathrm{SOAE}} \approx \frac{1}{2}\lambda_{\mathrm{peak}}, \tag{5}$$

where λ_{peak} is the wavelength of the TW in its peak region. Consequently, the predicted normalized spacing between SOAE frequencies is

$$f/\Delta f \approx 2l/\lambda_{\mathrm{peak}}, \tag{6}$$

where l is the cochlear length scale, the distance over which the best frequency changes by a factor of e, shown for the model in Fig. 5(b). The normalized spacing is defined as the ratio of the geometric mean of two adjacent SOAE frequencies, f_a and f_b, divided by their difference,

$$f/\Delta f = \frac{\sqrt{f_a f_b}}{|f_a - f_b|}. \tag{7}$$

The PMD in humans is approximately 15 when expressed in terms of $f/\Delta f$ (Shera, 2003).

In order to calculate the wavelength of the TW for a given position and frequency in the state space model, it is necessary to return to the wave equation (de Boer, 1996):

$$\frac{\partial^2 p(x,\omega)}{\partial x^2} + \kappa_{\mathrm{TW}}^2(x,\omega)p(x,\omega) = 0, \tag{8}$$

where p is the pressure across the BM and κ_{TW} is the wave number of the TW, both functions of position and frequency. The wave number is related to the cochlear partition impedance, Z_{cp}, by the following:

$$\kappa_{\mathrm{TW}}^2(x,\omega) = \frac{-2j\omega\rho}{HZ_{\mathrm{cp}}(x,\omega)}, \tag{9}$$

where ρ is the density of the fluid, and H is the height of the scala vestibule and scala tympani above and below the cochlear partition. By definition,

$$\mathrm{Re}(\kappa_{\mathrm{TW}}) = \frac{2\pi}{\lambda_{\mathrm{TW}}}, \tag{10}$$

where λ_{TW} is the wavelength of the TW.

It is now possible to relate the wavelength of the TW in its peak region to the cochlear partition impedance at a given place, x, with characteristic frequency, ω_{cf},

$$\lambda_{\mathrm{peak}}(x) = \mathrm{Re}\left[\sqrt{\frac{HZ_{\mathrm{cp}}(x,\omega_{\mathrm{cf}})}{-2j\omega_{\mathrm{cf}}\rho}}\right]2\pi. \tag{11}$$

This is shown in Fig. 5(c), and is approximately 0.9 mm across most of the cochlea and slowly increases near the apex, thus breaking scaling symmetry. This trend is also consistent with physiological measurements made at the base and apex in animals (Robles and Ruggero, 2001).

Given the cochlear length scale and the wavelength of the TW at its peak as a function of position, the predicted spacing between unstable frequencies, $f/\Delta f$, can now be calculated as in Eq. (7). This result is shown in Fig. 5(d). The model's predicted SOAE spacing is approximately the measured PMD in humans ($f/\Delta f \approx 15$) for most of the length of the cochlea and decreases toward the apex.

III. SPATIALLY VARYING GAIN

It has been previously reported that deviations from a smoothly varying set of micromechanical parameters can cause instability in cochlear models. It is believed that the frequencies of cochlear instability represent the frequencies of potential SOAEs. Elliott *et al.* (2007) demonstrated that these models are most sensitive to rapid changes in the gain as a function of position. In the current paper, greater consideration is given to the nature of the inhomogeneities introduced and the resultant characteristics of the unstable frequencies. The feedback gain as a function of position along the cochlea, $\gamma(x)$, has been chosen as the parameter to be perturbed. In order to compare the relative level of instability present in a cochlea, it is instructive to examine the number of unstable frequencies present. However, to further quantify the magnitude of a cochlear model's instability, the concept of a pole's damping ratio is reviewed.

A second-order system can be described by its damping ratio, ζ, a dimensionless quantity that describes the rate at which system oscillations decay following an initial perturbation. This is related to the poles of a system, $s=\sigma+j\omega$, in the following manner:

$$\zeta = \cos(\alpha) = \frac{-\sigma}{\sqrt{\sigma^2+\omega^2}}, \tag{12}$$

where α is the angle formed between the positive-real half-axis of the s-plane and the pole in question. When poles with nonzero imaginary components cross into the positive-real half-plane [$\zeta(s)<0$], the response of a linear system will diverge exponentially. The rate of this divergence is given by $e^{-\zeta\omega_n t}$, where t is the time and ω_n is the resonant frequency of the pole in units of angular frequency. ω_n is determined by calculating the imaginary component of the pole. The damping ratio of an unstable pole is useful as it relates the rate at which the system will become unstable; the average value of many poles can also be compared across different cochlear models. This quantity is referred to as the *undamping ratio* in this paper, in the context of discussing unstable poles, and is assigned as ξ:

$$\xi = -\zeta. \tag{13}$$

A step change in gain is employed as a starting point for the discussion of cochlear stability analysis. From there, sinusoidal spatial variations and the band-limited random spatial variations are applied as gain distributions. It is important to note that the step and sinusoidal distributions of $\gamma(x)$ are introduced to understand the underlying mechanisms of SOAE generation and should not be interpreted as an attempt to model what necessarily exists in a human cochlea.

A. Step change in gain

A step change in gain gives rise to a discontinuity in the variation of BM impedance as a function of position along the cochlea. An ideal step in space has a well-distributed wave number spectrum, and thus should reflect wavelets across a wide range of wavelengths. One additional consequence of varying the gain as a function of position, $\gamma(x)$, is that the underlying properties of the TW are affected. For instance, a higher gain results in a shorter λ_{peak}. To minimize this effect, a relatively small amplitude step was chosen with a $\pm 3\%$ deviation from nominal gain on either side of the step. The stability plot for the cochlear model with such a step imposed on the gain at 18.2 mm from the base of the cochlea is shown in Fig. 6.

Three distinct frequencies are found to be unstable in this cochlea, at 1.478, 1.577, and 1.669 kHz. These frequencies are all close to the characteristic frequency at the location of the discontinuity, which is 1.550 kHz. According to Zweig and Shera (1995), only the frequencies whose responses peak in this region may become unstable since not enough energy is reflected otherwise; this is seen in Fig. 6 as

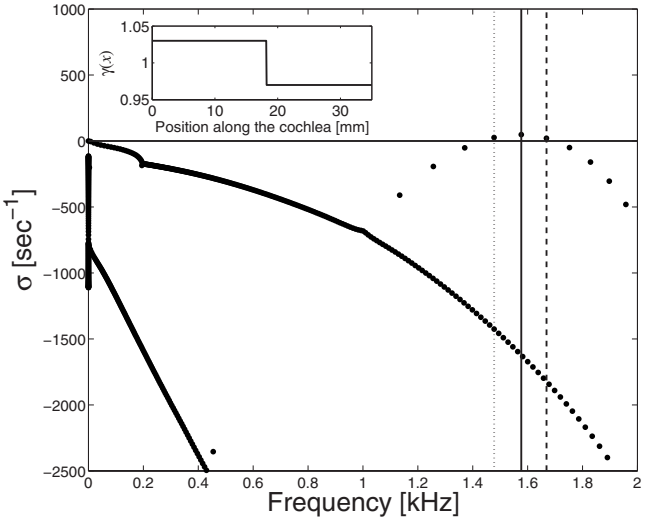


FIG. 6. Stability plot for a cochlea with the stepped gain inset: $\gamma(x<18.2\ \text{mm})=1.03$ and $\gamma(x\geq 18.2\ \text{mm})=0.97$. Note the frequency scale has been shorted to emphasize the locations of the unstable poles. Vertical lines are the frequencies of the unstable poles: 1.478 kHz (dotted), 1.577 kHz (solid), and 1.669 kHz (dashed).

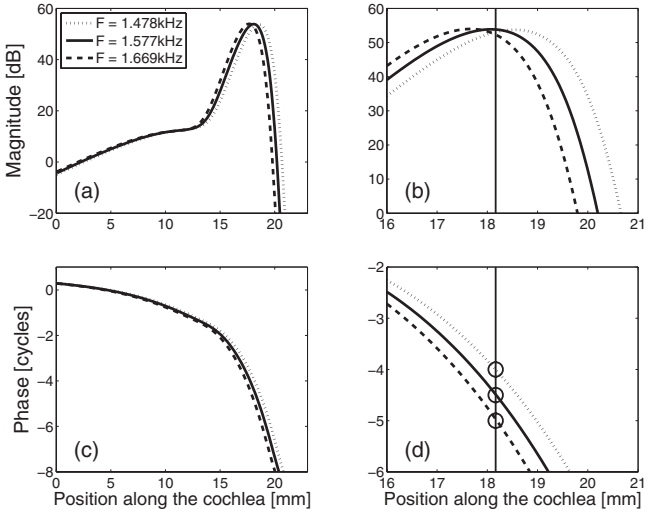


FIG. 7. Magnitude [(a) and (b)] and phase [(c) and (d)] of basilar membrane velocity for excitation at 1.478 kHz (dotted), 1.577 kHz (solid) and 1.669 kHz (dashed) given a base line model with nominal gain, $\gamma(x)=1$. (b) and (d) show the expanded axes for clarity of interpretation. A vertical line is drawn at the location of the discontinuity of Fig. 6 in the zoomed-in panels [(b) and (d)]. Circles in the phase plot (d) indicate phase shifts of −4, −4.5, and −5 cycles at this location.

only three frequencies near the discontinuity's characteristic frequency are unstable. Furthermore, there is a range of successively more stable poles that follow an arc leading away from the three unstable poles, both higher and lower in frequency. Presumably, the TWs of these frequencies are not reflected strongly enough by the discontinuity to cause instability.

The resultant spacings between the two pairs of adjacent unstable frequencies, $f/\Delta f$, are approximately 15 for the pair lower in frequency, and approximately 17 for the pair higher in frequency. This is consistent with the expectations given a slightly lower γ value apical of the discontinuity, and a slightly higher γ value basal to the discontinuity. To better understand why only these specific frequencies become unstable, Fig. 7 shows the magnitudes and phases of the BM velocity responses at these frequencies, for which a nominal gain throughout the cochlea is used, $\gamma(x)=1$.

A vertical line through Fig. 7(b) and Fig. 7(d) denotes the location along the cochlea of the discontinuity applied in Fig. 6. This line intersects with the phase responses of the 1.478, 1.577, and 1.667 kHz stimulus tones at −4, −4.5, and −5 cycles, respectively, within an accuracy of 1%. This is consistent with the "cochlear laser" theory of SOAE generation which requires that the phases of the unstable frequencies must undergo an integer number of cycles of total phase change between the reflection site and the middle ear boundary in order to combine constructively over successive reflections. For the unstable frequencies shown above, the "round-trip" phase change would equal 8, 9, and 10 cycles. Reexamining Fig. 6 in light of this feature, the stable poles that follow the same arc as the unstable poles must also represent frequencies that scatter wavelets which constructively combine, but perhaps are too weak to overcome the damping basal to the inhomogeneity.

Shera and Zweig (1993) and Zweig and Shera's (1995) concept of SOAE generation assumes wave amplification and multiple reflections between the middle ear boundary and the region of backscattering. A simple test of this theory involves changing the middle ear boundary so that it is less reflective.

Figure 8. shows the stability plot of a cochlear model with the same step change introduced in Fig. 6, but with a resistive boundary in the place of the human middle ear boundary, as shown in Fig. 3. The imaginary parts of the poles of Fig. 8 are almost identical to those of Fig. 6, but the real parts of the poles affected by the discontinuity are more stable. Whereas the base line model with a step change in gain was unstable, the model with the revised boundary and the same discontinuity is now stable.

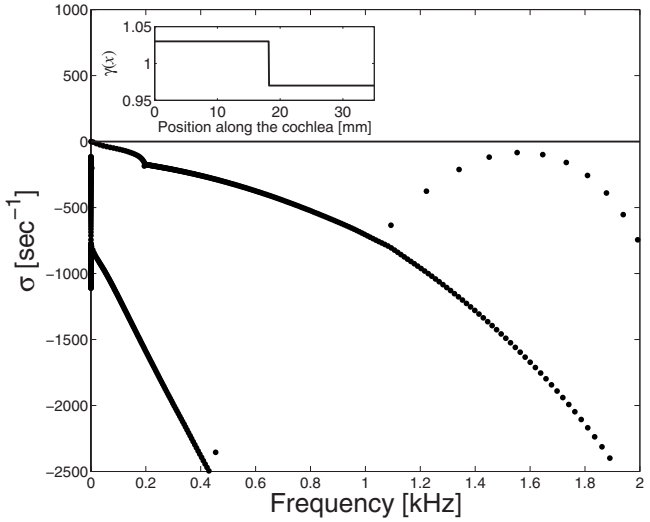


FIG. 8. Stability plot of a cochlear model with stepped gain distribution inset: $\gamma(x<18.2\ \text{mm})=1.03$ and $\gamma(x\geq 18.2\ \text{mm})=0.97$. The base line middle ear has been replaced with a resistive boundary, the reflection coefficient of which is shown in Fig. 3.

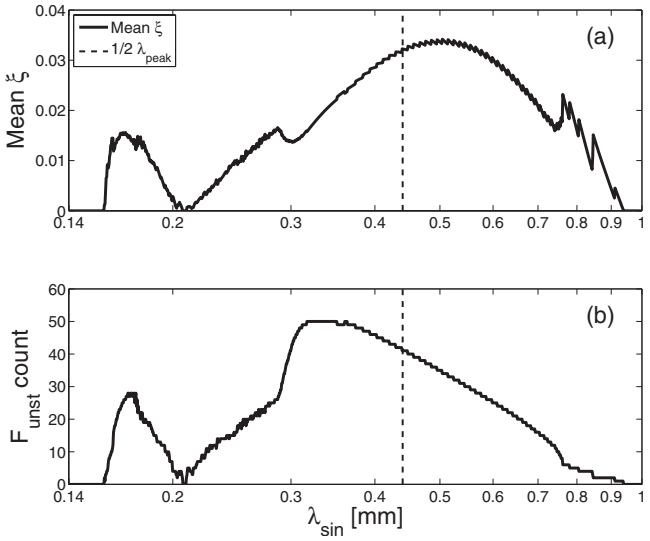


FIG. 9. Average undamping ratio ξ (a) and number of unstable frequencies (b) for a sinusoidal distribution of gain with varying wavelength, λ_{sin}. A vertical line marks the location of half the wavelength of the TW at its peak.

B. Sinusoidal variations in gain

A distribution of gain that is sinusoidal as a function of position is of interest because its wave number spectrum is concentrated at a single wave number, just as a sinusoidal wave form that is a function time has a frequency spectrum that is concentrated at a single frequency. This set of simulations follows the theory outlined by Strube (1989), which assumes uniform corrugations in gain along the BM. A range of wavelengths was chosen for the sinusoidal variation gain as a function of position along the cochlea, varying from 1 mm down to 0.14 mm, the latter being the spatial Nyquist limit of the model. A 10% peak-to-peak variation in amplitude about nominal gain generated instabilities over most of this spatial range, while maintaining stability for sinusoidal wavelengths greater than approximately 0.95 mm.

Figure 9 summarizes the level of instability in these cochleae by plotting both the mean undamping ratio, ξ, and the number of unstable frequencies as a function of the gain's sinusoidal wavelength. As expected, given the theories of Strube (1989), Shera and Zweig (1993), and Zweig and Shera (1995), the strongest instability occurred when the wavelength of the sinusoid, λ_{sin}, was approximately half the peak wavelength; this value occurs at 0.44 mm in the model. In addition, there was a region of greatly decreased instability in the model, centered about a periodicity of approximately one-fourth peak wavelength. This is thought to be due to destructive interference between the reflection sites, as the backscattered wavelets are out of phase with each other given this spatial periodicity.

The locally jagged aspect of the mean undamping ratio curve in Fig. 9(a) at approximately 0.75 mm is due to the periodic introduction of "new" unstable poles with low-undamping ratios that are generated as the wavelength of the sinusoid is varied. The average undamping ratio peaks at approximately 0.5 mm, which is slightly longer than half the peak wavelength for most of the length of the cochlea in this model. It is of note that the maximum in the total number of unstable poles, shown in Fig. 9(b), is located at a sinusoidal wavelength somewhat shorter than half the peak wavelength. As the sinusoidal wavelength of the gain variations is shortened, the number of peaks in the gain (and thus reflection sites) along the cochlea increases, creating more unstable poles. Even for sinusoidal periods less than half the peak wavelength, the rate at which unstable poles are being generated per unit decrease in λ_{sin} is still outpacing the rate at which they are returning to stability; this explains the location of the peak in Fig. 9(b).

C. Band-limited random gain distributions

Shera and Zweig's (1993) theory of SOAE generation assumes that the cochleae of normal-hearing humans contain a dense but random array of inhomogeneities. Each of these place-fixed perturbations reflects energy from the forward TW (Talmadge *et al.*, 1993; Shera and Zweig, 1993; Zweig and Shera, 1995). In this section, the stability of cochlear models with band-limited, spatially random gain distributions is used to approximate what is postulated to exist in a human cochlea. A fifth order Butterworth filter was employed to band-limit gain distributions in the wave number domain (Lineton, 2001). The low wave number cutoff frequency was fixed at the length of the cochlea itself, in order to prevent any dc shifts in the gain. The high wave number cutoff frequency was initially set to 6.6 radians/mm and slowly increased, thus generating cochlear models with successively more densely spaced reflection sites. The average filter bandwidths have been plotted below in terms of 2π times inverse wave number; this quantity has units of length (mm) and is directly comparable to the wavelength of the TW at its peak.

Figure 10 summarizes the results of simulations of 400 different cochleae, each with unique, spatially random gain distributions. The (a) panels show a typical stability plot from each group. The averaged power spectrum of the gain distributions is shown in the (b) panels, the two Roman numeral sets (I and II) having different high wave number (low wavelength) cutoffs; the dashed vertical line represents half the wavelength of the TW at its peak. A 5 mm sample of a gain distribution at this cutoff wavelength is inset. The (c) panels show the histograms of the average number of unstable poles per cochlea, sorted into logarithmic frequency bins. Figure 10(IIc) demonstrates that a lower cutoff wavelength, and thus a more rapid spatial variation in gain, is necessary to generate instability at frequencies below 2 kHz. This is believed to be due to the lower level of enhancement toward the apex of the cochlea and the lower magnitude of the basal reflection coefficient in this frequency band.

The histogram of normalized spacings of adjacent unstable frequencies per cochlea is shown in the (d) panels. The data for the (c) and (d) panels are presented for all instabilities (gray, thick bars) and also in a restricted range of 0.5–2 kHz (thin, black bars). This smaller range represents the frequency bandwidth where the middle ear's reverse-pressure transfer function is most efficient (Puria, 2003), and thus where one might expect the most SOAEs to be detected. The results for λ_{cutoff}=0.19 mm [Fig. 10(IId)] show a peak in

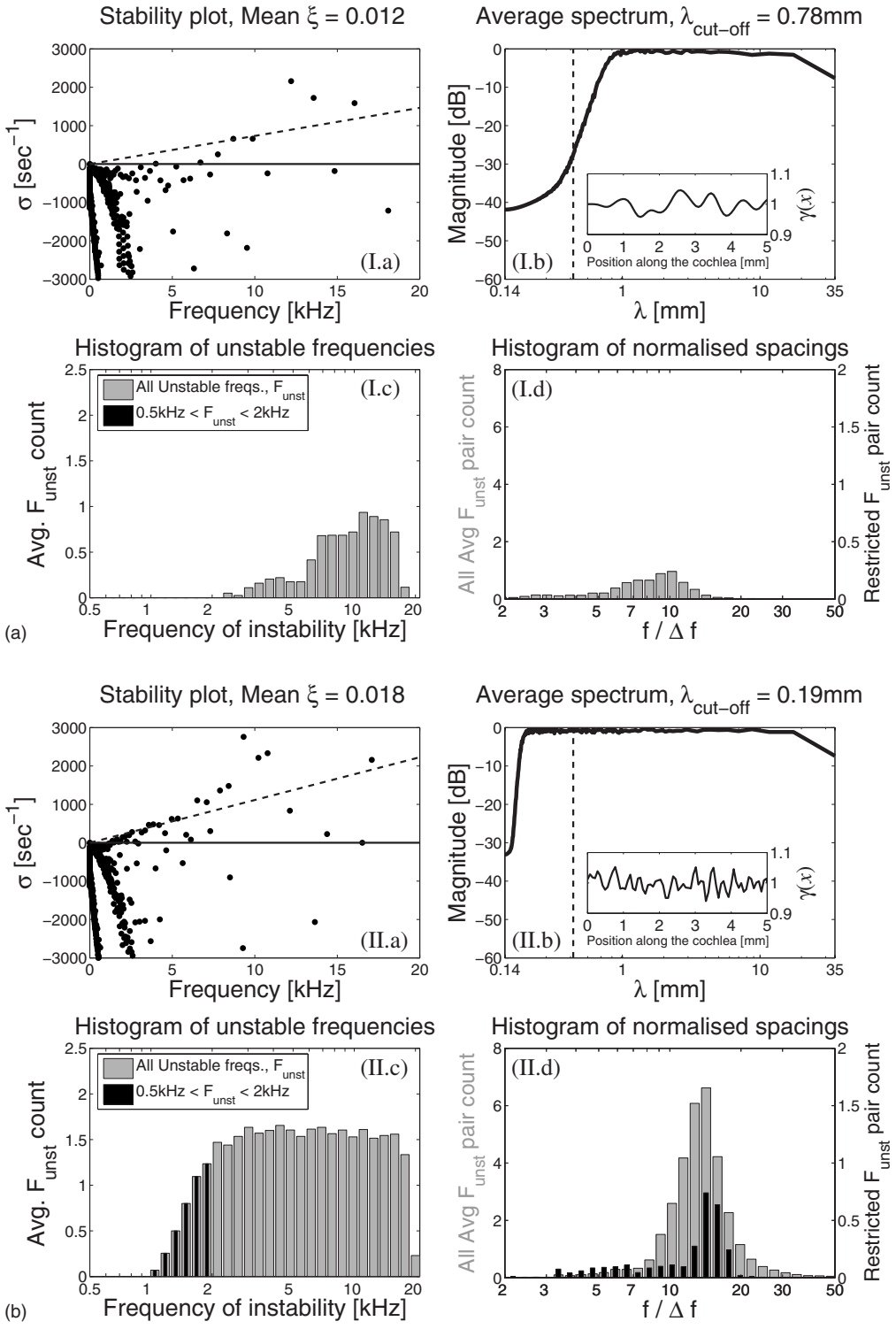


FIG. 10. The collected results from 2×200 cochlear models with randomly generated gain distributions. Each Roman numeral subset has been filtered with a different cutoff wavelength: [(Ia)–(Id)] λ_{cutoff}=0.78 mm, [(IIa)–(IId)] λ_{cutoff}=0.19 mm. A peak-to-peak amplitude of 15% was applied to these gain distributions. (a) A characteristic stability plot taken from the set. The average undamping ratio for that single case, ξ, is given and superimposed (dotted line). (b) Averaged inverse wave number spectrum of the gain distributions. Half the peak wavelength is indicated by a dotted vertical line and the first 5 mm of a characteristic gain distribution are inset. (c) Averaged histogram of all unstable frequencies per cochlea sorted in logarithmic frequency bins (gray). The instabilities occurring between 0.5 and 2 kHz are superimposed in thin, black bars. (d) Averaged histogram of normalized spacings ($f/\Delta f$) per cochlea in gray. The histogram of spacings of unstable frequencies per cochlea occurring between 0.5 and 2 kHz is again superimposed in black. Note the different left- and right-vertical scales.

the normalized spacing at $f/\Delta f\approx15$ in the frequency range of 0.5–2 kHz. These results are consistent with the Shera and Zweig's (1993) theory which assumes a dense array of reflection sites, represented in these simulations by a low cutoff wavelength.

Figure 11 summarizes data from the above calculations, while also presenting data from many other simulations which have different cutoff wavelengths and peak-to-peak variations in gain. The mean unstable frequency count and the mean undamping ratio, ξ, vary directly with the ampli-

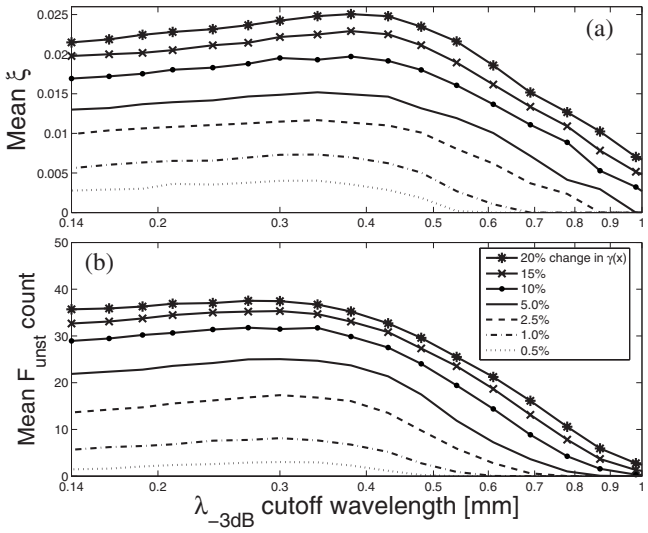


FIG. 11. Variation of average undamping ratio (a) and total unstable pole count (b) with cutoff wavelength for five amplitudes of peak-to-peak random variations in gain. At each amplitude, 20 cutoff wavelengths were each applied to 200 models with randomly generated gain distributions. A total of 28,000 stability tests were performed.

tude of the variation in $\gamma(x)$. This result is consistent with the findings of Elliott *et al.* (2007). In contrast to the sinusoidal case (see Fig. 9), no distinct notch in either the average undamping ratio or the number of instabilities is apparent at a cutoff wavelength of approximately one-quarter of the peak wavelength. In the sinusoidal simulations, all of the spatial spectral energy was concentrated at a particular wave number; this potentially generated strong, destructive interference when the sinusoidal wavelength was one-quarter the peak wavelength. The spectral energy in the random spatial variations in gain is comparatively much more diffuse, perhaps reducing the amount of both constructive and destructive interferences. The statistics of the spacings of instabilities is thus largely independent of the exact form of the spatial variations, provided they have a significant component at the wave number corresponding to one-half λ_{peak}. Peak-to-peak variations in $\gamma(x)$ as small as 0.5% can give rise to instabilities, provided λ_{cutoff} is less than approximately 0.5 mm, near the half peak wavelength.

IV. DISCUSSION

The findings of this paper, based on a numerical model of the human cochlea, are consistent with the multiple-reflection theory of Zweig and Shera (1995). The state space formulation is able to predict the frequencies at which a linear, active cochlear model will become unstable. Elliott *et al.* (2007) presented a nonlinear time-domain simulation demonstrating that an isolated unstable pole will evolve into a limit cycle within the cochlea at the expected frequency. Direct measurements in animals have shown that spontaneous basilar membrane oscillation is associated with SOAEs in the ear canal (Nuttall *et al.*, 2004). Consequently, comparisons are drawn in this paper between measured SOAE characteristics and the instabilities generated in the cochlear model. However, it is worth highlighting the similarities and differences between measured data and these simulation results.

This model predicts that instabilities exist all along the cochlea and across a wide band of frequencies, given a dense array of inhomogeneities in the cochlea. In contrast, SOAEs in normal-hearing individuals are only routinely detected between 0.5 kHz and 6 kHz (Probst *et al.*, 1991). Even if instabilities exist in all regions along the average human cochlea, however, it is likely that only a subset of these will be detected in the ear canal. It is believed that the inefficient reverse-transmission characteristics of the middle ear hinder the detection of SOAEs outside of its best transmissibility range, given its steep drop-off below and above resonance, of approximately −40 dB/decade. The limited bandwidth of normally detected SOAEs is also potentially reduced by physiological noise and the current limitations of sensor technology. Just as improved measurement techniques have revealed increasingly sharp active BM enhancement through the years, refinements in recording technique have exposed a higher prevalence of SOAEs in more recent studies (Probst *et al.*, 1991; Penner and Zhang, 1997).

The average number of unstable frequencies shown in Fig. 11 for a 10% peak-to-peak variation in gain is similar to the maximum number of emissions detected in a single ear, some in excess of 30 SOAEs (Talmadge *et al.*, 1993). It has been shown that the level and number of instabilities in the state space model depend on the amplitude of the variations in BM impedance and the spatial density of the inhomogeneities. When nonlinear effects are incorporated into time-domain simulations, it is anticipated that the total number of instabilities may differ from those predicted by linear stability analysis.

It has been demonstrated by numerous experimentalists (e.g., Zwicker and Schloth, 1984) that externally applied stimuli can frequency-lock, phase-synchronize, suppress, or otherwise affect a SOAE. Some modelers have used Van der Pol oscillators to account for these phenomena (Bialek and Wit, 1984; Wit, 1986; van Hengel *et al.*, 1996). Further work is needed to examine the nonlinear interaction of limit cycles and external stimuli in the state space model presented here.

The current linear model predicts a distribution of unstable frequency spacings that is similar to physiologically compiled data in several respects. Although the current model's results do not accurately match the observed variation in SOAE spacings with frequency (Shera, 2003), the spacing results presented in Fig. 10(IId) are consistent with the predictions shown in Fig. 5(d). Furthermore, the peak in the normalized spacings of Fig. 10(IId) is correctly located at the PMD when sufficient spectral content is present in the inhomogeneities at half the peak wavelength, as predicted by Zweig and Shera (1995).

When the current understanding regarding hearing sensitivity, the various forms of OAEs and pathology are combined, a convincing picture regarding the generation of SOAEs begins to emerge. As many authors have pointed out, SOAEs in humans appear to be a natural by-product of the species' sharply tuned sense of hearing. Normal hearing individuals that do not exhibit SOAEs typically have an audio-

gram which underperforms those with SOAEs by approximately 3 dB in the standard 1–6 kHz range (McFadden and Mishra, 1993). Pélanová *et al.* (2007) also reported that the high-frequency audiogram of normal-hearing children without SOAEs underperformed those with SOAEs by approximately 5 dB through the 10–16 kHz range. In the "laser-cochlea" theory of OAE generation, it is the portion of the cochlea basal to the reflection site that is crucial to sustaining the limit cycle oscillation. If the losses in this region are not overcome by the active enhancement provided by the outer hair cells, no spontaneous emission can occur.

V. CONCLUSIONS

Simulations using the state space model of the human cochlea show patterns of SOAE production that can be explained by Zweig and Shera's (1995) theory. As demonstrated by the step change in gain, only frequencies with a TW that undergoes an integer round-trip phase change between the middle ear boundary and the inhomogeneity will become unstable. Instabilities are detected along the entire cochlea given spatially random changes in gain, but it is believed that only a subset of these unstable frequencies become measurable as SOAEs due to the middle ear's inefficient reverse transmission characteristics. The spectral content of the inhomogeneities in the BM impedance also has a strong impact upon the level and frequency spacings of the resultant instabilities.

A 10% variation in gain as a function of position generated the most instability in the model when a sinusoidal inhomogeneity with a wavelength roughly equal to half the wavelength of the TW at its peak was applied; instability was eliminated when the sinusoid's wavelength was reduced to roughly one-fourth the wavelength of the TW at its peak. When random inhomogeneities are simulated, the expected PMD between adjacent unstable frequencies is strongly expressed in the results only when there is sufficient spectral content at one-half the wavelength of the TW at its peak.

Nonlinear time-domain simulations, such as those introduced in Elliott *et al.* (2007), are expected to provide a method of explaining the more subtle interactions that exist in human cochleae due to multiple instabilities and externally applied stimuli. However, it is clear that this linear model can provide a great deal of insight into the mechanisms underlying the generation of SOAEs as numerical results presented here are in good agreement with the theory of Zweig and Shera (1995).

ACKNOWLEDGMENTS

The authors would like to thank Dr. Sunil Puria for sharing his research data, and two anonymous reviewers for their helpful comments and suggestions. This work was partially supported by a Fulbright Postgraduate Award.

APPENDIX: HELICOTREMA BOUNDARY CONDITION

The structure of the noncoupled micromechanical matrices is identical to Elliott *et al.* (2007), except at the helicotrema for which

$$\dot{\mathbf{x}}_N(t) = \mathbf{A}_N \mathbf{x}_N(t) + \mathbf{B}_N p_N(t), \tag{A1}$$

where the boundary condition is now taken to be a mass-damper system, so that

$$\mathbf{x}_N(t) = [\dot{w}'_N(t) \quad w'_N(t)]^T, \tag{A2}$$

$$\mathbf{A}_N = \begin{bmatrix} \dfrac{-C_H}{M_H} & 0 \\ 1 & 0 \end{bmatrix}, \tag{A3}$$

and

$$\mathbf{B}_N = \begin{bmatrix} \dfrac{1}{M_H} & 0 \end{bmatrix}^T. \tag{A4}$$

In order to incorporate this change into the macromechanical formulation, it was necessary to insert an additional term in the finite difference fluid-coupling matrix, **F**, such that it is still invertible. The expanded matrices represented in Eq. (1) of this work now become

$$\frac{H}{2\rho\Delta^2}\begin{bmatrix} -\dfrac{\Delta}{H} & \dfrac{\Delta}{H} & & & & & & 0 \\ 1 & -2 & 1 & & & & & \\ 0 & 1 & -2 & 1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 1 & -2 & 1 & & 0 \\ & & & & 1 & -2 & & 1 \\ 0 & & & & & \dfrac{\Delta}{H} & & -\left(\dfrac{\Delta}{H}+\dfrac{\Delta^2}{H^2}\right) \end{bmatrix} \times \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-1}(t) \\ p_N(t) \end{bmatrix} - \begin{bmatrix} \ddot{w}_{SR}(t) \\ \ddot{w}_2(t) \\ \vdots \\ \ddot{w}_{N-1}(t) \\ \ddot{w}'_N(t) \end{bmatrix} = \begin{bmatrix} \ddot{w}_{SO}(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{A5}$$

where H is the height of the channel, ρ is the density of the fluid, and Δ is the length of a cochlear segment. The physical meaning of this additional term in the fluid-coupling matrix can be determined by relating this revised equation to the boundary condition at the apex.

The last row in Eq. (A5) represents the helicotrema boundary condition and can be written as

$$\frac{H}{\Delta^2}\left[\frac{\Delta}{H}p_{N-1} - \left(\frac{\Delta}{H}+\frac{\Delta^2}{H^2}\right)p_N\right] = 2\rho\ddot{w}'_N, \tag{A6}$$

where p_{N-1} and p_N are the pressures adjacent to and at the helicotrema, and $\ddot{w}'_N$ represents the "effective" helicotrema acceleration. Rewriting Eq. (A6) gives

$$\frac{p_{N-1}-p_N}{\Delta} = 2\rho\ddot{w}'_N + \frac{1}{H}p_N. \tag{A7}$$

However, the physical boundary condition is defined as

$$\left.\frac{dp}{dx}\right|_{x=L} = 2\rho\ddot{w}_N, \tag{A8}$$

where $\ddot{w}_N$ is the true helicotrema acceleration. Using a finite difference approximation in Eq. (A8) and expressing the acceleration at the helicotrema as a velocity yields

$$\frac{p_{N-1} - p_N}{\Delta} \approx 2j\omega\rho\dot{w}_N. \tag{A9}$$

The true admittance at the helicotrema is the volume velocity at the helicotrema divided by the local pressure:

$$Y_N = \frac{A\dot{w}_N}{p_N}, \tag{A10}$$

where A is the area of the helicotrema. Relating the approximated boundary condition in terms of the admittance gives

$$\frac{p_{N-1} - p_N}{\Delta} \approx \frac{2j\omega\rho}{A} Y_N p_N. \tag{A11}$$

The effective velocity at the helicotrema can also be expressed in terms of an effective admittance at the helicotrema, Y'_N [defined by the parameters in Eqs. (A1)–(A4)]:

$$\dot{w}'_H = \frac{Y'_N p_N}{A}. \tag{A12}$$

Substituting Eqs. (A11) and (A12) back into Eq. (A7) results in an equation that relates the true helicotrema admittance to its effective value:

$$\frac{2j\omega\rho}{A} Y_N p_N = \frac{2j\omega\rho}{A} Y'_N p_N + \frac{1}{H} p_N. \tag{A13}$$

Simplifying Eq. (A13) reveals

$$Y_N = Y'_N + \frac{A}{2j\omega\rho H}. \tag{A14}$$

The term $A/2j\omega\rho H$ is equivalent to the admittance of an acoustic mass of $m=2\rho H/A$. The acoustic mass of a short tube of length L and area A is $2\rho L/A$. In this case, $H=L$ corresponds to the assumed length of the helicotrema opening. The assigned value of 1 mm corresponds well with the value quoted by Fletcher (1953). The added term in the finite difference fluid-coupling matrix can be interpreted as an inertial term in parallel with the effective helicotrema impedance which is defined by the state space model in Eqs. (A1)–(A4). It should be noted that the change to the helicotrema boundary condition has a negligible effect on the model's response above approximately 200 Hz. Below this frequency, the reflections from the apex are more strongly attenuated than when using the pressure release boundary condition presented in Elliott *et al.* (2007).

[1] Z_{out} is referred to as M3 in Puria (2003).

Allen, J. B., Shaw, G., and Kimberley, B. P. (**1995**). "Characterization of the nonlinear ear canal impedance at low sound levels," Assoc. Res. Otolaryngol. Abstr. **18**, 190.

Bialek, W., and Wit, H. (**1984**). "Quantum limits to oscillator stability: theory and experiments on acoustic emissions from the human ear," Phys. Lett. **104**, 173–178.

Braun, M. (**1997**). "Frequency spacing of multiple spontaneous otoacoustic emissions shows relation to critical bands: A large-scale cumulative study," Hear. Res. **114**, 197–203.

Dallmayr, C. (**1985**). "Spontane oto-akusticsche Emissionen: Statistik und Reaktion auf akustische Störtöne," Acustica **59**, 67–75.

Dallmayr, C. (**1986**). "Stationäre und dynamische Eigenschaften spontaner und simultan evozierter oto-akustischer Emissionen," dissertation, Technische Universität, Munich.

de Boer, E. (**1996**). "Mechanics of the cochlea: modeling efforts," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 258–317.

Diependaal, R. J., de Boer, E., Viergever, M. A. (**1987**). "Cochlear power flux as an indicator of mechanical activity," J. Acoust. Soc. Am. **82**, 917–926.

Elliott, S. J., Ku, E. M., and Lineton, B. (**2007**). "A state space model for cochlear mechanics," J. Acoust. Soc. Am. **122**, 2759–2771.

Fletcher, H. (**1953**). in *Speech and Hearing in Communication*, edited by J. B. Allen (Acoustical Society of America, New York), pp. 248.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Gold, T. (**1948**). "Hearing. II. The physical basis of the action of the cochlea," Proc. R. Soc. London, Ser. B **135**, 492–498.

Kemp, D. T. (**1979**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Lineton, B. (**2001**). "Testing a model of the stimulus frequency otoacoustic emissions in humans," Ph.D. thesis, University of Southampton, Southampton.

Martin, P., and Hudspeth, A. J. (**1999**). "Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli," Proc. Natl. Acad. Sci. U.S.A. **96**, 14380–14385.

McFadden, D., and Mishra, R. (**1993**). "On the relation between hearing sensitivity and otoacoustic emissions," Hear. Res. **71**, 208–213.

Neely, S. T. (**1981**). "Finite difference solution of a two-dimensional mathematical model of the cochlea," J. Acoust. Soc. Am. **69**, 1386–1393.

Neely, S. T., and Kim, D. O. (**1986**). "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1480.

Nobili, R., Vetesnik, A., Turicchia, L., and Mammano, F. (**2003**). "Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model," J. Assoc. Res. Otolaryngol. **4**, 478–494.

Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). "Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig," J. Assoc. Res. Otolaryngol. **5**, 337–348.

Pélanová, J., Groh, D., Popear, J., Kabelka, Z., and Syka, J. (**2007**). "Presence and characteristics of spontaneous otoacoustic emissions in children and adolescents," Inner Ear Biology, Vol. **113**, p. 26.

Penner, M. J., and Zhang, T. (**1997**). "Prevelance of spontaneous otoacoustic emissions in adults revisted," Hear. Res. **103**, 28–34.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2066.

Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.

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

Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.

Shera, C. A., Tubis, A., and Talmadge, C. L. (**2003**). "Stimulus-spectrum irregularity and the generation of evoked and spontaneous otoacoustic emissions: comments on the model of Nobili *et al.*" Last viewed 12/20/2007. http://otoemissions.org/whitepapers/biophysics/chris_shera.html

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 Zweig, G. (**1990**). "Reflection of retrograde waves within the cochlea and at the stapes," J. Acoust. Soc. Am. **89**, 1290–1305.

Shera, C. A., and Zweig, G. (**1993**). "Order from chaos: resolving the paradox of periodicity in evoked otoacoustic emission," *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.

Strube, H. W. (**1989**). "Evoked otoacoustic emissions as cochlear Bragg reflections," Hear. Res. **38**, 35–46.

Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). "New

off-line method for detecting spontaneous otoacoustic emissions in human subjects," Hear. Res. **71**, 170–182.
Talmadge, C. L., and Tubis, A. (**1993**). "On modeling the connection between spontaneous and evoked otoacoustic emissions," in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 25–32.
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.
van Hengel, P. W. J., Duifhuis, H., and van den Raadt, M. P. M. G. (**1996**). "Spatial periodicity in the cochlea: The result of interaction of spontaneous emissions?" J. Acoust. Soc. Am. **99**, 3566–3571.
Wit, H. P. (**1986**). "Statistical properties of a strong spontaneous otoacoustic emission," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. E. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, Berlin), pp. 221–228.
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., and Peisl, W. (**1990**). "Cochlear preprocessing in analog models, in digital models, and in human inner ear," Hear. Res. **44**, 209–216.
Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Medial olivocochlear efferent inhibition of basilar-membrane responses to clicks: Evidence for two modes of cochlear mechanical excitation

John J. Guinan, Jr.[a)]
Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, 243 Charles Street, Boston, Massachusetts 02114 and Harvard-MIT HST Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

Nigel P. Cooper
School of Life Sciences (ISTM/Neuroscience), Keele University, Staffordshire ST5 5BG, United Kingdom

(Received 15 November 2007; revised 28 May 2008; accepted 29 May 2008)

Conceptualizations of mammalian cochlear mechanics are based on basilar-membrane (BM) traveling waves that scale with frequency along the length of the cochlea, are amplified by outer hair cells (OHCs), and excite inner hair cells and auditory-nerve (AN) fibers in a simple way. However, recent experimental work has shown medial-olivocochlear (MOC) inhibition of AN responses to clicks that do not fit with this picture. To test whether this AN-initial-peak (ANIP) inhibition might result from hitherto unrecognized aspects of the traveling-wave or MOC-evoked inhibition, MOC effects on BM responses to clicks in the basal turns of guinea pig and chinchilla cochleae were measured. MOC stimulation inhibited BM click responses in a time and level dependent manner. Inhibition was not seen during the first half-cycle of the responses, but built up gradually, and ultimately increased the responses' decay rates. MOC stimulation also produced small phase leads in the response wave forms, but had little effect on the instantaneous frequency or the waxing and waning of the responses. These data, plus recent AN data, support the hypothesis that the MOC-evoked inhibitions of the traveling wave and of the ANIP response are separate phenomena, and indicate that the OHCs can affect at least two separate modes of excitation in the mammalian cochlea. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2949435]

PACS number(s): 43.64.Kc [BLM] Pages: 1080–1092

I. INTRODUCTION

The sensitivity and frequency selectivity of mammalian hearing are enhanced by the active amplification of hydrodynamic traveling waves that takes place in the cochlea. This amplification results from mechanical feedback between the outer hair cells (OHCs) in the organ of Corti and the basilar membrane (BM), but the details of this mechanism are poorly understood. At the macromechanical level, our knowledge is limited to inferences from model-based mathematical "inversions" of the measured BM motion (e.g., Zweig, 1991; de Boer and Nuttall, 2000). At the micromechanical level, two separate forms of OHC motility are presently being debated as candidates for the cochlear amplifier (for reviews, see Dallos *et al.*, 2006 and Fettiplace, 2006). The present investigation seeks to inform such debates by providing observations as to what OHCs do, and do not do, to the mechanics of the living mammalian cochlea.

According to the simplest models of cochlear mechanics, sound-evoked ossicular motion produces BM traveling waves that evoke a single pattern of vibration at each place along the length of the cochlea. This vibration pattern couples the BM's motion to the bending of inner-hair-cell (IHC) stereocilia and leads to the excitation of auditory-nerve (AN) fibers with little or no postmechanical filtering (see Narayan *et al.*, 1998). More complex, so-called "micromechanical" models have more than one mechanical degree of freedom within each cochlear cross section, and hence permit larger differences between BM motion and AN excitation (reviewed by Patuzzi, 1996). Such models find support in recent measurements that demonstrate complex, frequency-dependent patterns of vibration in real cochleae, either on the BM itself (e.g., Russell and Nilsen, 1997) or within the organ of Corti and tectorial membrane TM complex (Fridberger *et al.*, 2004; Nowotny and Gummer, 2006; Karavitaki and Mountain, 2007b). However, excitation of the IHCs in an intact cochlea by a mechanical vibration mode that significantly affects AN coding but is not seen in the BM's traveling wave has not yet been demonstrated.

The present study provides new insights into cochlear amplification by investigating BM responses to clicks with and without electrical stimulation of the medial olivocochlear (MOC) efferent fibers. The MOC fibers innervate the OHCs via large cholinergic synapses, and provide a convenient way to affect the OHCs in a reversible, physiologic way. MOC effects on BM motion have previously been studied only in responses to tones, where their main effect is to turn down the gain of the traveling-wave amplifier (for reviews, see Guinan 1996; Cooper and Guinan, 2006). One reason to study the effects of the MOC fibers on click-evoked motion is to investigate how this gain change is pro-

[a)]Author to whom correspondence should be addressed. Electronic mail: jjg@epl.meei.harvard.edu. Tel.: 617-573-4236. FAX: 617-720-4408.

duced. The click stimuli, which have wideband spectra but are punctuate in time, are advantageous because they allow the BM to respond at, and reveal, its own resonant frequencies (in contrast, tones force the BM to follow the externally applied frequency). If, for instance, the MOC-induced changes in the gain of the traveling-wave amplifier were brought about by changes in the effective stiffness of the BM (Dallos *et al.*, 1997), then the frequency content as well as the time course of the click responses would be expected to change. Another advantage of clicks is the spreading out in time (i.e., in response cycles) of the effects of different stages of cochlear amplification, so that the MOC-induced changes in the click responses should give us insight into the buildup and decay of traveling-wave amplification.

A further reason to study MOC effects on BM motion is to search for a mechanical correlate of a newly discovered form of AN inhibition (Guinan *et al.*, 2005). "AN-initial-peak" (ANIP) inhibition is manifested as a reduction in the initial peak of an AN fiber's response to a click stimulus, and differs substantially from the inhibition produced by reducing the gain of traveling-wave amplification (although traveling-wave inhibition is also apparent in the AN click responses, as shown in Guinan *et al.*, 2005). ANIP inhibition primarily affects responses at moderate-to-high click levels; it inhibits the initial peak of a response more than subsequent peaks; and its effects differ with the polarity of the clicks. Guinan *et al.* (2005) hypothesized that the ANIP response is due to an OHC-produced motion that is distinct from, and not present in, the classical BM traveling wave. This proposal contrasts strongly with the earlier inferences that only minor signal transformations intervene between BM vibration and AN excitation (e.g., see Narayan *et al.*, 1998). However, there are clear differences between the data on which these two viewpoints were based, the most striking being that they originate at the opposite "ends" of the cochlea: ANIP inhibition has only been demonstrated clearly in the apical half of the cochlea [in cat AN fibers with characteristic frequencies (CFs) of up to ~6 kHz; see Guinan *et al.*, 2005], while BM motion has only been compared rigorously with AN excitation in the basal half of the cochlea (most notably in the 8–10 kHz CF region of chinchillas; see Narayan *et al.*, 1998). Despite these differences, it is generally assumed that the classic BM traveling wave is present, in much the same form, throughout the cochlea (e.g., see Patuzzi, 1996). Thus, if the traveling wave is little changed throughout the cochlea and there are only minor transformations between BM motion and AN firing, then an inhibition corresponding to the ANIP inhibition should be seen in the early part of basal-turn BM click responses. One goal of the present work is to determine whether such an inhibition is seen in BM motion.

II. METHODS

Experiments were performed on deeply anesthetized animals in accordance with NIH, UK and US guidelines. Guinea pigs (320–550 g) were anesthetized using either sodium pentobarbitone (25 mg/kg, I.P.) and Hypnorm (0.6 ml/kg, I.M.; each milliliter of Hypnorm contains 10 mg fluanisone and 0.315 mg fentanyl citrate), or ketamine (50 mg/kg, I.M) and xylazine (10 mg/kg, I.M). Chinchillas (329–393 g) were anesthetized using sodium pentobarbitone alone (65 mg/kg, I.P.). Maintenance doses of anesthetics were given whenever needed, and the animals were killed humanely without recovery from anesthesia at the end of the experiments. Artificial ventilation was used as required to maintain end-tidal CO_2 levels near 4.5%. Core temperatures were maintained near 37.6 °C.

Acoustic stimuli were produced by a reverse-driven Brüel and Kjaer $\frac{1}{2}$ in. microphone and delivered to the ear canal via a closed sound system that produced very little ringing [spectra from a similar acoustic system are shown by Wilson and Johnstone (1975)]. Tone amplitudes are expressed as sound pressure levels (SPLs) in dB with regard to 20 μPa, and were calibrated using a microphone (Brüel and Kjær 4134) with a probe tube placed within 2 mm of the tympanic membrane. Click amplitudes are expressed in peak-equivalent SPLs (pSPL) and were determined from the tone calibration and the spectrum of the electronic pulse that produced the click; the pSPL of a click is the SPL of a tone that produces the same peak sound pressure as the click.

The cochlea was exposed via a dorsolateral bulla opening, and its physiological condition was monitored using AN compound action potential (CAP) thresholds recorded from a fine silver electrode near the round window. CAP thresholds usually deteriorated from their initial values during the opening of the cochlea. Moderate hearing deterioration (e.g., 10–25 dB) decreased the magnitude of the efferent effects but did not appear to change the patterns of the effects that were seen on the BM. Nonetheless, the data that we will consider extensively in this report were selected to show the largest MOC effects, and came from animals with minimal threshold losses (0–10 dB, near CF).

BM responses were monitored in the first turn of the cochlea using a displacement-sensitive interferometer (Cooper, 1999) with the output typically sampled at 200 kHz. The BM was exposed by shaving a small hole into the scala tympani. Gold-coated polystyrene microbeads (PolySciences Inc. 15–25 μm diameter) were dropped through the fluid onto the BM to provide enhanced reflectivity. A small glass cover slip was placed over the cochlear hole to avoid a movable air-to-fluid interface that might create interferometric artifacts (Cooper and Rhode, 1992).

MOC efferents were stimulated via a bipolar electrode at the floor of the fourth ventricle (Guinan and Stankovic, 1996). Shock stimuli were pulse trains with 0.3 ms pulse widths, ac coupled by a transformer, presented at 200–300 pulses/s, in 100 ms long bursts at 330 ms intervals. Pulse amplitudes were limited to near or just below the threshold for visible muscle twitches so that shock-induced motion did not interfere with the BM measurement. Control measurements of ossicular motion with and without shocks were made to ensure that middle-ear-muscle contractions did not affect the results.

Click stimuli were 20 μs long pulses, alternating in polarity, presented once every 11 ms with 15 clicks per burst, as shown in Fig. 1(a). Click bursts and MOC-shock bursts were presented every 330 ms with the first MOC shock delayed 30 ms after the first click. With this timing, the first

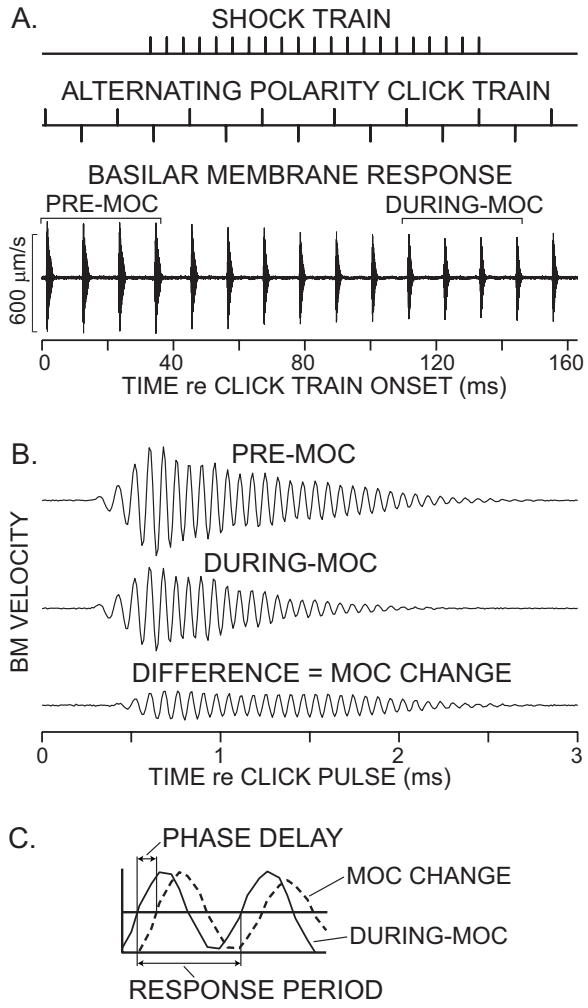


FIG. 1. The stimulus paradigm (A), derivation of the MOC change (B), and determining relative phases (C). (A) Schematic showing the timing of shock and click stimuli as well as the sets of four click responses used to get pre-MOC and during-MOC averages. (B) BM click-response wave forms measured prior to (pre-MOC) and during (during-MOC) the effects of the shock train, as well as their difference, the MOC change. (A) and (B) from GP3071, CF=14.5 kHz. The scale bar in (A) also applies to (B). (C) Phase delay calculated from the timing of the MOC-change zero crossings relative to the during-MOC zero crossings.

four click responses occurred before the effects of the efferent shocks had built up, and were called the "pre-MOC" responses. The four responses that were most affected by the MOC shocks were from clicks 11 to 14, and these were called the "during-MOC" responses [Fig. 1(a)]. Since the responses to the rarefaction clicks mirrored those to the condensation clicks almost perfectly, each set of four responses was averaged (after inverting the rarefaction-click responses) to increase the signal-to-noise ratios of the pre-MOC and during-MOC responses that will be considered throughout this paper. The MOC-induced change of the BM click response was obtained by subtracting the during-MOC response wave form from the pre-MOC response wave form, yielding the "MOC-change" wave form, as illustrated in Fig. 1(b).

MOC efferents can inhibit BM motion on two time scales, fast (time constant <100 ms) and slow (time constant ~ 10 s) (Sridhar *et al.*, 1995; Cooper and Guinan, 2003). Our MOC-change wave forms are only sensitive to MOC fast effects, because they compared the motion before each shock burst to the motion during that shock burst, as described above (Cooper and Guinan, 2003). Slow effects were also reduced by our data collection paradigm: Data were normally obtained in runs in which the click level was systematically varied (usually from high to low levels), with clicks and shocks paired as in Fig. 1. Slow effects decrease to near zero if MOC stimulation continues for several minutes, and after a run that produced a slow effect, it takes many minutes of recovery before a substantial MOC slow effect can be elicited again (Sridhar *et al.*, 1995). We reduced slow effects by doing the runs close enough in time that there was not sufficient recovery from the previous run to permit a substantial slow effect. To test for the presence of slow effects, occasional runs were done without efferent shocks, and responses from runs with versus without shocks were compared. Generally, this comparison showed little evidence of slow effects, so it seems likely that there was a very little slow effect during the runs reported here.

A. Analysis of click responses

Click responses were averaged across multiple presentations of a given stimulus before being stored to disk for subsequent analysis. In order to reduce the effects of any artifactual base line position shifts, such as those associated with breathing or shock-evoked animal motion, the first stage of our analysis was to convert the waveforms from displacement to velocity.

To quantify the BM click-response amplitude, we used the amplitude of the largest component of short-term fast Fourier transforms (FFTs) of the response, as calculated by MATLAB software's standard spectrogram function. To measure the initial part of the response with good time resolution, we used an analysis time of 1.5 CF cycles and a temporal overlap of 75%. A duration longer than a CF period was used to capture below-CF energy at the beginning of the response (using an integral number of CF periods was avoided because it produced a near-CF-period cyclic variation of the amplitude). As the response decayed into the noise in the later parts of the click response, we extended the analysis time to 4.5 and then 8.5 cycles, effectively averaging more cycles to increase the signal-to-noise ratio. Since the frequency content in this part of the response is very close to CF, the filtering aspect of this increase in the analysis window had little effect. However, with a multicycle analysis window and responses that are decaying with time, the initial part of the response is largest and tends to dominate the FFT. The resulting bias is greater for longer windows and faster decays; this prevented us from increasing the signal-to-noise ratio by extending the analysis time even further. In theory, a Hilbert transform method would be less biased by the fast decay of the response, but Hilbert transforms would not produce more accurate results (Lin and Guinan, 2004) because they use a derivative of the response, which is very sensitive to noise.

To characterize the phase changes produced by the MOC stimulation, we used the zero-crossing times of the click responses [Fig. 1(c)]. The time between adjacent (op-

positely directed) zero crossings was taken to be one-half period of the instantaneous response frequency. The advance (or delay) of a test wave relative to a reference wave was taken to be the time that its zero crossing came before (or after) the same-direction zero crossing of the reference wave [as shown in Fig. 1(c)]. Positive phase was the advance divided by the period (twice the half-period) of the response. Zero crossings are accurate only when the signal-to-noise ratio is good, so phase was only determined when the response amplitude was significantly larger than the background amplitude. By trial-and-error adjustment to remove aberrant points, we defined a significant amplitude to be 1.3 times larger than the maximum amplitude in the 1 ms period immediately preceding the onset of incus motion (when the incus response reached 10% of its peak value), or in the period between 6 and 11 ms after the click pulse. With the zero-crossing method, the response frequency was taken to be the reciprocal of 2× the time between adjacent zero crossings. The response frequency was also determined from the largest-amplitude component of the short-term FFT, as described above. The FFT and zero-crossing methods produced similar frequency measurements except near a sharp dip in the response. While the short-term FFT method was good for measuring the response frequency as a function of time, it was not satisfactory for comparing the phases of two response wave forms because the frequencies of their maximal responses could be different and phases measured from different frequencies cannot be unambiguously compared. Thus, for phase comparisons, only the zero-crossing method was used.

III. RESULTS

A. Click responses without MOC stimulation

Before MOC stimulation, BM responses to clicks (Figs. 1–4) were consistent with those previously reported from the cochlear basal turn of squirrel monkeys, guinea pigs, and chinchillas (Rhode and Robles, 1974; Robles *et al.*, 1976; LePage and Johnstone, 1980; Ruggero and Rich 1990, 1991a, 1991b; Ruggero and Rich, 1991a, 1991b; 1992a, Ruggero, *et al.*, 1992b 1993, 1996; Nuttall and Dolan, 1993; de Boer and Nuttall, 1997; Recio *et al.* 1998; Recio and Rhode, 2000). The instantaneous frequencies of the responses increased from well below the local CF, to the local CF, over approximately the first millisecond of the BM click response [Fig. 3(c)], in a characteristic pattern called a "glide" (de Boer and Nuttall, 1997; Shera, 2001a). The amplitude of the click response grew quickly, peaked, and then declined. As click level was increased, the initial half-cycle of the response grew approximately linearly with level, while later cycles showed varying degrees of compressive growth, particularly at higher sound levels (Figs. 2 and 3), a response pattern that is consistent with previous reports (Robles *et al.*, 1976; Recio *et al.* 1998). The declining part of the click response (the click "skirt") often showed waxing and waning [Fig. 3(a)] that were more prominent in animals with good thresholds. The predominance in animals with good thresholds indicates that waxing and waning are not due to pathology (Recio *et al.*, 1998).

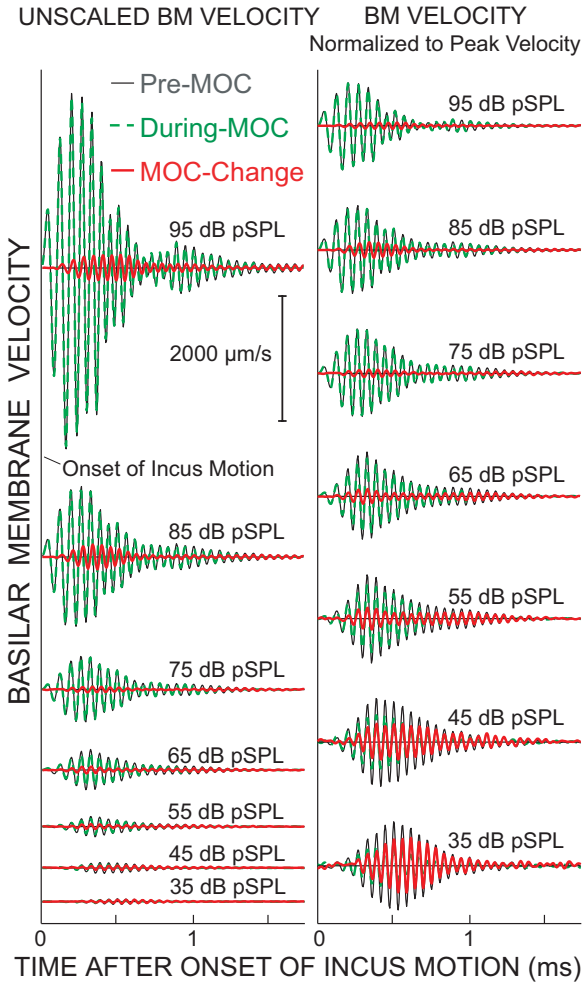


FIG. 2. Waveforms showing MOC effects on BM click responses as a function of click level. BM velocity in response to clicks at seven levels, prior to (thin black) and during (dashed green) the effects of MOC stimulation, and the MOC change (thick red; pre-MOC minus during-MOC velocities). Velocities all on the same scale (left) or normalized to the peak of each pre-MOC response (right). GP3096. CF=18 kHz.

B. Click responses with MOC stimulation

To successfully study the effects of MOC stimulation on BM click responses, the preparation had to have both a sensitive BM response and well-positioned MOC electrodes that resulted in a substantial MOC effect. A metric for the quality of the preparation that includes both factors is the MOC-induced level shift in the BM response to low-level clicks [Fig. 3(b)]. The data considered here are from the six guinea pigs with the largest MOC-evoked level shifts of BM click responses (7–15 dB) and the chinchilla with the largest level shift (5 dB). Figures 3, 6, and 7 show data from the guinea pig with the largest level shift (GP3144, shift=15 dB). Data from the two guinea pigs with the next highest level shifts (GP3096, shift=11 dB, and GP3131, shift=7.5 dB) and from the chinchilla are shown in the Supplementary Material (http://www.aip.org/pubservs/epaps.html).

The overall features of the MOC effects on BM click responses were similar across all animals, although there were considerable variations in click-response features across animals. MOC stimulation reduced the overall ampli-

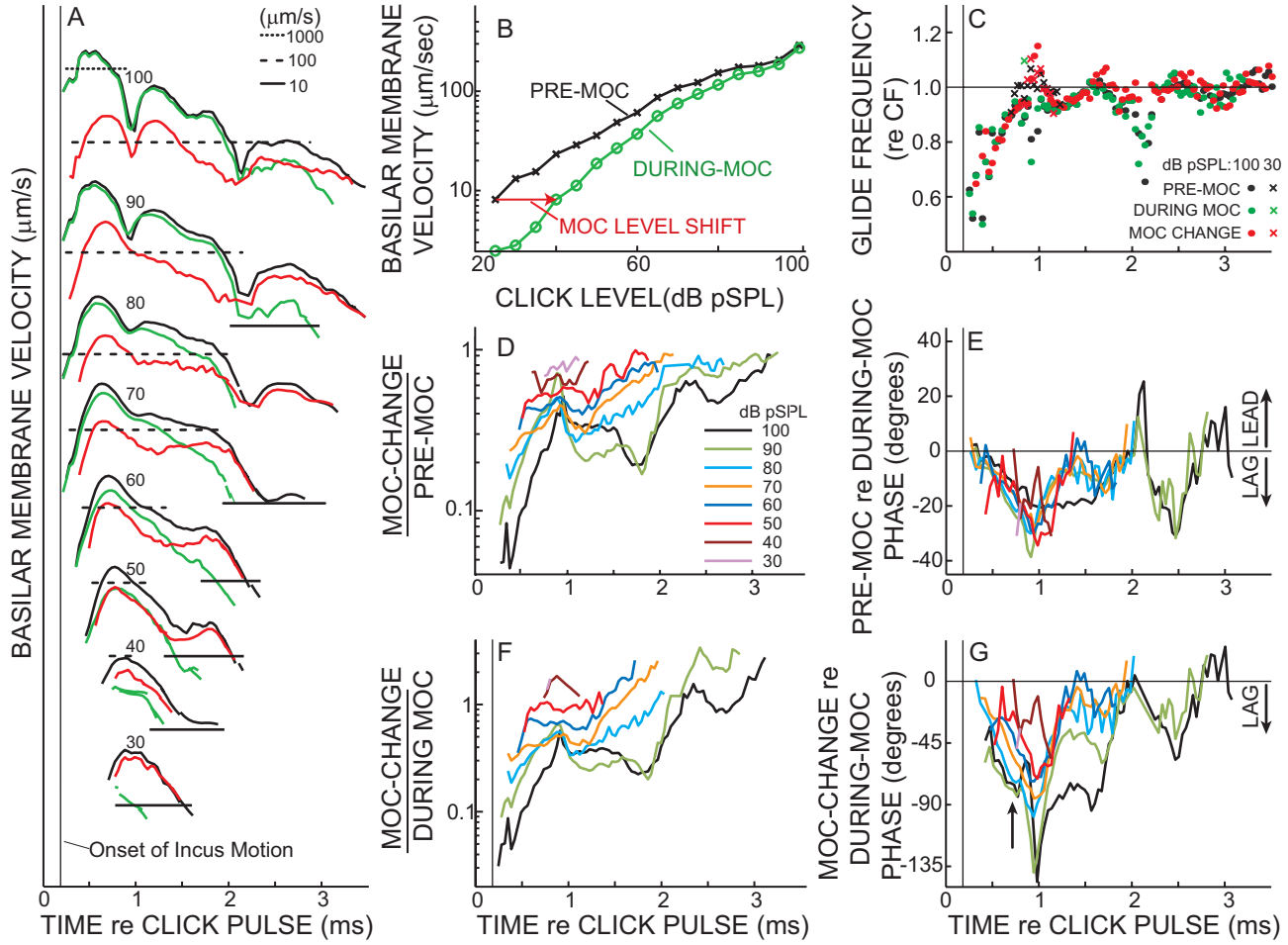


FIG. 3. Analyses of MOC effects on BM click responses in one experiment. (A) BM click-response amplitudes from short-term FFTs for pre-MOC (black), during-MOC (green), and MOC-change (red) [same color code applies to (B) and (C)] for eight click levels. The numbers show the click level in dB pSPL. Responses at different levels are shown on the same scale, but are displaced vertically for clarity. The horizontal lines show amplitude registrations (key at top). (B) Pre-MOC (crosses) and during-MOC (circles) click-response amplitudes vs click level, showing the level shift produced by MOC stimulation (arrow). The click-response amplitude is the BM velocity at the frequency with the most energy in the FFT of the entire response [see Fig. 7(a) for full spectra]. (C) Glide frequency as a function of time for the highest and lowest click levels (see key). (D) Relative inhibition as a function of time, as shown by the MOC-change amplitude divided by the pre-MOC amplitude, both derived from short-term FFTs (see See. II) with click level as a parameter [color key at right also applies to (E)–(G)]. (E) The phase of the pre-MOC response relative to the during-MOC response with click level as a parameter. (F) Relative inhibition shown by the normalized MOC change, which is the MOC-change amplitude divided by during-MOC amplitude (during-MOC chosen instead of pre-MOC because it is closer to the passive response). (G) The phase of the MOC change relative to the during-MOC response, with click level as a parameter. The arrow indicates the time when the phase difference at the highest levels best shows the phase of traveling-wave amplification (see Sec. IV). GP3144. CF=13.5 kHz.

tude of the click response, but sometimes briefly increased the instantaneous amplitude near an envelope minimum [Fig. 3(a)]. The biggest percentage reductions (i.e., the MOC change as percentage of the pre-MOC response) occurred towards the declining end of the responses, primarily at low sound levels [Fig. 2 (right), and Figs. 3(d) and 3(f)], but the biggest absolute reductions (i.e., the absolute values of the MOC change) occurred at high click levels near the peaks of the responses [Fig. 2 (left), and Fig. 3(a) red lines]. We did not see any "dc" base line shifts that could be attributed to MOC effects, although we were only measuring MOC "fast effects" so any base line shifts accompanying MOC "slow effects" would not have been seen. Artifactual base line shifts were sometimes induced by shock-evoked animal motion, but these were effectively removed by expressing the motion as BM velocity.

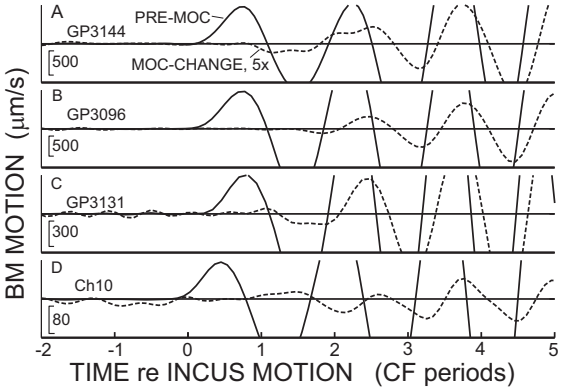


FIG. 4. Expanded views of the BM click-response onsets in three guinea pigs and one chinchilla (rows), illustrating the absence of significant MOC changes in the first half-cycle of the response. Each row shows the pre-MOC response (solid line) and the MOC-change (dashed line), which is multiplied by 5. In each row, the amplitude scales at the bottom left are for the pre-MOC responses and the animal is indicated in the top left. The CF's (top to bottom), are 13.5, 18, 12, and 8.5 kHz.

C. Inhibition near response onsets

Since ANIP inhibition occurs primarily at the onset of AN responses, we looked particularly closely for MOC inhibition near the onset of BM click responses. No significant inhibition was seen during the first half-cycle of the BM responses, and after that the inhibition grew progressively, at least until the first minimum in the envelope of the response (Figs. 2–4). The MOC inhibition started near zero at the beginning of the response, whether viewed as the MOC change [Fig. 3(a) red lines] or as MOC change normalized to the during-MOC response [Fig. 3(f)] (the during-MOC response was used for normalization because it is closer to the passive response than the pre-MOC response). For comparison, in Fig. 3(d) we also show the MOC change normalized by the pre-MOC response. The values for the first half-cycles are missing in Fig. 3 because none of the MOC-change responses were bigger than our criterion of 1.3 times the base line variation of the responses. The beginning parts of the responses from three guinea pigs and the chinchilla are shown on an expanded scale in Fig. 4. Significant MOC changes do not occur until the third or fourth half-cycle of the responses. The data show that, within the resolution of the experiment (a few percent of the pre-MOC response), there is no change on the first half-cycle of the response. Thus, there is no MOC-induced change in basal-turn BM motion that corresponds to the MOC inhibition of the ANIP response.

Since the effect of MOC stimulation on ANIP responses was much greater for rarefaction than condensation clicks (Guinan *et al.*, 2005), we looked closely throughout the BM click responses for differences in the responses to rarefaction versus condensation clicks. Other than the fact that the responses had opposing polarities, we were unable to find any differences between the responses to condensation and rarefaction clicks.

D. Inhibition in the declining part of the click response

During the declining part of click responses, MOC stimulation increased the overall decay rate of the responses [with greater effects when the responses reached low levels, as shown in Figs. 3(a) and 5] although this pattern was complicated somewhat by the responses' waxing and waning. The MOC-induced decrease depended primarily on the instantaneous level of the response rather than the click level (Fig. 5). Put another way, no matter whether the response was evoked by a high-level click and monitored several milliseconds after the click, or by a low-level click and monitored near its peak, whenever the instantaneous value of the pre-MOC click-response amplitude reached a given value [e.g., 30 μm/s in Fig. 5(b)] the value of the during-MOC response measured at the corresponding time and click level was approximately the same [e.g., 10 μm/s in Fig. 5(b)]. Aberrations from this trend were mostly related to the response waxing and waning [Fig. 5(a), inset]. In some animals, the inhibition appeared to increase when the response

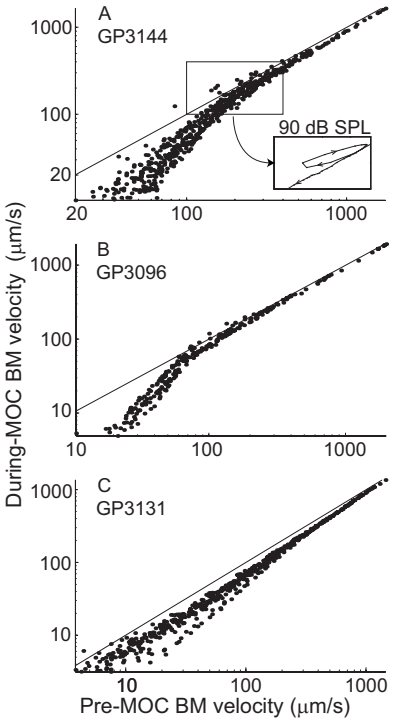


FIG. 5. BM velocity during MOC stimulation vs before MOC stimulation for responses to clicks in three guinea pigs. Each panel shows data from all click levels superimposed. Note that, in each animal, the points from different click levels fall approximately along the same locus. The inset in (A) shows the data from 90 dB SPL clicks to illustrate the pattern across the period of an envelope minimum. The aberrations induced by such response waning ranged from a loop, as shown in the inset in (A), to a small jump to the side (not illustrated). The diagonal lines illustrate equality, where the during-MOC and pre-MOC velocities are the same.

decayed to a certain amplitude of BM motion [Fig. 5(b)] whereas in others the inhibition changed only gradually with BM-motion amplitude [Fig. 5(c)].

MOC stimulation had little effect on either the timing or the prominence of waxing and waning in the click responses [Fig. 5(a)]. As shown by the normalized MOC change, the inhibition had regions of growth and regions of relative constancy, especially at moderate-to-high click levels [Fig. 3(f) also see Figs. S1(F), and S2(F)]. Furthermore, similar "plateau" levels of inhibition were reached across a variety of click levels [e.g., see the plateaus at fractional changes of 0.1–0.4 in Figs. 3(F), S1(F), and S2(F)]. These regions of approximately constant (relative) inhibition occurred despite substantial variation in the amplitude of the underlying motion {e.g., almost an order-of-magnitude variation in click-response amplitude at the highest level in GP3131 [Fig. S2(F)], and even greater variations across click levels–compare Figs. 3(a) and 3(f)}. This finding does not seem to fit with the common picture that the gain of the traveling-wave amplifier increases steadily as the motion decreases, for BM response amplitudes that show compressive nonlinearity. However, this view came from BM responses to tones and from the whole of click responses (e.g., Recio *et al.*, 1998), whereas here we are looking at short periods within the decay of click responses, decays which show prominent waxing and waning. Despite the periods of constant versus increasing relative inhibition, if the dispersion introduced by

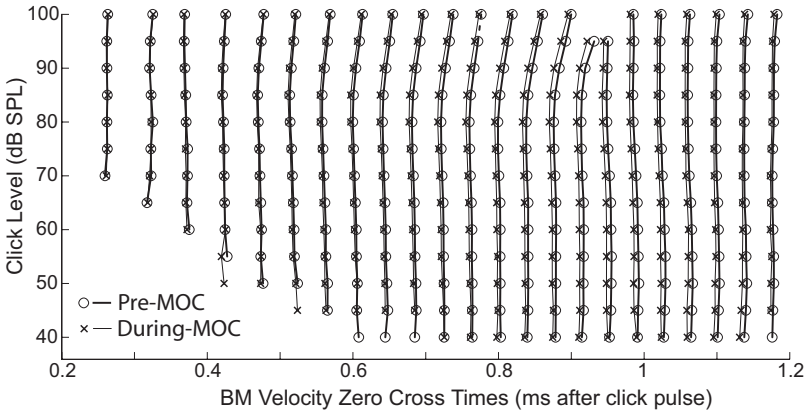


FIG. 6. Click-response zero-crossing times before (circles and thick lines) and during (crosses and thin lines) MOC effects for click levels at 5 dB intervals. Note that the pattern of changes produced by MOC stimulation does not correspond to the pattern of changes produced by changing the sound level. Guinea pig 3144.

the waxing and waning is ignored, the overall trend is for more inhibition at lower sound levels. Since the inhibitory plateaus are most prominent at high click levels, one might think that constant MOC inhibition is produced when the response amplitude is above some critical level. However, responses to the highest-level clicks in GP3144 and GP3131 [Figs. 3(f) and S2(F)] show a second plateau later in the response, when the response amplitudes were much smaller. What causes waxing and waning and regions of constant versus increasing inhibition remains to be elucidated.

E. MOC-evoked changes in response phase

The phase changes evoked by MOC stimulation are of considerable interest because they can provide insight into how traveling-wave amplification is accomplished. MOC stimulation produced substantial changes in the phases of click-evoked BM responses [Figs. 3(e) and 3(g); also see 6, S1(E), S1(G), and S2(E), and S2(G)], but these phase changes generally occurred over many cycles so that the response frequencies were little changed. MOC stimulation therefore produced little change in onset glide frequencies or in click-skirt frequencies [Figs. 3(c), S1(C), and S2(C)]; some details of these changes are presented in the Supplementary Material. Figure 3(g) shows the phase of the MOC change relative to the during-MOC response. We used the during-MOC response as the reference, instead of the pre-MOC response, because it is closer to the passive response and normalizing in this way is better for showing the changes in the phase of traveling-wave amplification relative to the passive drive that elicited the amplification (see Sec. IV).

MOC-induced changes in phase varied widely with click level and the time after the click, especially near any minima in the response envelopes [Figs. 3(f) and 3(g)]. The most consistent phase changes occurred during the period from the response onset to just before the first minimum in the response envelope. During this period, the phase of the normalized MOC change grew from near zero to a lag of 70°–90° at the highest sound levels, with the change being less at lower sound levels [Fig. 3(g), arrow]. The finding of variation in the MOC-change phase with level might appear to conflict with the generalization that the click-response zero crossings do not change with level in the onset glide (de Boer and Nuttall, 1997; Recio *et al.*, 1998; Shera, 2001b). However, this generalization is only an approximation (e.g., Fig. 6). Zero crossings at high levels that occur shortly after the start of the response are delayed relative to those at low levels in guinea pigs (Fig. 6), chinchillas, and cats (Recio *et al.*, 1998; Recio and Rhode, 2000; Lin and Guinan, 2000). Note that the patterns of the zero-crossing delays with increasing level versus the advances with MOC stimulation are quite different. The MOC-induced changes are relatively level insensitive whereas the level-induced changes include both small advances as level increases at low-to-moderate levels (up to 80 dB in Fig. 6) and larger delays at high levels (above 80 dB in Fig. 6).

The MOC-induced phase change can be seen in a different way from the comparison of the pre-MOC and during-MOC phases [Fig. 3(e)]. Before the first minimum in the response envelope, the pre-MOC phase consistently lagged the during-MOC phase with the lag starting near zero and growing over time to reach 20–30° at about the time of the first envelope minimum. This pattern changed little with sound level [Figs. 3(e) and 6]. Overall, in both sets of phase data [Figs. 3(e) and 3(g)], the phase change started near zero and increased until about the time of the first minimum. Since there is little traveling-wave amplification at the response onset, where the dominant frequency is far below CF, the lack of a change in the phase near the onset is not surprising. After the onset, the MOC-induced phase changes built up approximately in parallel with the glide frequency approach to CF [Fig. 3(c)]. In the declining part of the response after the first envelope minimum, the phase change decreased and generally approached zero in a pattern that varied widely across level and across animals.

F. Spectra, and spectral changes in BM click responses

Before MOC stimulation, the spectra of our BM click responses [Fig. 7(a), also see S3(A) and S3(C)] were very similar to those described for the chinchilla by Recio *et al.* (1998). At low levels, the spectra were centered near CF and were relatively smooth functions of frequency, but at higher levels the energy spread out in the low-frequency direction, the peak energy moved to below CF, and deep dips appeared near and above CF. During MOC stimulation, the overall spectra of BM click responses looked similar to the spectra

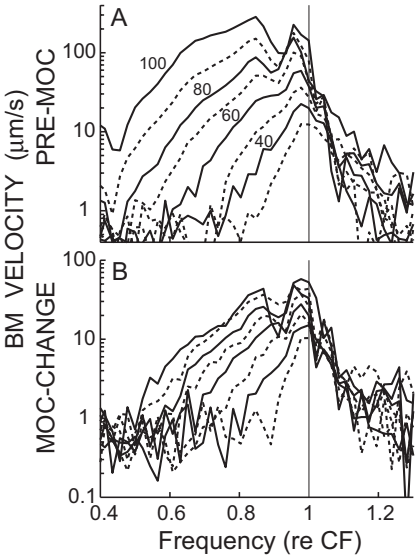


FIG. 7. The spectra of click responses (top) and of the MOC change (bottom) for guinea pig 3144. The lines are for click levels at 10 dB intervals, alternating solid and dashed, with the numbers indicating the pSPLs of the adjacent solid lines.

before MOC stimulation, but there were systematic differences between these as shown by the MOC-change spectra.

The spectra of the MOC changes [Fig. 7(b)] resembled previously described patterns of MOC-induced changes in BM responses to tones (reviewed by Cooper and Guinan, 2006). The largest changes (relative to the pre-MOC responses) were near CF and at low sound levels. The MOC-evoked changes increased in absolute terms, but decreased in a relative sense as sound level increased, with complicated effects at high levels, particularly above CF (Fig. 7). Interestingly, the MOC changes extended to almost an octave below CF [i.e., to near 0.5 CF in Fig. 7(b)], although the changes at these low frequencies were only small fractions of the pre-MOC responses.

IV. DISCUSSION

Our major findings are as follows: (1) MOC stimulation inhibited BM click responses in a time- and intensity-dependent manner (Figs. 2–4). (2) No inhibition was seen during the first half-cycle of BM click responses (Fig. 4) (in stark contrast to the strong inhibition observed in recent ANIP studies). (3) MOC stimulation caused the click responses to decay faster than normal, particularly at low amplitudes of BM motion (Fig. 5), but had little effect on the timing of waxing and waning [Figs. 2 and 3(a)]. (4) MOC-evoked changes in click-response spectra were largest near CF at low click levels, but spread well below CF at high click levels (Fig. 7). (5) The initial part of during-MOC click responses showed a small ($\leqslant 30°$) fine-structure phase advance relative to pre-MOC click responses, with the result that the MOC-change phase-lagged the during-MOC response by 70°–90° at the highest levels [Figs. 3(e), 3(g), and 6].

A. MOC effects in the basal turn

The MOC effects on click responses in the basal turn show exactly what would be expected for MOC inhibition of amplification of the classic traveling wave, but did not show any correlate of MOC inhibition of the ANIP response found by Guinan *et al.* (2005), which is contrary to what would be expected if the classic traveling wave continued to the apex and produced the ANIP response. As expected for MOC inhibition of the classic traveling wave, the inhibition was greatest (in a relative sense) at low click levels, where the whole response was inhibited. At higher click levels, the pattern of MOC inhibition (none during the first half-cycle, growing over the next few cycles, and waxing and waning later) was similar to the pattern of traveling-wave amplification predicted from the nonlinearities of click responses (Recio *et al.*, 1998). In tone responses, MOC inhibition is greatest at low levels near CF and decreases for frequencies above and below CF (reviewed by Cooper and Guinan, 2006). A similar pattern was seen in the spectra of MOC effects on click responses, when the MOC change is considered relative to the pre-MOC response (Fig. 7). The MOC-change spectra are above the noise floor down to an octave below CF where they are an order-of-magnitude less than the pre-MOC response (Fig. 7, see S3). This click-response MOC change extends to further below CF than has been reported for MOC effects on tone responses (Dolan *et al.*, 1997; Murugasu and Russell, 1996; Cooper and Guinan 2006). In part, this is due to the fact that the MOC change includes MOC effects on both the magnitude and the phase of the responses, whereas only magnitude changes were considered for MOC effects on tone responses However, MOC effects of 1 dB or less on the magnitude of tone responses are not easily distinguished from noise and could have been overlooked in previous studies.

In contrast to the clear MOC inhibition of traveling-wave amplification, there was no inhibition of the first peak of the click response that might correspond to the MOC inhibition of the ANIP response that has been found in the middle and apical cochlea by Guinan *et al.* (2005). Guinan *et al.* (2005) could not determine if ANIP inhibition was present in the cochlear base because basal-turn AN responses have inadequate synchrony compared to the local CF. However, basal-turn AN responses show MOC inhibition at certain tail frequencies and this inhibition may be a tone correlate of ANIP inhibition (Guinan *et al.*, 2005). Previous studies of MOC effects on BM tone responses have reported finding no MOC inhibition at tail frequencies (Dolan *et al.*, 1997; Murugasu and Russell, 1996; Cooper and Guinan, 2006); however, these studies did not focus on looking for the BM tail-frequency inhibition that corresponds to the neural tail-frequency inhibition, which is more than an octave below CF, and they could easily have missed small, frequency-specific tail effects. Furthermore, the signal-to-noise ratio of laser-based BM measurements is considerably lower at tail frequencies than at basal-turn CFs. Thus, whether or not there is MOC inhibition of BM motion at tail frequencies more than one octave below CF is still an open question. In summary, in the cochlear base, MOC inhibition

of BM motion appears to be due only to inhibition of traveling-wave amplification and a BM-motion correlate of the ANIP motion has not been found.

B. The origin of waxing and waning

The waxing and waning in the click-response envelope appear to be due to beats between vibrations at two near-CF frequencies (Recio *et al.*, 1998; Lin and Guinan, 2000), but little is known about these vibrations. The waxing and waning period was not changed by MOC stimulation, which suggests that the frequencies involved were not changed. Note that two near-CF resonances were found in the BM impedance calculated from an inverse solution based on BM responses to tones (Zweig, 1991). If one presumes that there are two resonances that are independent and are excited with the same phase at response onset, then the two resonances should become out of phase and cancel at times: $T_c=(2n+1)\pi/\Delta F$, where n is an integer, and ΔF is the difference in frequency between the two resonances. Thus, the first cancellation should be after one unit of time, with subsequent cancellations at intervals of twice this. This is not the pattern of BM click responses (Figs. 2 and 3, S1, and S2). Hypotheses that fit the data better are that the resonances are excited in opposite phases or that they are strongly coupled and exchange energy.

Interference of two motion components has also been seen in high-level tone data as response dips in BM motion (Rhode and Recio, 2000; Guinan and Cooper, 2003; Rhode, 2007). MOC stimulation inhibits one of these components more than the other (Guinan and Cooper, 2003) in a way that accounts for the MOC-induced *increase* in BM motion first noted by Dolan *et al.* (1997). It might seem that the two factors that produce the tone-response dips also produce the click-response waxing and waning. Arguing against this possibility are the following: (1) The tone-response dips are seen only at high levels whereas the click-response waxing and waning can also be found at low levels. (2) Tone-response dips are seen each time the traveling-wave phase crosses the same value (plus or minus an integer number of cycles), indicating that the non-traveling-wave component has a very high velocity and may be an evanescent or compression wave (Cooper and Rhode, 1996; Dong and Cooper, 2006; Rhode, 2007). Since evanescent waves are not generally "tuned" to a particular CF, they are highly unlikely to produce the waxing and waning in the click response. A more likely solution is that other forms of tuned motion exist within the organ of Corti and TM complex, perhaps motions similar to those that have been observed in various *in vitro* preparations (e.g., Nowotny and Gummer, 2006; Karavitaki and Mountain, 2007a). The most general conclusion that can be drawn from this is that both tone and click data do not fit with the idea that a single traveling wave excites a single mode of BM and organ of Corti vibration, even in the base of the cochlea.

C. Phase changes and traveling-wave amplification

The phases of BM click responses were changed in complex ways by MOC stimulation, but at least some aspects of these changes can be understood. For instance, it might seem surprising that MOC stimulation decreased the phase (i.e., decreased the latency of the waveform fine structure) of the BM click response, considering that MOC stimulation *increases* the latency of AN CAPs (Gifford and Guinan, 1987). However, the BM latency change can be understood as being consistent with the hypotheses that (1) traveling-wave amplification produces a delay in the BM click-response wave form and (2) the MOC reduction of this amplification reduces this delay, i.e., produces a phase advance. It seems likely that the decreased delays of human transient-evoked otoacoustic emissions produced by MOC activity evoked by contralateral sound (Ryan *et al.*, 1991; Giraud *et al.*, 1996) were due to decreases in BM response delays similar to those found here.

One motivation for looking at MOC-induced changes in BM response phase is that they may give clues to the processes of cochlear amplification. The best estimates of the phase of traveling-wave amplification come from mathematical "inversion" of BM measurements (e.g., de Boer and Nuttall, 2000), which show that traveling-wave amplification for a tone is due to a cochlear-partition impedance that has negative damping (i.e., energy injection in phase with the motion) just basal to the BM CF place. In contrast, the impedance of the cochlear partition at the CF place has positive damping (energy absorption). In a simplified vector explanation, in response to pressure, the negative damping impedance would produce a response phase of −180°, while the stiffness-dominated passive response would have a phase of −90°. Thus, in this simplified view, the response component from cochlear amplification would lag the passive response by 90°.

How do we compare the results of this mathematical inversion based on tone responses with our click data? In an attempt to glean what we can about cochlear amplification phase from our click responses, we restrict our attention to the most telling part of the response. First, we consider only high-level responses where the during-MOC response is most heavily influenced by the passive response, but the MOC change still represents removal of the contribution from the cochlear amplifier. Second, we restrict consideration to times when the glide frequency is at or near CF, but avoid comparisons of phases during the decaying portion of the click responses because these are complicated by the waxing and waning of the response [the phase of the main Fourier component of the response almost reverses at a deep minimum, as illustrated by Recio *et al.*, 1998]. Under these conditions [shown by the arrow in Fig. 3(g)], the phase of the normalized MOC change was a lag of 70°–90°, which is similar to the 90° predicted by our simplified analysis that presumed cochlear amplification is produced by negative damping. Although it focuses on only one part of the click response, this analysis of the MOC change provides an independent measure consistent with the hypothesis that traveling-wave amplification is accomplished by negative damping of the cochlear partition.

D. Implications for cochlear mechanics in the middle and apical cochlea

Our finding of almost no MOC inhibition of the earliest cycle of the basal-turn BM click response contrasts with the strong inhibition found in the ANIP responses in the middle and apex of the cochlea (Guinan *et al.*, 2005). We presume that this BM-AN difference is not due to the species difference (guinea pig versus cat), but this must be checked in future work. Another possible difference is that the BM data were from somewhat damaged preparations whereas the AN data were from undamaged preparations. This seems unlikely to explain the BM-AN difference because we saw no change in the pattern of BM effects with different degrees of damage, and because there was less than a 10 dB threshold deterioration in some animals (e.g., guinea-pig 3144 and chinchilla 10) so the condition of these animals was similar to the condition for much of the AN data.

Several other factors can be ruled out as candidates for producing the ANIP inhibition. A small (~1 dB) inhibition at tail frequencies is produced by the MOC-induced decrease in endocochlear potential (EP), but this change in EP cannot account for the ANIP inhibition because it is too small and cannot act only on the first peak (similarly, it cannot account for the 10 dB inhibition of AN responses at certain tail frequencies because it is too small and not frequency selective; Stankovic and Guinan, 1999). Alternately, it has been suggested that MOC activation produces a stiffness change in OHCs (He and Dallos, 2000; but see Hallworth, 2007) and this stiffness change might be thought to produce the ANIP inhibition. However, the putative OHC stiffness change occurred slowly and appeared to be associated with MOC slow effects, which our methods exclude. Furthermore, any fast stiffness change would be expected to alter the first cycle of the BM response, which we did not see. In addition, any MOC-evoked stiffness change would be present throughout the click response and should affect much more than just the first cycle of the responses: in particular, the characteristic frequency of the click response would be affected strongly, which is inconsistent with our data.

There are two important differences in the AN versus BM data: (1) the CF difference, i.e., the ANIP response was found in the apical half of the cochlea (CFs < 6 kHz) whereas the BM click responses were measured in the basal turn, and (2) the measurement location, i.e., AN versus BM. We note that nothing corresponding to the ANIP inhibition was seen in BM responses at the chinchilla 8.5 kHz place [Fig. 4(d)], a frequency only slightly above 6 kHz where ANIP inhibition was found in the cat. Unfortunately, there are no data available for MOC effects on BM motion in the middle or apical part of the cochlea. Furthermore, for the middle or apical part of the cochlea, there are no BM data from preparations that have been demonstrated to have normal thresholds (the lowest CF with good BM data is 6 kHz—Rhode, 2007).

E. The classic traveling-wave view is not sufficient to account for the ANIP data

The common conception of cochlear mechanics throughout the cochlea is an extrapolation of BM motion in the basal turn where it can be measured and the normality of the preparation can be checked with AN CAPs. In models, the classic traveling wave is the motion of the cochlear partition produced by a slow, apically moving, sound-frequency pressure difference across the cochlear partition acting on the impedance of the cochlear partition (e.g., see Patuzzi, 1996). Based on the gradual change in cochlear anatomy from base to apex, classic cochlear models assume that the traveling wave exists, and scales with the local CF, throughout the cochlea. In this conception, the traveling wave produces all of the excitation of the AN, with little or no room for any intervening filtering or active processes (e.g., Patuzzi, 1996; Narayan *et al.*, 1998). However, this conception is not compatible with the finding that there is no MOC inhibition of the first peak of basal-turn BM motion.

If a frequency-scaled version of the traveling wave shown by basal-turn BM motion contained all of the motion that produces AN firing throughout the cochlea, then the first peak of the classic traveling wave must be the drive for the ANIP response, and the MOC inhibition of the ANIP response would be present throughout the cochlea. This follows because with the classic traveling-wave hypothesis, the form of BM motion in the base is the form of the traveling wave that must exist throughout the cochlea. Thus, if there is no MOC inhibition of the first peak of the traveling wave in the base, then the same is true throughout the cochlea. However, since we see MOC inhibition of the first peak of the AN response in the middle and apex of the cochlea, then the classic traveling wave cannot be directly responsible for the ANIP response. We infer that some other motion must produce the ANIP response.

Note that with the classic-traveling-wave hypothesis, the earliest peak of the neural click response must come from the first peak of the traveling-wave click response, because the amplitude of the traveling-wave response grows gradually from one peak to the next. If the second or a later traveling-wave peak excites a neural response, then the first peak of the traveling wave will excite a response simply by using a higher click level. Similarly, if one supposes that the ANIP response is *not* produced by motion that forms a part of the traveling wave, then this means it must come from some other motion. Either way, we conclude that some motion other than the traveling wave must produce the ANIP response.

F. Could the traveling-wave change from base to apex enough to account for the data?

One might argue that the traveling wave changes in form from base to apex and this change in form is sufficient for the traveling wave in the middle and apex to account for the ANIP inhibition. For this to be true, the first peak of the traveling wave must change enough to be inhibited by MOC stimulation, but the rest of the traveling wave must stay sufficiently the same to account for the other aspects of the apical AN responses, which show the expected properties of the classic traveling wave. Traditional cochlear models show that the first peak of the traveling wave is from below-CF energy and is passive (e.g., Shera, 2001a). Such cochlear models would need serious revision to accommodate a "trav-

eling wave" that shows MOC inhibition of the first peak. While it cannot be said that this "proves" that the traveling wave could not change enough from the classic traveling wave to make ANIP inhibition possible, the necessity of a drastic change in the first peak (or first two half-cycles) but little or no change in the rest of the traveling wave poses a great difficulty for this hypothesis. The alternative, that there is a non-traveling-wave motion that produces the ANIP response, is far more attractive. The conceptual framework for this conclusion is that cochlear motion is the sum of vibrational modes: The traditional traveling wave is one vibrational mode and the inferred ANIP motion is another. Since there is little evidence for the ANIP motion in the base, we assume that in the base the traditional traveling wave is the dominant motion, but as one moves apically, the role of the ANIP motion gradually increases. Thus, although not compelling, the most parsimonious explanation of the data is that in addition to the classic traveling wave (which exists in scaled forms throughout the cochlea—Patuzzi, 1996), there is another motion, the ANIP motion, that is negligible (or not discernable) in the base and grows in prominence towards the apex.

With this conceptual framework, the BM motion in the apex may, or may not, show the ANIP motion, depending on whether the ANIP motion vibrational mode includes significant movement of the BM. Mechanical measurements of click responses in the cochlear apex show weakly nonlinear growth of the first peak (Cooper and Rhode, 1996), which may have the same origin as the ANIP motion, a conjecture that awaits experimental verification.

If both the traveling wave and the inferred ANIP motion contribute to BM motion in the apical half of the cochlea, then the nonlinearity seen in BM motion will depend on both motions. Strong MOC inhibition of traveling-wave amplification occurs when there is strong nonlinearity in the BM input-output function. If this is also true for the ANIP motion, then the strong ANIP inhibition may indicate that there is strong nonlinearity in the growth of the ANIP motion. If the first peak of the apical BM motion is a combination of a linear first-peak response to the classic traveling wave and a highly nonlinear ANIP motion vibrational mode, then the strong ANIP motion nonlinearity might be considerably watered down by the linear first peak of the traveling wave, when seen in BM motion.

G. Two cochlear amplifiers?

The MOC inhibitions of the classic traveling wave and the inferred ANIP motion appear to be separate phenomena, and raise the question of whether there are actually two mechanical amplifiers in the mammalian cochlea. Traveling-wave amplification is shown by basal-turn BM motion and AN responses throughout the cochlea, and all aspects of this amplification appear susceptible to MOC inhibition. The ANIP motion, inferred from AN responses in the apical half of the cochlea, is also strongly inhibited by MOC stimulation, which indicates that it comes from, or is strongly influenced by, OHCs. Since these are due to two separate motions (as argued above), they reveal the presence of two separate ways in which OHCs influence cochlear motions that drive IHC stereocilia and excite auditory nerve fibers. If the inferred ANIP motion is produced or enhanced by the action of OHCs, it could be called a second cochlear amplifier. The MOC inhibition of this motion is consistent with such a hypothesis. However, whether OHCs amplify, or simply influence, the inferred ANIP motion remains to be determined.

With two modes of OHC-influenced motion that excite AN responses, and two known forms of OHC motility, a natural question is whether each form of motility primarily affects one type of motion. One possibility is that somatic motility produces traveling-wave amplification, and hair-bundle motility amplifies or influences the inferred ANIP motion. Somatic motility extends well beyond 50 kHz (Scherer and Gummer, 2004), high enough in frequency to provide basal-turn amplification. Hair-bundle motility shows asymmetry with the direction of motion, at least in nonmammalian vertebrates (Jaramillo and Hudspeth, 1993; Ricci *et al.*, 2000), potentially making it able to amplify or influence the inferred ANIP motion and produce more inhibition for rarefaction clicks than for condensation clicks. However, the ANIP asymmetry may have other origins (e.g., in IHC mechanoelectrical transduction, or in the IHC-AN synapse) and may not require an asymmetric mechanical drive. Alternately, it may be that the two forms of OHC motility may interact (Kennedy *et al.*, 2006). More evidence is needed to determine the relationship between the forms of OHC motility and the traveling wave and ANIP motions.

H. What is the ANIP motion?

There are several possibilities for what the inferred ANIP motion might correspond to physically, but no data that reveal its identity. Two possibilities for the ANIP motion that involve OHC somatic motility are fluid flow past IHC stereocilia due to reticular-lamina tilting (Nowotny and Gummer, 2006), and/or motion from fluid flow along the tunnel of Corti due to OHCs squeezing the organ of Corti (Karavitaki and Mountain, 2007a). Another possibility for the ANIP motion is radial motion of the TM (Hemmert *et al.*, 2000), which has recently been shown to propagate longitudinally along the TM (Ghaffari *et al.*, 2007). TM radial motion might originate from, or be influenced by, stereocilia motility. The ANIP motion may arrive at a given cochlear place before the classical traveling wave (Guinan *et al.*, 2005) but this does not help us distinguish the origin of the ANIP response because either fluid flow along the tunnel of Corti or fast propagation of a radial shear wave along the TM might account for this. In summary, work in excised preparations shows several motions that may correspond to the ANIP motion, but there are no data tying any of these to the ANIP motion, so the exact identity of the ANIP motion, and whether a component of it is apparent in BM motion, are unknown.

I. What have we learned?

Previous data show cochlear motions other than the classic traveling wave and indicate that cochlear scaling breaks down in the apex (e.g., Nowotny and Gummer, 2006;

Karavitaki and Mountain, 2007a,b; Shera and Guinan, 2003), so one might ask the following: "What is new here?" The breakdown of scaling in the apex is usually thought of as indicating that the apical traveling wave loses its basal-turn form, not, as indicated here, that a second motion becomes important enough to drive AN firing. Further, the breakdown of scaling is usually thought to be restricted to the extreme apex (in the cat, a 1 kHz transition point is typically given), whereas the data of Guinan *et al.* (2005) indicate that the ANIP motion extends at least up to 6 kHz, which includes the middle and apex of the cochlea. Although work in excised preparations has shown that the cochlea can support non-traveling-wave cochlear motions (at least when driven electrically), the importance of these motions in the intact cochlea is unknown. This paper, plus the data of Guinan *et al.* (2005), provides the first credible arguments from intact, normally functioning, acoustically stimulated cochleae, that a non-traveling-wave motion that excites AN fibers is produced or amplified by the action of OHCs. In summary, the present manuscript (1) rules out that the classic traveling wave can fully explain cochlear mechanics in the middle and apex of the cochlea, (2) puts strong constraints on models that attempt to explain both basal and apical cochlear mechanics, and (3) provides evidence from normal cochleae that strongly points to there being two modes of cochlear excitation affected by active processes, thus raising the possibility that there is another "cochlear amplifier" in addition to the classic traveling-wave amplifier.

ACKNOWLEDGEMENTS

We thank Dr. M.C. Liberman and Dr. C. Shera for comments on an earlier version of the manuscript. Supported by NIDCD RO1DC00235, the Royal Society, and Deafness Research UK.

NOMENCLATURE

AN	=	Auditory nerve
ANIP	=	Auditory nerve initial peak
BM	=	Basilar membrane
CAP	=	Compound action potential
CF	=	Characteristic frequency
during-MOC	=	During the MOC effect
EP	=	Endocochlear potential
FFT	=	Fast Fourier transform
IHC	=	Inner hair cell
MOC	=	Medial olivocochlear
MOC-change	=	Pre-MOC minus during MOC
OHC	=	Outer hair cell
pre-MOC	=	Before MOC effect
pSPL	=	Peak equivalent SPL
SPL	=	Sound pressure level
TM	=	Tectorial membrane

Cooper, N. P. (**1999**). "An improved heterodyne laser interferometer for use in studies of cochlear mechanics," J. Neurosci. Methods **88**, 93–102.
Cooper, N. P., and Guinan, J. J., Jr. (**2003**). "Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity," J. Physiol. (London) **548**, 307–312.
Cooper, N. P., and Guinan, J. J., Jr. (**2006**). "Efferent-mediated control of basilar membrane motion," J. Physiol. (London) **576**, 49–54.
Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.
Cooper, N. P., and Rhode, W. S. (**1996**). "Fast travelling waves, slow travelling waves and their interactions in experimental studies of apical cochlear mechanics," Aud. Neurosci. **2**, 289–299.
Dallos, P., He, D. Z., Lin, X., Sziklai, I., Mehta, S., and Evans, B. N. (**1997**). "Acetylcholine, outer hair cell electromotility, and the cochlear amplifier," J. Neurosci. **17**, 2212–2226.
Dallos, P., Zheng, J., and Cheatham, M. A. (**2006**). "Prestin and the cochlear amplifier," J. Physiol. (London) **576**, 37–42.
de Boer, E., and Nuttall, A. L. (**1997**). "The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592.
de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform of the basilar membrane. II. From data to models—and back," J. Acoust. Soc. Am. **107**, 1487–1496.
Dolan, D. F., Guo, M. H., and Nuttall, A. L. (**1997**). "Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation," J. Acoust. Soc. Am. **102**, 3587–3596.
Dong, W., and Cooper, N. P. (**2006**). "An experimental study into the acousto-mechanical effects of invading the cochlea," J. R. Soc., Interface **3**, 561–571.
Fettiplace, R. (**2006**). "Active hair bundle movements in auditory hair cells," J. Physiol. (London) **576**, 29–36.
Fridberger, A., Widengren, J., and Boutet de Monvel, J. (**2004**). "Measuring hearing organ vibration patterns with confocal microscopy and optical flow," Biophys. J. **86**, 535–543.
Ghaffari, R., Aranyosi, A. J., and Freeman, D. M. (**2007**). "Longitudinally propagating traveling waves of the mammalian tectorial membrane," Proc. Natl. Acad. Sci. U.S.A. **104**, 16510–16515.
Gifford, M. L., and Guinan, J. J., Jr. (**1987**). "Effects of electrical stimulation of medial olivocochlear neurons on ipsilateral and contralateral cochlear responses," Hear. Res. **29**, 179–194.
Giraud, A. L., Perrin, E., Chery Croze, S., Chays, A., and Collet, L. (**1996**). "Contralateral acoustic stimulation induces a phase advance in evoked otoacoustic emissions in humans," Hear. Res. **94**, 54–62.
Guinan, J. J., Jr. (**1996**). "The physiology of olivocochlear efferents," in *The Cochlea*, edited by P. J. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 435–502.
Guinan, J. J., Jr., and Cooper, N. P. (**2003**). "Fast effects of efferent stimulation on basilar membrane motion," in *The Biophysics of the Cochlea: Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 245–251.
Guinan, J. J., Jr., and Stankovic, K. M. (**1996**). "Medial efferent inhibition produces the largest equivalent attenuations at moderate to high sound levels in cat auditory-nerve fibers," J. Acoust. Soc. Am. **100**, 1680–1690.
Guinan, J. J., Jr., Lin, T., and Cheng, H. (**2005**). "Medial-olivocochlear-efferent inhibition of the first peak of auditory-nerve responses: Evidence for a new motion within the cochlea," J. Acoust. Soc. Am. **118**, 2421–2433.
Hallworth, R. (**2007**). "Absence of voltage-dependent compliance in high-frequency cochlear outer hair cells," J. Assoc. Res. Otolaryngol. **8**, 464–473.
He, D. Z., and Dallos, P. (**2000**). "Properties of voltage-dependent somatic stiffness of cochlear outer hair cells," J. Assoc. Res. Otolaryngol. **1**, 64–81.
Hemmert, W., Zenner, H. P., and Gummer, A. W. (**2000**). "Three-dimensional motion of the organ of Corti," Biophys. J. **78**, 2285–2297.
Jaramillo, F., and Hudspeth, A. J. (**1993**). "Displacement-clamp measurement of the forces exerted by gating springs in the hair bundle," Proc. Natl. Acad. Sci. U.S.A. **90**, 1330–1334.
Karavitaki, K. D., and Mountain, D. C. (**2007a**). "Evidence for outer hair cell driven oscillatory fluid flow in the tunnel of corti," Biophys. J. **92**, 3284–3293.
Karavitaki, K. D., and Mountain, D. C. (**2007b**). "Imaging electrically evoked micromechanical motion within the organ of Corti of the excised gerbil cochlea," Biophys. J. **92**, 3294–3316.
Kennedy, H. J., Evans, M. G., Crawford, A. C., and Fettiplace, R. (**2006**). "Depolarization of cochlear outer hair cells evokes active hair bundle motion by two mechanisms," J. Neurosci. **26**, 2757–2766.
LePage, E. L., and Johnstone, B. M. (**1980**). "Nonlinear mechanical behav-

ior of the basilar membrane in the basal turn of the guinea pig cochlea," Hear. Res. **2**, 183–189.
Lin, T., and Guinan, J. J., Jr. (**2000**). "Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.
Lin, T., and Guinan, J. J., Jr. (**2004**). "Time-frequency analysis of auditory-nerve-fiber and basilar-membrane click responses reveal glide irregularities and non-characteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.
Murugasu, E., and Russell, I. J. (**1996**). "The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea," J. Neurosci. **16**, 325–332.
Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). "Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae," Science **282**, 1882–1884.
Nowotny, M., and Gummer, A. W. (**2006**). "Nanomechanics of the subtectorial space caused by electromechanics of cochlear outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **103**, 2120–2125.
Nuttall, A. L., and Dolan, D. F. (**1993**). "Basilar membrane velocity responses to acoustic and intracochlear electric stimuli," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 288–294.
Patuzzi, R. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, edited by P. J. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
Rhode, W. S. (**2007**). "Basilar membrane mechanics in the 6–9 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **121**, 2792–2804.
Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.
Rhode, W. S., and Robles, L. (**1974**). "Evidence from Mössbauer experiments for nonlinear vibration in the cochlea," J. Acoust. Soc. Am. **55**, 588–597.
Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2000**). "Active hair bundle motion linked to fast transducer adaptation in auditory hair cells," J. Neurosci. **20**, 7131–7142.
Robles, L., Rhode, W. S., and Geisler, C. D. (**1976**). "Transient response of the basilar membrane measured in squirrel monkeys using the Mössbauer effect," J. Acoust. Soc. Am. **59**, 926–939.
Ruggero, M. P., and Rich, N. C. (**1990**). "Systemic injection of furosemide alters the mechanical response to sound of the basilar membrane," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, D. B.Matthews, M. Ruggero, and C. R. Steele(Springer-Verlag, Berlin).
Ruggero, M. A., and Rich, N. C. (**1991a**). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., and Rich, N. C. (**1991b**). "Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration," Hear. Res. **51**, 215–230.
Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992a**). "Basilar membrane responses to two-tone and broadband stimuli," Philos. Trans. R. Soc. London, Ser. B **336**, 307–315.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1992b**). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," J. Neurophysiol. **68**, 1087–1099.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1993**). "Alteration of basilar membrane responses to sound by acoustic overstimulation," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 258–265.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1996**). "The effect of intense acoustic stimulation on basilar-membrane vibrations," Aud. Neurosci. **2**, 329–345.
Russell, I. J., and Nilsen, K. E. (**1997**). "The location of the cochlear amplifier: spatial representation of a single tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.
Ryan, S., Kemp, D. T., and Hinchcliffe, R. (**1991**). "The influence of contralateral acoustic stimulation on click-evoked otoacoustic emission in humans," Br. J. Audiol. **25**, 391–397.
Shera, C. A. (**2001a**). "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion," J. Acoust. Soc. Am. **109**, 2023–2034.
Shera, C. A. (**2001b**). "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," J. Acoust. Soc. Am. **110**, 332–348.
Shera, C. A., and Guinan, J. J., Jr. (**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.
Scherer, M. P., and Gummer, A. W. (**2004**). "Vibration pattern of the organ of Corti up to 50 kHz: Evidence for resonant electromechanical force," Proc. Natl. Acad. Sci. U.S.A. **101**, 17652–17657.
Sridhar, T. S., Liberman, M. C., Brown, M. C., and Sewell, W. F. (**1995**). "A novel cholinergic "slow effect" of olivocochlear stimulation on cochlear potentials in the guinea pig," J. Neurosci. **15**, 3667–3678.
Stankovic, K. M., and Guinan, J. J., Jr. (**1999**). "Medial efferent effects on auditory-nerve responses to tail-frequency tones I: Rate reduction," J. Acoust. Soc. Am. **106**, 857–869.
Wilson, J. P., and Johnstone, J. R. (**1975**). "Basilar membrane and middle-ear vibration in guinea pig measured by capacitive probe," J. Acoust. Soc. Am. **57**, 705–723.
Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.

Comparison of behavioral and auditory brainstem response measures of threshold shift in rats exposed to loud sound

Henry E. Heffner,[a] Gimseong Koay, and Rickye S. Heffner
Department of Psychology, University of Toledo, Toledo, Ohio 43606

(Received 15 April 2008; revised 28 May 2008; accepted 29 May 2008)

The purpose of this study was to determine how closely the auditory brainstem response (ABR) can estimate sensorineural threshold shifts in rats exposed to loud sound. Behavioral and ABR thresholds were obtained for tones or noise before and after exposure to loud sound. The results showed that the ABR threshold shift obtained with tone pips estimated the initial pure-tone threshold shifts to within ±5 dB 11% of the time and the permanent pure-tone threshold shifts 55% of the time, both with large errors. Determining behavioral thresholds for the same tone pips used for the ABR did not improve the agreement between the measures. In contrast, the ABR obtained with octave noise estimated the initial threshold shifts for that noise to within ±5 dB 25% of the time and the permanent threshold shifts 89% of the time, with much smaller errors. Thus, it appears that the noise-evoked ABR is more accurate in estimating threshold shift than the tone-evoked ABR.
 [DOI: 10.1121/1.2949518]

PACS number(s): 43.64.Ri, 43.64.Wn, 43.66.Sr, 43.66.Gf [BLM] Pages: 1093–1104

I. INTRODUCTION

The first step in studying hearing loss is to determine the degree of the loss and the frequencies at which it occurs. Although it is usually not difficult to obtain a behavioral audiogram for adult human subjects, this can be a time-consuming process with animals. For this reason, hearing loss in animals is often assessed with a physiological measure, such as the compound action potential or auditory brainstem response (ABR). The question, then, is how accurately do such physiological measures reflect changes in behavioral thresholds?

To determine the accuracy of physiological measures of hearing loss, it is necessary to compare behavioral and physiological measures in the same animals, which four previous studies have done. Three of these studies recorded the neural responses evoked by sound from electrodes either in the cochlea (Dallos *et al.*, 1978) or in the inferior colliculus (Davis and Ferraro, 1984; Henderson *et al.*, 1983). The results of these studies indicated that although the physiological measures agreed closely with some of the behavioral threshold shifts that resulted from ototoxic drugs or exposure to loud sound, there were significant differences with no way to determine which estimates were accurate and which were not. The fourth study compared the ABR with behavioral thresholds before and after exposure to an ototoxic drug or loud sound and found better agreement between the behavioral and physiological measures than the previous three studies, although their results were based on only two animals (Borg and Engström, 1983). Nevertheless, these results suggested that the ABR is a promising technique for assessing hearing loss in animals.

Indeed, the ABR has been used for many years to assess hearing loss in humans, especially in infants and individuals with developmental disabilities who cannot be tested behaviorally. As a result, a number of studies have been conducted to determine the accuracy of the ABR for estimating behavioral thresholds at different frequencies [for reviews, see Gorga (1999), Gorga and Neely (2002), and Stapells (2000a and 2000b)]. In general, the ABR appears to be 10–20 dB less sensitive than pure-tone behavioral thresholds in adults with normal hearing (e.g., Stapells, 2000a, 2000b). Interestingly, for individuals with sensorineural hearing loss, the ABR usually falls within 5–15 dB of the behavioral thresholds (Gorga and Neely, 2002; Stapells, 2000a, 2000b). In other words, the relation between the ABR and behavioral thresholds changes following sensorineural hearing loss. However, it should be noted that such comparisons have only been done for frequencies in the human midrange (500 Hz to 4 kHz), and it has been shown in mice that the ABR diverges significantly from the behavioral thresholds at the high and low-frequency ends of the audiogram (Heffner and Heffner, 2003), an effect also seen in other animals (Finneran and Houser, 2006; Szymanski *et al.*, 1999).

Recently, we have been studying tinnitus in animals caused by exposure to loud sound (Heffner and Harrington, 2002; Heffner and Koay, 2005). In doing so, we have used the ABR to estimate the accompanying hearing loss to determine whether the hearing loss, rather than the tinnitus, is associated with the increased activity in the dorsal cochlear nucleus that occurs following such exposure (Zhang *et al.*, 2004). The purpose of this study, then, was to determine how well the ABR estimates behavioral hearing loss in rats exposed to loud sound. As will be seen, our results indicate that the ABR evoked by octave noise provides a much more accurate estimate of hearing loss than the tone-evoked ABR.

[a]Address correspondence to: Henry E. Heffner, Ph.D. Department of Psychology, University of Toledo, Toledo, OH 43606, Phone: 419/530-2684, FAX: 419/825-1659, E-mail: hheffne@pop3.utoledo.edu.

II. METHODS

Behavioral and ABR thresholds for tones and noise were obtained on monaural rats. For optimal accuracy, behavioral and ABR thresholds were obtained for only one sound at a time. The animals were then exposed to a loud tone for 10 min, followed 1 h later by behavioral and ABR testing to determine the resulting threshold shift. Both thresholds were then tracked over subsequent days until they had stabilized, and the ABR threshold shifts were compared with the behavioral threshold shifts. To avoid potential bias during testing, the behavioral and ABR results for an animal were not compared until testing was complete.

A. Subjects

The subjects were 18 male Long Evans laboratory rats (*Rattus norvegicus*) ranging in age from 70 to 115 days at the beginning of the experiments. The animals had been bred in the Department of Psychology of the University of Toledo and were thus known to have no previous history of exposure to loud sound, such as transportation noise. They were given free access to rodent blocks. Water was available only during the daily training and test sessions. The use of animals in this study was approved by the University of Toledo Animal Care and Use Committee.

B. Surgical procedure

Prior to training and testing, each animal was deafened in its left ear so that all testing was conducted on its right ear. This involved anesthetizing an animal with halothane, removing the left eardrum and middle ear bones, and packing the bulla with a piece of foam rubber earplug (E-A-R Classic earplug, Aearo Corp.) to prevent sound from entering the bulla. The cochlea was purposely left intact to avoid affecting the vestibular system, which could affect behavioral auditory thresholds by causing an animal to hold its head in a tilted position. Subsequent ABR testing failed to reveal any response in the deafened ear to the sounds used in this study.

C. Behavioral apparatus

Testing was conducted in a carpeted, double-walled sound chamber (IAC model 1204; Industrial Acoustics Co., Bronx, NY, USA; $2.55 \times 2.75 \times 2.05$ m), the walls and ceiling of which were lined with egg crate foam. The equipment for behavioral control and stimulus generation was located outside the chamber, and the animals were observed over closed-circuit television.

The animals were tested in a cage (28 cm long $\times$ 13 cm wide $\times$ 16 cm high) constructed with 1 in. (2.54 cm) wire mesh [for a drawing of the test cage, see Heffner *et al.* (1994)]. The cage was mounted on a camera tripod and raised 92 cm above the floor. A water spout (15-gauge stainless steel tubing) was mounted vertically up through the floor of the front of the cage so that it projected 5 cm above the cage floor. An oval brass disk (1.2×2.0 cm) was mounted on top of the spout at a 30 deg angle. This arrangement permitted an animal to lick water off the spout while holding its head in a normal position facing the front of the cage.

The water spout was connected via plastic tubing to a syringe pump (NE 1000, New Era, Wantagh, NY). A contact switch, connected between the cage and the water spout, operated the syringe pump whenever the animal was in contact with the spout. The syringe pump was set to dispense at a rate of 42 ml/h, and a rat received 8–14 ml of water per daily session. Mild electric shock was provided by a Coulbourn ac-resistive small animal shocker connected between the water spout and the cage floor. A 25 W light bulb located beneath the cage was turned on and off with the shock.

D. Acoustic apparatus

Behavioral thresholds were obtained for pure tones ranging from 2 to 45 kHz as well as for octave-band noise (approximately 20–40 kHz); these signals had a duration of 400 ms and a rise-fall time of 10 ms. In addition, thresholds were obtained for two of the same sounds used to obtain the ABR: a 16 kHz tone and the 20–40 kHz noise, both of which were 1 ms total duration with a rise-fall time of 0.5 ms (no plateau).

The 400 ms pure tones were digitally produced, gated with a 10 ms rise-fall time, amplified, and sent to either a Motorola piezoelectric speaker (2 kHz) or a Foster ribbon tweeter (4, 8, 16, and 45 kHz). The speaker was located 90E to the right of an animal's head at a distance of 1 m when it was drinking from the water spout. The noise was generated using Tucker-Davis Technologies (TDT) SIGGEN software. The output of the digital to analog converter (TDT, model DA3) was passed to a programable attenuator (TDT, model PA4), filtered, amplified, and sent to the ribbon tweeter. Sound pressure levels were measured using a Bruel & Kjaer (B&K) 1/4 in. (0.64 cm) microphone (Model 4135, B&K, Naerum, Denmark), a measuring amplifier (B&K model 2608), and a spectrum analyzer (Zonic 3525). The measuring equipment was calibrated with a pistonphone (B&K model 4230). The spectra of the 20–40 kHz noise and the 16 kHz tone stimuli are shown in Fig. 1. (The ABR stimuli are described below.)

E. Behavioral procedure

A standard conditioned suppression procedure was used to obtain the behavioral thresholds (Heffner *et al.*, 2006). A thirsty animal was placed in the test cage and allowed to drink from the water spout. Sounds were presented at random intervals and followed at their offset by a mild electric shock delivered through the spout. The animal quickly learned to avoid the shock by breaking contact with the spout whenever it heard a tone.

Test sessions were divided into 2.0 s intervals separated by 1.0 s intertrial intervals. Each trial contained either a sound ("warning" signal) or silence ("safe" signal), with 22% of the trials containing a sound. A response was recorded if an animal broke contact for more than half of the last 150 ms of a trial. The response was classified as a hit if the trial contained a sound and as a false alarm if no sound

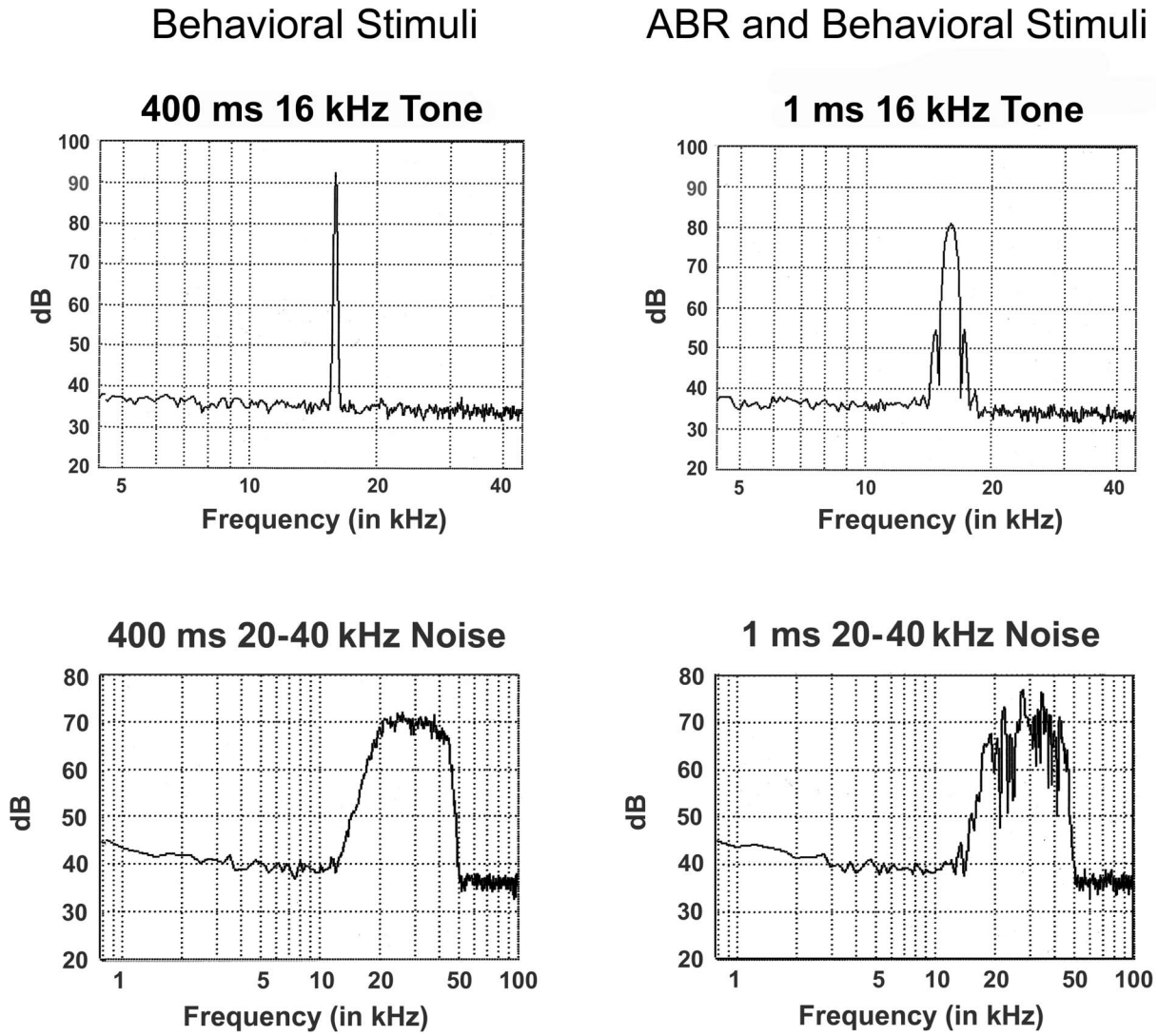


FIG. 1. Spectra of the noise and the 16 kHz tone stimuli used in the tests. Note that the 1 ms tone and noise stimuli are the same as those used for the ABR. The side lobes of the 1 ms, 16 kHz tone burst, which are caused by the rapid onset and offset, peak at 14.75 and 17.25 kHz.

was presented. Both the hit and false alarm rates were determined for each block of six to eight warning trials (which also included approximately 25 safe trials) for each stimulus condition. The hit rate was corrected for false alarms according to the formula: performance=hitrate-(false alarm rate $\times$hit rate), with the hit and false alarm rates expressed in proportions of 1. Absolute thresholds were determined by reducing the intensity of a tone in successive blocks of six to eight warning trials until the animal no longer responded to the signal above the 0.01 chance level (binomial distribution). Thresholds were obtained for only one stimulus at a time so that an entire session could be devoted to obtaining a reliable threshold.

F. Recording the auditory brainstem response

ABR testing was conducted in a double-walled sound chamber identical to that used for behavioral testing. To obtain the ABR, a rat was anesthetized with isoflurane, and subdermal electrodes were inserted at the vertex and behind the right ear, with the ground electrode in the animal's hind leg. The speaker was positioned directly above the animal's ear at a height of 12 cm. Body temperature was maintained by electrically heating the chamber. Because thresholds were obtained for only one sound, the procedure was usually completed in 15–25 min.

The sound noise were generated using the same equipment and loudspeaker used to obtain the behavioral thresholds. The main difference was that the stimuli were 1 ms in duration, 0.5 ms rise-fall time (no plateau), and pulsed 27.7 times/s. The spectra of the 16 kHz tone and the noise are shown in Fig. 1.

Data were collected using a Nicolet model CA 2000 electrodiagnostic system (Nicolet Instrument Corporation, Madison WI). The biological signal was bandpass filtered (0.15–3.0 kHz) and amplified (sensitivity setting of 25 μV) with the artifact rejection level set at 10 μV. The recording window was 10 ms in duration and was triggered by a timing pulse from the TDT system at the stimulus onset. Thresholds were determined by reducing the intensity of the stimulus in 10 dB steps until no latency-appropriate responses were evident. The intensity of the stimulus was then increased in 2.5 or 5 dB steps until a response could once again be discerned. Threshold was then defined as the lowest intensity at which a

latency-appropriate response with an amplitude greater than 0.05 μV could be detected. The number of samples per average varied with the clarity of the response, ranging from a minimum of 1000 at higher stimulus intensities to 6000–8000 around the threshold. At least two recordings were taken above and below the threshold and were compared to see if the peaks matched. The traces were then combined and the amplitude determined.

G. Exposure to loud sound

For exposure, an animal was anesthetized with isoflurane and its right ear exposed to a loud tone for 10 min. The exposure tones used were 1.4, 2.8, 5.6, 11.2, 16, and 31.5 kHz at intensities of 110, 115, or 120 dB sound pressure level (SPL). The 16 kHz tone was chosen because we had previously used it to induce tinnitus in rats (Imig *et al.*, 2007). The other frequencies were chosen because Davis *et al.* (1950) had found that the maximum hearing loss caused by exposure to loud tones generally, although by no means always, occurred half an octave above the frequency of the exposing tone. Thus, in measuring threshold shifts at frequencies half an octave above the frequency of the exposing tone, we expected to see differing degrees of hearing loss, which we did.

Some rats were exposed more than once as part of a study of the cumulative effects of exposure to loud sound. Specifically, rat 06-07 was exposed again 32 days after the first exposure, rat 07-08 was exposed 20 days after the first exposure, rat 06-01 was exposed 20 days after the first exposure and again 33 days after the second exposure, and rat 06-02 was exposed 16 days after the first exposure and again 34 days after the second exposure.

The tone was produced by a digital signal generator (Model 3525, Zonic, Tokyo, Japan), amplified (Model MPA, 100-w/channel, Radio Shack, Fort Worth, TX), and sent either to an Electro-Voice Model 1823M driver (for frequencies of 1.4 and 2.8 kHz) or to a Motorola KSN 1005A piezoelectric loudspeaker (for frequencies of 5.6 kHz and higher). The sound was directed to an animal's ear through a plastic funnel with a 4 mm inner diameter tip that was attached to the loudspeaker with thermoplastic adhesive. The sound was measured with the $\frac{1}{4}$ in. microphone placed at the tip of the plastic tube.

A behavioral threshold was obtained 1 h after the exposure, following which the animal was reanesthetized and its ABR threshold obtained. Behavioral and ABR thresholds were then obtained daily until they had stabilized, with the ABR threshold taken immediately following the behavioral threshold.

III. RESULTS

The results consist of a comparison between behavioral and ABR measures of threshold shift following a 10 min exposure to a loud tone.

To obtain maximum reliability, each animal was tested daily on the same stimulus until thresholds had stabilized. Because initial threshold shifts were obtained beginning 1 h after exposure, a control test was conducted to determine whether any lingering effects of the anesthesia might have affected the thresholds.

The results of the exposures are described in terms of (1) the size of the initial hearing loss determined behavioraly 1 h after the exposure (followed immediately by the ABR recording, (2) the time to recover from the temporary portion of the hearing loss (defined as the number of days it took for a threshold to fall to within 3 dB of its final value), and (3) the magnitude of the permanent hearing loss

A. Behavioral and ABR threshold shifts for tones

The behavioral threshold shifts for 400 ms pure tones are compared with the ABR threshold shifts (1 ms tone pips) in Figs. 2 and 3, where they are arranged by the frequency of the test tone. Because preliminary tests showed that the ABR threshold did not vary from day to day, only one or two pre-exposure ABR thresholds were obtained to minimize the number of times that a rat had to be anesthetized. Pre-exposure behavioral thresholds were also quite stable, generally varying by less than 3 dB.

As can be seen in Figs. 2 and 3, the ABR does not provide a reliable estimate of the initial behavioral threshold shift. Differences between the ABR and behavioral thresholds ranged from an underestimate of more than 28 dB [Fig. 2(a)] to an overestimate of 28.8 dB [Fig. 2(e)] with the ABR as likely to overestimate as to underestimate the behavioral threshold shift.

With regard to the time to recover from the temporary threshold shift (defined as the number of days it took for a threshold to fall to within 3 dB of its final value), in only 3 of the 11 cases did the ABR agree with the behavioral recovery time [Figs. 2(a), 3(c), and 3(e)].

An analysis of the permanent hearing loss, on the other hand, shows some agreement between the ABR and the behavioral measure, with the final ABR falling within ± 5 dB of the final behavioral measure of hearing loss in 6 of the 11 cases [Figs. 2(a), 2(b), 2(d), 3(a), 3(b), and 3(f)]. However, it should be noted that the two cases showing the best agreement were two exposures on the same animal that did not have a permanent hearing loss [Figs. 3(a) and 3(b)]. In other cases, the ABR underestimated the permanent hearing loss by 7.8 dB [Fig. 3(c)] and overestimated it by up to 37.3 dB [Fig. 2(e)]. In short, it would appear that the tone ABR does not provide a reliable estimate of pure-tone sensorineural hearing loss.

B. Anesthesia controls

The consistent disagreement between the behavioral and ABR measures of the initial hearing loss raised the possibility that the behavioral measure might have been affected by lingering effects of the anesthesia, even though the animals were given an hour to recover before testing. To investigate this possibility, we determined the behavioral thresholds of four rats for the 16 kHz, 1 ms tone pips used in the ABR test. The animals were then given a "sham" exposure in which they were anesthetized for 10 min, but not exposed to sound, and then tested 1 h later. As shown in Fig. 4, the anesthesia

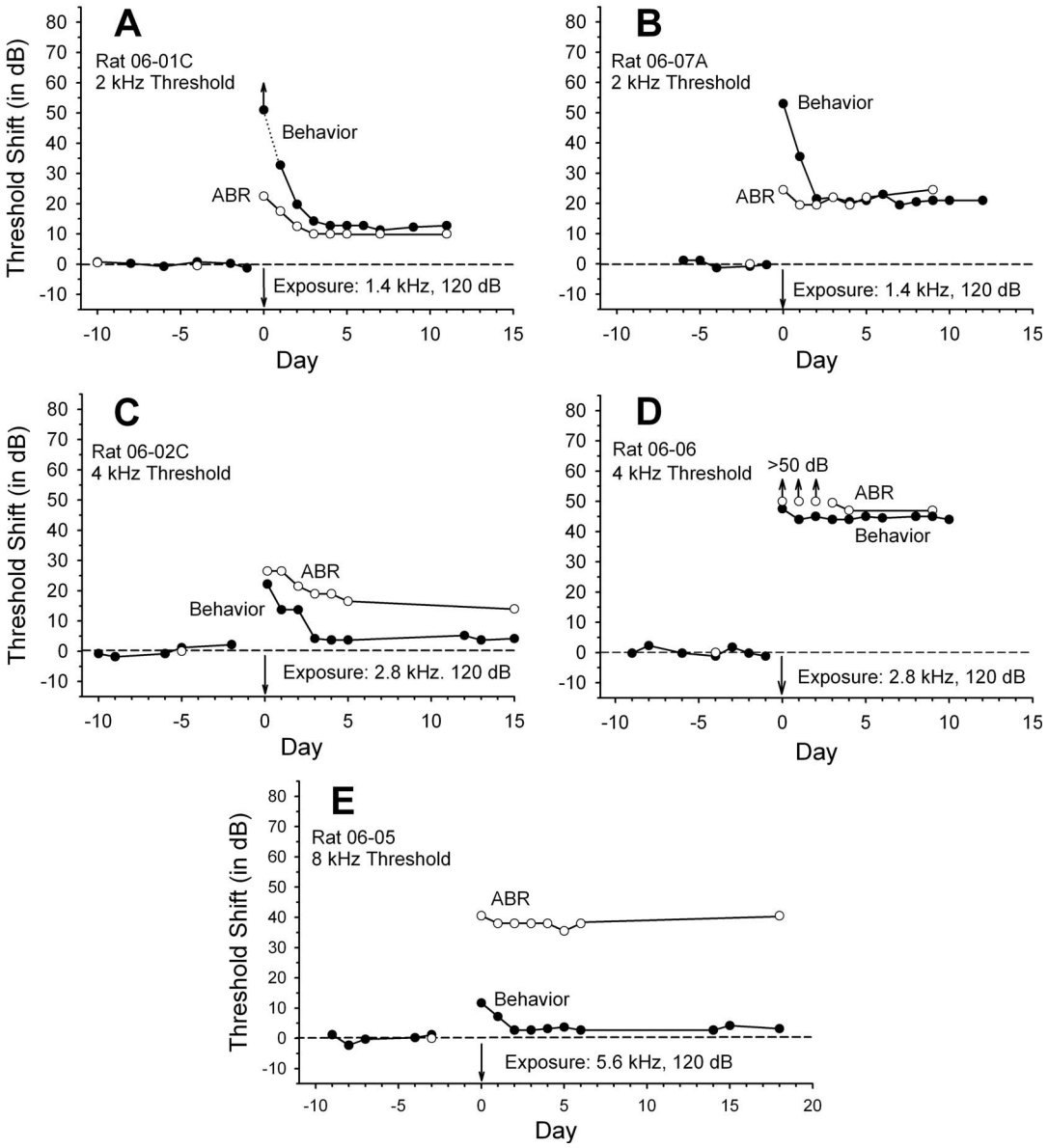


FIG. 2. Behavioral (closed circles) and ABR (open circles) threshold shifts for 2, 4, and 8 kHz tones. Upward pointing arrows and dotted lines indicate that the threshold that day was greater than the maximum stimulus intensity that could be produced. The intensity and frequency of the exposure tone is listed in each graph; all exposure durations were 10 min. Note that here and in Fig. 3 the letters A, B, and C following a rat's designation indicate that the results are from the animal's first, second, and third exposures; most rats were exposed only once.

alone had no effect on their behavioral thresholds, nor, for that matter, did it affect the subsequent ABR threshold. Thus, the large discrepancies observed between behavioral and ABR measures of the initial hearing loss was not caused by any lingering effects of the anesthesia.

C. Behavioral and ABR threshold shifts using the same 1 ms 16 kHz stimulus

Because the tone ABR uses a different stimulus than the pure-tone behavioral audiogram (Fig. 1), it is possible that there might be better agreement between the two measures if the same auditory stimulus were used for both. To test this possibility, behavioral thresholds were obtained for four rats using the same 1 ms, 16 kHz tone pips used for the ABR. The rats were tested before and after exposure to 11.2 kHz at 120 dB for 10 min.

Despite using the same stimuli, the ABR still did not provide a reliable estimate of the behavioral threshold shift for tones (Fig. 5). Differences between the initial behavioral and ABR threshold shifts ranged from 5.1 to 18.6 dB (Fig. 5). With regard to the time to recover from the temporary portion of the hearing loss, the ABR indicated that recover occurred by day 1 in each case, but the behavioral thresholds of rats 08-01 and 08-02 took 2 and 3 days, respectively, to recover to within 3 dB of their final value (Fig. 5).

Finally, the ABR and behavioral measures of the permanent threshold shift fell to within ±5 dB of each other for

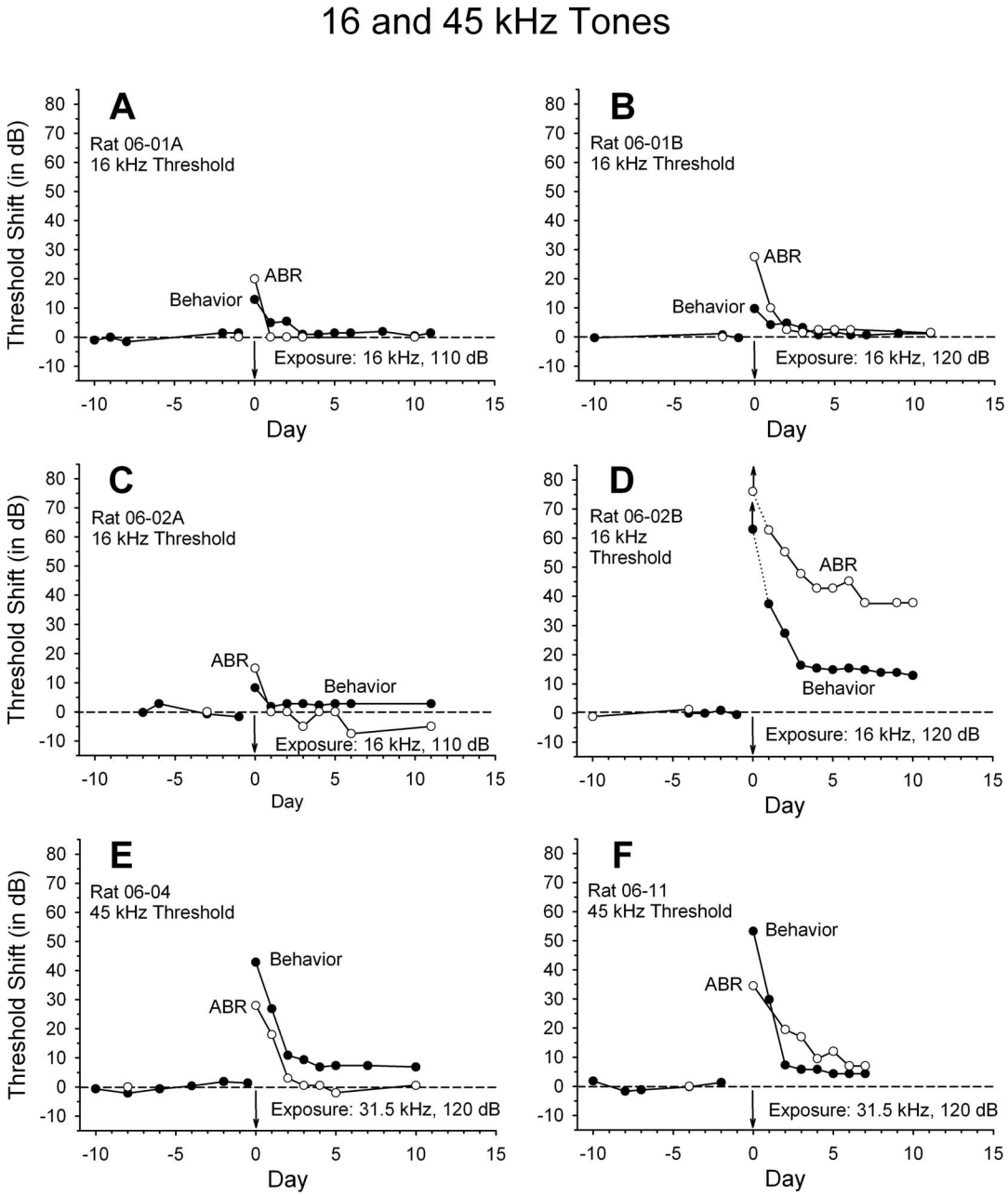


FIG. 3. Behavioral (closed circles) and ABR (open circles) threshold shifts for 16 and 45 kHz tones. Upward pointing arrows indicate that the threshold was greater than the maximum stimulus intensity that could be produced. Exposure frequency and intensity are listed in each graph; all exposure durations were 10 min.

two of the animals [Figs. 5(b) and 5(c)]. In the other two cases, the ABR overestimated the behavioral threshold shift by 15.1 dB in one case [Fig. 5(a)] and underestimated it by 15.6 dB in another [Fig. 5(d)]. Thus, the ABR did not estimate the behavioral threshold shift for the 1 ms tone pip used in the ABR any better than it estimated the 400 ms pure-tone threshold shift.

D. Behavioral and ABR threshold shifts for 20–40 kHz noise

In recent studies of tinnitus induced by exposure to loud sound, we estimated the accompanying hearing loss by determining the shift in the ABR threshold using band filtered noise (Heffner and Harrington, 2002; Heffner and Koay, 2005; Imig *et al.*, 2007). Therefore, we were interested in determining whether the ABR reliably estimated the behavioral hearing loss for the 20–40 kHz noise we have used elsewhere (Imig *et al.*, 2007). The behavioral thresholds in this experiment were obtained using the 400 ms duration noise.

In contrast to the tests using tonal stimuli, the results of this test indicated relatively good, although not perfect, agreement between the ABR and behavioral measures of threshold shift (Fig. 6). As with the tone ABR, the noise-evoked ABR was least accurate in estimating the initial hearing loss, with none of the four estimates within ±5 dB of the behavioral threshold shifts. However, it was fairly accurate in estimating the time course, being off by one day in one case in which the behavioral thresholds took 3 days to stabilize whereas the ABR threshold stabilized in 2 days [Fig.

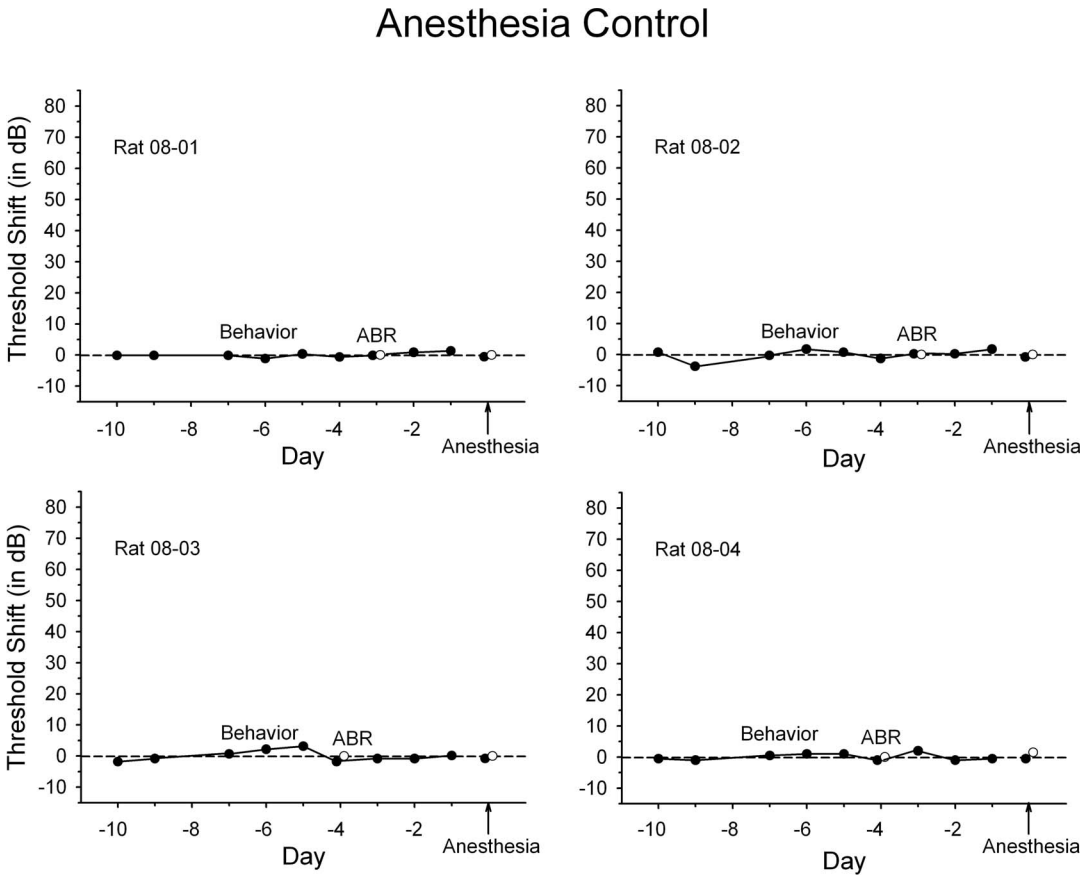


FIG. 4. Behavioral thresholds are not affected by 10 min of gas anesthesia (isoflurane) delivered 1 h before behavioral testing. The behavioral auditory stimulus was the same 1 ms, 16 kHz stimulus used in the 16 kHz ABR. Closed circles indicate behavioral thresholds; open circles indicate ABR thresholds.

6(c)]. Similarly, it was fairly accurate in estimating the permanent threshold shift, with the final scores for all four animals differing by no more than ±3.7 dB.

E. Behavioral and ABR threshold shifts using the same 1 ms noise stimulus

Given the relatively good agreement between the ABR and behavioral measures of threshold shift using the noise stimulus, we investigated whether the agreement would be greater if the same 1 ms noise burst were used for both the behavioral and the ABR thresholds.

With regard to the initial hearing loss, two of the four ABR estimates of threshold shift fell within ±5 dB of the behavioral threshold shifts (Fig. 7)—which is a slightly better agreement than was found for the 400 ms noise thresholds. With regard to the time course of the recovery, the ABR disagreed with the behavioral recovery in two of the five cases where in one it indicated a five versus a four day recovery [Fig. 7(a)], and in the other it indicated a one versus a three day recovery [Fig. 7(e)]. Finally, with regard to the permanent threshold shift, the ABR estimate showed slightly less agreement than was found for the 400 ms noise thresholds with four of the five threshold estimates falling within ±5 dB.

To determine if using the same stimulus for the behavioral and ABR thresholds resulted in a better agreement, the results of this test were compared with those of the previous one using the Mann–Whitney *U* test in which the difference between the behavioral and ABR threshold shifts were rank ordered. The results of the analysis indicated that using the same stimuli for both tests did not significantly improve the agreement ($p > 0.2$). However, given the small sample sizes, we cannot rule out the possibility that increasing the number of animals tested might yield a statistically reliable difference.

F. Hearing loss

It can be seen from these results that the magnitude of a threshold shift often varied between animals exposed to the same loud tone, a phenomenon that has been seen in both humans and animals (e.g., Davis *et al.*, 1950; Heffner and Harrington, 2002). One possibility is that the magnitude of the threshold shift is related in some way to pre-exposure absolute sensitivity. For example, perhaps those animals with better sensitivity are more susceptible to the traumatic effects of loud sound, or the reverse. Because of the variety of exposing tones and test stimuli that were used, we do not have many instances for comparison. Thus, Table I shows the two sets of data for which three or more animals were exposed and tested in the same way. In the first group, in which four animals were exposed to 11.2 kHz at 120 dB and tested on the 1 ms 16 kHz tone pip (Fig. 5), the rank ordering of the rats by pre-exposure sensitivity indicates that the better the pre-exposure sensitivity, the greater the hearing loss. However, the second group, in which three animals were exposed to 16 kHz at 120 dB and tested on 1 ms noise (Fig. 7), shows

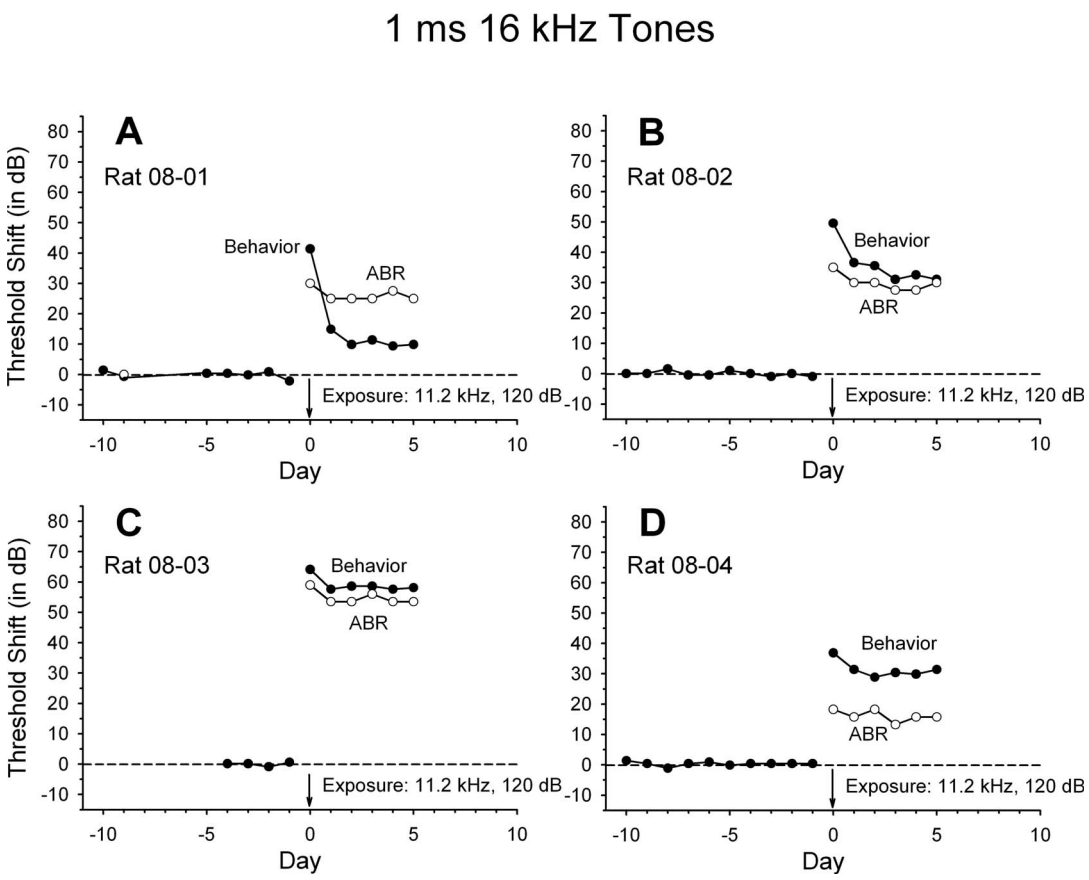


FIG. 5. Behavioral (closed circles) and ABR (open circles) threshold shifts for the 1 ms, 16 kHz stimulus. Using the same tonal stimuli for the behavioral and ABR thresholds does not noticeably improve the agreement between the two procedures. (The last of two pre-exposure ABRs for rats 08-02 and 08-03 were done 14 days before exposure; the last one for rat 08-04 was done 26 days before exposure.)

the opposite, that the less sensitive animals had the greater hearing loss. However, it should be noted that the animals' absolute thresholds do not vary that much and that whereas the pre-exposure thresholds of rats 08-01 and 08-02 differed by only 0.1 dB, their threshold shifts differed by over 20 dB (Table I). Although it is possible that there is some relationship between absolute sensitivity and threshold shift that varies with the frequency of the exposing and test stimuli, we do not have sufficient data to address this question. Similarly, we do not yet have sufficient data to determine if previous exposure to loud sound affects the results of subsequent exposures.

IV. DISCUSSION

A. The behavioral thresholds

The interpretation of these results rests on the degree of confidence in the behavioral thresholds. The method of conditioned suppression used here is a relatively simple procedure that has long been used to determine the auditory thresholds of mammals (Masterton *et al.*, 1969). Indeed, thresholds obtained for rats with this method in different laboratories and many years apart show remarkably good agreement (cf. Heffner *et al.*, 1994 and Kelly and Masterton, 1977). One factor contributing to this reliability is that the act of drinking from a spout fixes an animal's head in the sound field, thus making accurate measurement of the sound reaching its ears possible. Another is that an animal need only make the simple and natural response of freezing when it detects a sound. Finally, by devoting an entire test session to a single sound, we ensured that a sufficient number of trials could be obtained to precisely determine an animal's threshold. The stability over time also supports our confidence that the behavioral thresholds are both reliable and valid.

It should be noted, however, that exposing animals to a sound loud enough to cause a hearing loss may also cause tinnitus; moreover, given the levels of exposure used here, the severity of the tinnitus would be expected to be greatest immediately after the exposure and then to gradually decline (Heffner and Koay, 2005; Imig *et al.*, 2007). Thus, the question arises as to whether the greater initial difference between the behavioral and ABR measures of hearing loss for noise could be attributed to tinnitus. Although plausible, there are at least two reasons why this is probably not the case. First, the behavioral stimulus was pulsed to prevent it from being confused with tinnitus. Although some patients do describe their tinnitus as pulsing, it would seem unlikely that the animals would develop tinnitus that was close in pitch and pulsing at the same rate as the sounds on which they were tested. Thus, the characteristics of the physical sounds should have prevented them from being confused with tinnitus. Second, unlike human patients who typically have little experience in auditory testing, the rats in this study were trained observers, having received 30 or more

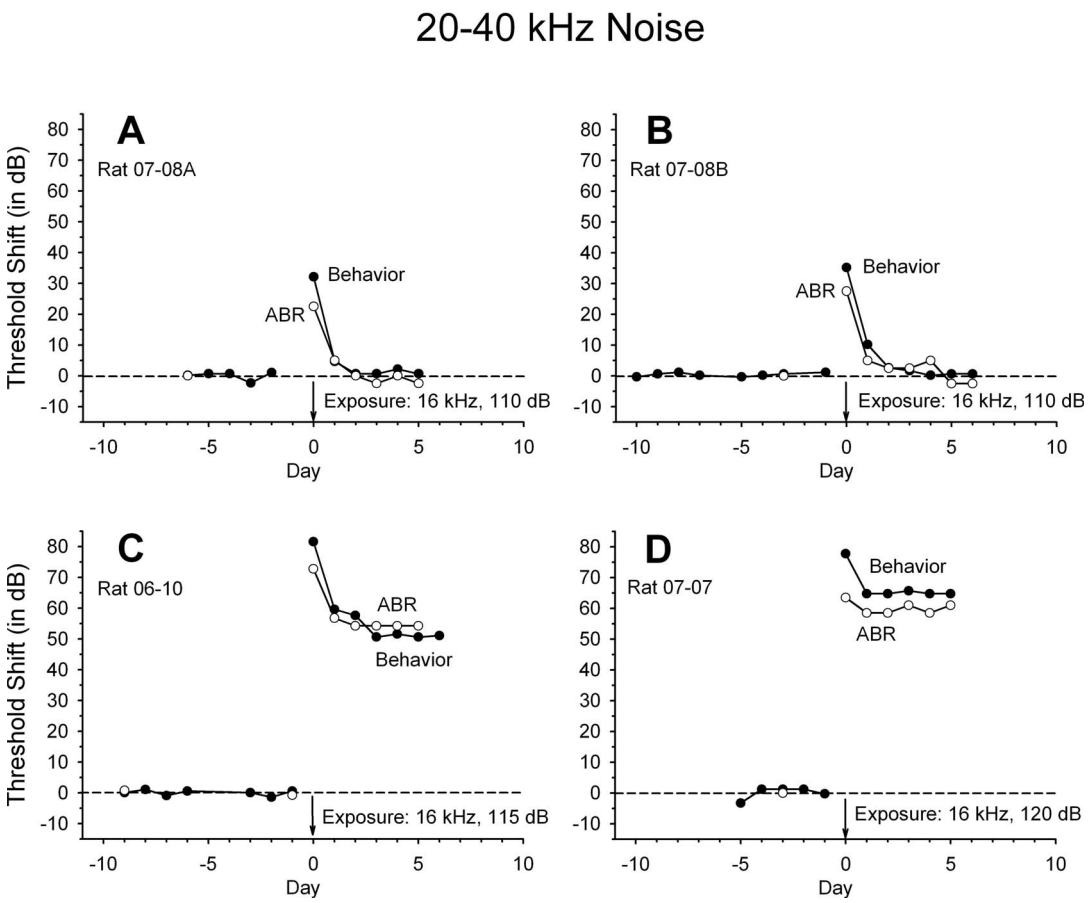


FIG. 6. Behavioral (closed circles) and ABR (open circles) threshold shifts for 20–40 kHz band noise in which behavioral thresholds were determined with 400 ms noise bursts and ABR thresholds with 1 ms noise bursts. Note the relatively good agreement between the two measures.

days of training to detect one specific sound prior to testing. Thus, we think it unlikely that the animals' thresholds were affected by tinnitus.

B. Estimating sensorineural hearing loss from the ABR

One of the main conclusions of this study is that tonal ABR thresholds do not provide a reliable estimate of sensorineural hearing loss for tones, regardless of whether the behavioral tests use pure tones or the ABR tone pips. Interestingly, the problem is not that the ABR errs by a consistent amount and direction (in which case a correction factor could be applied), but that its correspondence with the behavioral threshold shift is erratic. For example, the tone ABR estimated the permanent threshold shift to within ±5 dB in 8 of the 15 cases (see Figs. 2, 3, and 5), but over- or underestimated the other 7 cases by as much as 37 dB [Fig. 2(e)]. Because the tone ABR provides an accurate measure in about half of the cases, we re-examined the records of each of the animals to see if there was some factor, such as background noise level in the ABR, that might indicate whether or not the ABR was accurate, but could find none. Nevertheless, it is conceivable that there might be some other measure that, used in conjunction with the ABR, would indicate those animals for which the ABR provides an accurate estimate of threshold shift (e.g., evoked otoacoustic emissions, middle latency, and other responses).

In contrast, the noise-evoked ABR gave a more accurate picture of the behavioral threshold shift for that noise stimulus, regardless of whether the behavioral stimulus was 400 ms or 1 ms pulses (Figs. 6 and 7). As shown in Table II, rank ordering the animals on the ABR estimate of threshold shift for noise results in only one reversal for the initial hearing loss for noise and only a minor reversal for the permanent hearing loss. However, it should be noted that octave-band noise is a relatively broad frequency stimulus, and the question is whether the ABR evoked by narrow band noise would provide a reliable indication of the frequencies at which a hearing loss occurs.

Finally, we evaluated the noise-evoked ABR as an estimate of behavioral hearing loss because we have previously used it to estimate hearing loss in studies of tinnitus (Heffner and Harrington, 2002; Heffner and Koay, 2005; Zhang *et al.*, 2004). These studies suggested that the increase in spontaneous activity in the dorsal cochlear nucleus that follows exposure to loud sound is due not to tinnitus, but to the hearing loss resulting from the exposure (for a discussion of this issue, see Heffner and Koay, 2005). One outcome of the present study is that the noise-evoked ABR is a reliable measure of behavioral hearing loss. This finding supports the view that increase in spontaneous activity in the DCN, which begins to occur about a week after exposure to loud sound, is related not to tinnitus, but to the accompanying hearing loss.

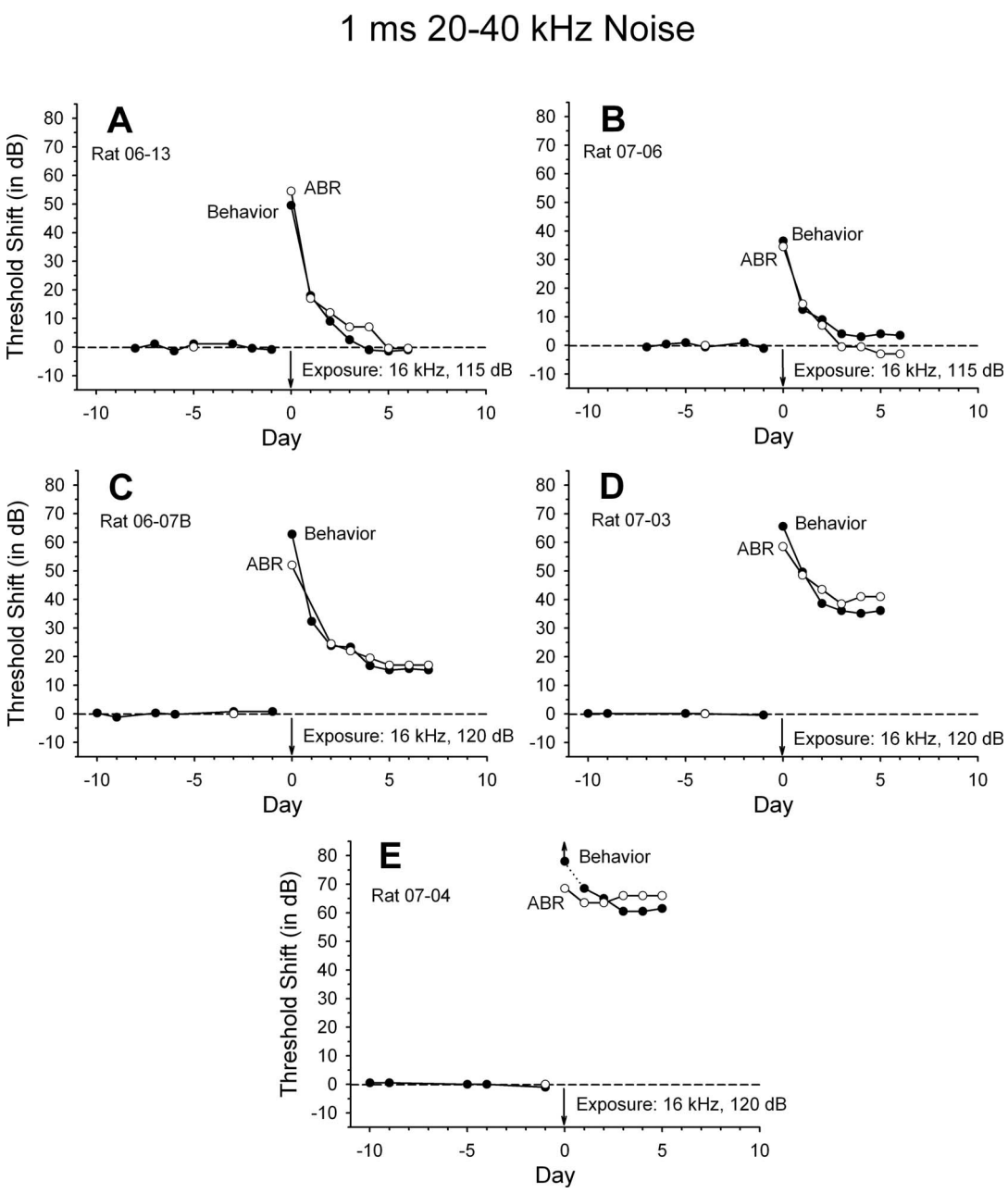


FIG. 7. Behavioral (closed circles) and ABR (open circles) threshold shifts for 1 ms high-frequency noise bursts in which both the behavioral and the ABR thresholds were determined for 1 ms noise bursts. Using the same noise stimulus in the behavioral and ABR tests did not improve the correspondence between the two measures beyond that seen in Fig. 6.

C. Previous comparisons of behavioral and physiological measures of hearing loss in animals

As noted in the Introduction, there have been four previous studies that compared behavioral and physiological measures of hearing loss in the same animals. These differed from the present study in several important ways. First, although the previous studies tested fewer animals, they obtained thresholds from each animal for a number of different frequencies. Second, with one exception, they measured the permanent but not the initial hearing loss. Third, again with one exception, the previous studies did not calculate threshold shifts but rather compared the posttreatment behavioral and physiological absolute thresholds with each other. Thus, where possible, we reanalyzed their data to determine threshold shifts. Finally, only one of the studies recorded the ABR; the others recorded either the compound action potential from inside the bulla or the evoked response recorded with bipolar electrodes implanted in the inferior colliculus.

The first study, conducted by Dallos *et al.* (1978), compared tone-evoked compound action potentials with behavioral thresholds in gerbils and chinchillas whose cochleas had been damaged with kanamycin. The physiological stimuli were short-duration tones with a 1 ms rise time, whereas the behavioral thresholds were obtained using 3.8 s tones with a 10 ms rise time. Because their physiological measures were obtained in terminal experiments, the comparisons were limited to posttreatment measures of the permanent effects of the kanamycin. Their results for four ger-

TABLE I. Pre-exposure sensitivity and the magnitude of postexposure threshold shift.

Rat	Pre-exposure threshold (in dB SPL)	Threshold shift (in dB)
Exposed to 11.2 kHz at 120 dB[a] Tested on 1 ms 16 kHz:		
08-03	19.9	58.1
08-04	22.6	31.4
08-02	23.5	31.1
08-01	23.6	9.9
Exposed to 16 kHz at 120 dB[a] Tested on 1 ms noise:		
06-07B	13.7	15
07-03	17.9	36.1
07-04	18.5	61.5

[a]All exposures were 10 min in duration.

bils and four chinchillas, tested on multiple frequencies, showed that the thresholds for the compound action potential paralleled the behavioral thresholds fairly well and could thus indicate the frequencies at which kanamycin caused a behavioral threshold shift. However, the compound action potential was less successful in indicating the magnitude of the behavioral threshold shift. In the case of the gerbils, the compound action potential sometimes indicated the actual behavioral threshold although in most cases it overestimated the hearing loss, sometimes by as much as 40 dB. In the case of the chinchillas, the compound action potential threshold was almost always higher than the behavioral threshold with an average difference of 18 dB. Thus, the compound action potential can indicate the relative pattern of a behavioral hearing loss but is less successful in indicating the magnitude of the loss.

Five years later, Henderson *et al.* (1983) compared auditory evoked potentials recorded from a bipolar electrode implanted in the inferior colliculus with behavioral thresholds in three monaural chinchillas. The animals were tested before and after a 1 h exposure to loud noise (a mixture of continuous and impulse 2–4 kHz band noise). The physiological and behavioral stimuli in this study were both 20 ms duration tones (5 ms rise-fall times). They obtained a threshold for seven different frequencies (ranging from 0.5 to 8 kHz) 24 h after exposure and again 30 days after exposure. Although they did not calculate threshold shifts, it is possible to derive threshold shifts from Table 2 of their paper. Their data, like ours, show that the evoked response does not reliably correspond to the initial behavioral threshold shift. Specifically only 12 of their 21 physiological estimates (57%) fell within ±5 dB of the behavioral threshold shifts with some being off by as much as 25 or 30 dB. With regard to the permanent behavioral hearing loss, the evoked response estimated 13 out of 21 threshold shifts to within ±5 dB. However, the chinchillas had little or no permanent hearing loss, with only a third of the thresholds elevated by 5 dB or more, and thus these results do not indicate how well this method estimates a permanent loss; for this it is necessary to turn to the next study.

The third study, by Davis and Ferraro (1984), also compared auditory evoked potentials recorded from a bipolar electrode implanted in the inferior colliculus with behavioral thresholds in monaural chinchillas before and after exposure to loud sound, in this case a 2 kHz tone (120 dB, 4 h). Although they did not measure the initial hearing loss, they determined the permanent behavioral hearing loss (5 weeks after exposure) for two sets of tones, the same tones used in the physiological measures (20 ms duration, 5 ms rise-fall) as well as longer duration tones (500 ms duration, 5 ms rise-fall). They obtained pre- and postexposure evoked potentials and behavioral thresholds for six chinchillas at seven frequencies from 500 Hz to 4 kHz. Their results, reanalyzed to reveal the amount of threshold shift for each measure, can be summarized as follows: First, the evoked response estimated the behavioral threshold shifts for 500 ms tones to within ±5 dB, 10 out of 35 times (29%) with misestimates as high as 25 dB or more. However, it was more accurate in estimating the threshold shifts for the 20 ms tones in which 26 of the 42 threshold estimates (62%) were within ±5 dB and the largest misestimates appeared to fall within 15–19 dB of the behavioral thresholds. Second, the evoked response accurately estimated the 20 ms behavioral threshold shifts for some animals, but not for others; for example, the tone be-

TABLE II. Rank ordering by ABR estimate of hearing loss for 20–40 kHz noise.

	Initial hearing loss (in dB)			Permanent hearing loss (in dB)			
Rat	ABR	Behavior	Difference	Rat	ABR	Behavior	Difference
400 ms noise:							
07-08A	22.5	32.2	−9.7	07-08A	−2.5	0.7	−3.2
07-08B	27.5	35.2	−7.7	07-08B	−2.5	0.7	−3.2
07-07	63.5	77.7	−14.2	06-10	54.3	50.6	3.7
06-10	72.8	81.6	−8.8	07-07	61.0	64.7	−3.7
1 ms noise:							
07-06	34.5	36.5	−2.0	07-06[a]	−3.0	3.5	−6.5
06-07B	52.0	62.8	−10.8	06-13	−0.5	−1.0	−1.7
06-13[a]	54.5	49.5	5.0	06-07B	17.0	15.3	0.5
07-03	58.5	65.6	−7.1	07-03	41.0	36.1	4.9
07-04	68.5	>78.0	−9.5+	07-04	66.0	61.5	4.9

[a]Indicates an incorrect ranking (in both cases, the ABR underestimated the behavioral hearing loss).

havioral threshold shifts of one chinchilla (animal 5) fell within ±5 dB of the evoked response shifts for all seven frequencies whereas only two of the seven thresholds of another (number 3) fell within ±5 dB. Finally, the evoked response was more accurate in estimating the behavioral threshold shift for lower frequencies (0.5, 0.75, and 1 kHz) than for higher frequencies (1.5, 2, 3, and 4 kHz).

The final study, by Borg and Engström (1983), recorded the ABR in rabbits using subcutaneous needle electrodes. The physiological stimuli were brief tone bursts, whereas the behavioral stimuli were 10 s tones. The authors reported threshold shifts for two rabbits, one that had been exposed to loud sound and the other that had been treated with kanamycin. The results of the animal exposed to loud sound indicated a fairly good agreement with the two measures differing by 7 dB or less at 0.5, 1, 2, and 4 kHz; behavioral thresholds at higher frequencies could not be obtained, so no further comparisons were available. The results of the animal given kanamycin also showed good agreement; with the exception of 0.5 kHz where the threshold shifts differed by 25 dB, the threshold shifts at 1, 2, 4, 8, and 16 kHz were within 4 dB of each other. The authors mentioned that they made two other noise exposure but did not report the results. Thus, their results are based on complete data from one animal and on partial data from another.

In conclusion, the most comprehensive study for which measures of threshold shift are available is that by Davis and Ferraro (1984). Their results indicate that the evoked response recorded from the inferior colliculus can estimate the permanent behavioral threshold shifts for short-duration (20 ms) pure tones to within ±5 dB about 60% of the time, with misestimates ranging near 20 dB. In comparison, our results indicate that the ABR evoked by octave noise estimated the permanent behavioral threshold shift to an accuracy of ±5 dB eight out of nine times (89%), with the largest misestimate being 6.5 dB (Table II). This suggests that the accuracy of physiological measures of hearing loss is improved by using noise rather than tone stimuli. However, to obtain frequency-specific information, it would be necessary to narrow the width of the noise band, and we do not know at this time if this would reduce its accuracy.

D. The hearing loss

One of the outcomes of this study is the observation that animals exposed to the same loud sound may develop different hearing losses. As can be seen in Table I, this variation cannot be easily accounted for by variation in pre-exposure sensitivity. This is not a new observation as we have seen such variation in hamsters exposed to loud tones (Heffner and Harrington, 2002). Indeed, in their classic study of temporary hearing loss caused by exposure to loud sound, Davis *et al.* (1950) noted that the same exposures to loud sound result in different patterns and degrees of hearing loss in different individuals; as they stated, "…some individuals are systematically more susceptible than others." This leads to the question of why some ears are more resistant to overstimulation by sound. Is there individual variation in some biochemical or physiological mechanism that protects the ear from loud sounds? Is there a way to identify those individuals with susceptible ears so that they might take precautions to protect them? Are there treatments that might make ears less susceptible to damage?

ACKNOWLEDGMENTS

We thank Dr. Harvard Armus for providing the animals. Supported by NIH Grant No. DC6629.

Borg, E., and Engström, B. (**1983**). "Hearing thresholds in the Rabbit, a behavioral and electrophysiological study," Acta Oto-Laryngol. **95**, 19–26.

Dallos, P., Harris, D., Özdamar, Ö., and Ryan, A. (**1978**). "Behavioral, compound action potential, and single unit thresholds: Relationship in normal and abnormal ears," J. Acoust. Soc. Am. **64**, 151–157.

Davis, R. I., and Ferraro, J. A. (**1984**). "Comparison between AER and behavioral thresholds in normally and abnormally hearing chinchillas," Ear Hear. **5**, 153–159.

Davis, H., Morgan, C. T., Hawkins, J. E., Jr., Galambos, R., and Smith, F. W. (**1950**). "Temporary deafness following exposure to loud tones and noise," Acta Oto-Laryngol., Suppl. **88**, 1–57.

Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncates*)," J. Acoust. Soc. Am. **119**, 3181–3192.

Gorga, M. P. (**1999**). "Predicting auditory sensitivity from auditory brainstem response measurements," Semin. Hear. **20**, 29–43.

Gorga, M. P., and Neely, S. T. (**2002**). "Some factors that may influence the accuracy of auditory brainstem response estimates of hearing loss," in *A Sound Foundation Through Early Amplification 2001, Proceedings of the Second International Conference*, edited by R. C. Seewald and J. S. Gravel (Phonak AG, Chicago), pp. 49–61.

Heffner, H. E., and Harrington, I. A. (**2002**). "Tinnitus in hamsters following exposure to intense sound," Hear. Res. **170**, 83–95.

Heffner, H. E., and Heffner, R. S. (**2003**). "Audition," in *Handbook of Research Methods in Experimental Psychology*, edited by S. F. Davis (Blackwell, Malden, MA), pp. 413–440.

Heffner, H. E., Heffner, R. S., Contos, C., and Ott, T. (**1994**). "Audiogram of the hooded Norway rat," Hear. Res. **73**, 244–247.

Heffner, H. E., and Koay, G. (**2005**). "Tinnitus and hearing loss in hamsters exposed to loud sound," Behav. Neurosci. **119**, 734–742.

Heffner, H. E., Koay, G., and Heffner, R. S. (**2006**). "Behavioral assessment of hearing in mice Conditioned suppression," in *Current Protocols in Neuroscience*, edited by J. Crowley (Wiley, NY), pp. 8.21D.1–8.21D.15.

Henderson, D., Hamernik, R. P., Salvi, R. J., and Ahroon, W. (**1983**). "Comparison of auditory-evoked potentials and behavioral thresholds in the normal and noise-exposed chinchilla," Audiology **22**, 172–180.

Imig, T., Heffner, H., Koay, G., and Durham, D. (**2007**). "Time course of recovery of spontaneous activity (SA) in the rat inferior colliculus (IC) following unilateral acoustic trauma," Assoc. Res. Otolaryngol. Abstr. **30**, 136–137.

Kelly, J. B., and Masterton, B. (**1977**). "Auditory sensitivity of the albino rat," J. Comp. Physiol. Psychol. **91**, 930–936.

Masterton, B., Heffner, H., and Ravizza, R. (**1969**). "The evolution of human hearing," J. Acoust. Soc. Am. **45**, 966–985.

Stapells, D. R. (**2000a**). "Frequency-specific evoked potential audiometry in infants," in *A Sound Foundation Through Early Amplification*, edited by R. C. Seewald (Phonak AG, Basel), pp. 13–31.

Stapells, D. R. (**2000b**). "Threshold estimation by the tone-evoked auditory brainstem response: A literature meta-analysis," J. Speech-Lang., Path. & Audiology **24**, 74–83.

Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**). "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms," J. Acoust. Soc. Am. **106**, 1134–1141.

Zhang, J., Heffner, H. E., Koay, G., and Kaltenbach, J. A. (**2004**). "Hyperactivity in the hamster dorsal cochlear nucleus: Its relationship to tinnitus," Assoc. Res. Otolaryngol. Abstr. **27**, 302.

Spectral integration of speech bands in normal-hearing and hearing-impaired listeners

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 21 December 2007; revised 12 May 2008; accepted 13 May 2008)

This investigation examined whether listeners with mild–moderate sensorineural hearing impairment have a deficit in the ability to integrate synchronous spectral information in the perception of speech. In stage 1, the bandwidth of filtered speech centered either on 500 or 2500 Hz was varied adaptively to determine the width required for approximately 15%–25% correct recognition. In stage 2, these criterion bandwidths were presented simultaneously and percent correct performance was determined in fixed block trials. Experiment 1 tested normal-hearing listeners in quiet and in masking noise. The main findings were (1) there was no correlation between the criterion bandwidths at 500 and 2500 Hz; (2) listeners achieved a high percent correct in stage 2 (approximately 80%); and (3) performance in quiet and noise was similar. Experiment 2 tested listeners with mild–moderate sensorineural hearing impairment. The main findings were (1) the impaired listeners showed high variability in stage 1, with some listeners requiring narrower and others requiring wider bandwidths than normal, and (2) hearing-impaired listeners achieved percent correct performance in stage 2 that was comparable to normal. The results indicate that listeners with mild–moderate sensorineural hearing loss do not have an essential deficit in the ability to integrate across-frequency speech information.
[DOI: 10.1121/1.2940582]

PACS number(s): 43.66.Dc, 43.66.Sr, 43.71.Ky [BCM] Pages: 1105–1115

I. INTRODUCTION

The present study investigated speech recognition associated with a single relatively narrow band of speech information or two relatively narrow bands that were widely separated in frequency. In normal-hearing listeners, narrow bands of speech from widely spaced spectral regions can combine to produce a percent correct well above the sum of the percent correct values associated with each band separately (e.g., Grant and Braida, 1991; Warren *et al.*, 1995; Lippmann, 1996; Kasturi *et al.*, 2002). For example, whereas a low- and a high-frequency region may each support less than 20% correct speech identification, both bands presented together may support 70% identification (e.g., Grant and Braida, 1991). Although such effects are not predicted adequately by the articulation theory (e.g., French and Steinberg, 1947), they should not be regarded as surprising. Grant and Braida commented that the substantial improvement in performance achieved when spectrally separated bands are presented together is due to the fact that the bands are likely to contain nonredundant, complementary information. One way to conceptualize this is in terms of the possible set of meaningful utterances that are compatible with the bandlimited information presented to the listener. Each band presented alone might be consistent with a large set of distinct utterances. However, when the bands are presented together the combined information is consistent with a reduced set of candidate utterances.

The ability to perceive speech on the basis of sparse cues that are separated in frequency could have importance for speech understanding in noisy backgrounds. For example, when the signal to noise ratio is very low, a listener may not have access to the entire spectrum of a speech target and good performance may depend upon the ability to integrate speech fragments that are separated in frequency (e.g., Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Assmann and Summerfield, 2004; Buss *et al.*, 2004; Cooke, 2006; Hall *et al.*, 2008). This may be particularly significant from the perspective of hearing impairment because some evidence indicates that listeners with sensorineural hearing loss may have a diminished ability to integrate synchronous, frequency-distributed information in the perception of speech (Turner *et al.*, 1995; Turner *et al.*, 1999; Healy and Bacon, 2002). Such evidence has arisen from studies using "vocoder" paradigms where speech is divided into a number of frequency bands and the envelope of each band is used to modulate a carrier stimulus in the same frequency region as the original speech, with either a tone or noise band serving as the carrier (e.g., Shannon *et al.*, 1995; Turner *et al.*, 1995; Dorman *et al.*, 1997; Turner *et al.*, 1999). In one such study, it was found that although hearing-impaired listeners could perform as well as normal for consonant perception based upon a single band containing temporal envelope information, the performance of the impaired listeners was worse than normal for two or more bands carrying quasi-independent temporal envelope information (Turner *et al.*, 1999). The normal performance for the single band case was interpreted as being consistent with temporal modulation transfer function studies (Viemeister, 1979) that have gener-

[a]Electronic mail: jwh@med.unc.edu

ally indicated a normal ability of hearing-impaired listeners to code the temporal envelope of a stimulus, provided that effects of stimulus audibility are taken into account (e.g., Bacon and Gleitman, 1992; Moore *et al.*, 1992). Turner *et al.* (1999) and Healy and Bacon (2002) have suggested that the speech results related to the combination of information from more than one spectral band containing temporal information may indicate a reduced ability of hearing-impaired listeners to combine spectrotemporal information across frequency.

One question of importance is whether the deficits in across-frequency integration suggested by the above-mentioned studies occur exclusively for the type of speech material they used (only temporal envelope information), or whether it also occurs for more conventionally filtered speech. A recent study by Grant *et al.* (2007) is pertinent to this question. That study examined the intelligibility of speech filtered into relatively narrow spectral bands for both normal-hearing listeners and listeners with sensorineural hearing impairment. They examined several conditions involving either audio only or audio-visual cues. The conditions most relevant to the present study involved either a relatively low-frequency band alone (298–375 Hz) or the low-frequency band plus a high-frequency band (4762–6000 Hz). Both normal-hearing and hearing-impaired listeners achieved approximately 20% correct performance for the low band alone, but whereas the normal-hearing listeners improved to approximately 60% for both low and high bands presented together, the hearing-impaired listeners improved to only about 40% correct with both bands present. Grant *et al.* noted several possible accounts for the poorer performance of the hearing-impaired listeners when both bands were present, including an essential deficit in the ability to integrate across-frequency information, relatively great hearing loss in the region of the high-frequency band, and poor processing of the high-frequency information due to upward spread of masking, a manifestation of poor frequency selectivity that is relatively common in listeners with sensorineural hearing loss (e.g., Tyler *et al.*, 1984; Gagné, 1988). One feature of the Grant *et al.* study that makes it somewhat difficult to compare to the previous vocoder-based studies is that although the Grant *et al.* study examined the impact of adding a high-frequency speech band to a low-frequency band, the study did not specifically determine the intelligibility associated with the high band presented alone.

In the present study, speech recognition was assessed for low and high bands presented alone, and for the bands presented together. Furthermore, stimulus features intended to minimize effects related to upward spread of masking were employed. The rationale was to use an approach that allowed a test of whether listeners with sensorineural hearing impairment have an essential deficit in the ability to integrate across-frequency speech information apart from factors that could be related to a peripherally based reduction in frequency selectivity. Two experiments were performed. The first experiment tested listeners with normal hearing, examining the intelligibility of speech presented in quiet and speech presented in a masking noise background that was intended to simulate a 40–50 dB hearing loss. The main purpose of this study was to obtain information about the robustness of across-frequency speech integration under unmasked and masked conditions and to provide baselines against which to compare data from hearing-impaired listeners tested in the second experiment.

II. EXPERIMENT 1: NORMAL-HEARING LISTENERS

A. Methods

1. Listeners

Two sets of listeners with normal hearing participated, one tested in quiet (four males and seven females) and the other tested in background noise (three males and five females). The mean age of the normal-hearing listeners was 35.2 years with a standard deviation of 12.3 years. All listeners were screened to have audiometric thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz.

2. Rationale and stimuli

The speech material consisted of Bamford–Kowal–Bench sentences (Bench *et al.*, 1979), with each sentence containing from three to five key words. There were 21 lists of 16 sentences each. This corpus of sentences allowed testing to be completed without replicating any sentence for all of the listeners tested in quiet and all but two of the listeners tested in noise. For these two listeners, parts of list 1 and list 2 were repeated. In some conditions, speech was filtered into a single band centered at one of two frequencies, and in other conditions bands were available at both center frequencies simultaneously. The two bands were arithmetically centered on 500 and 2500 Hz, and filtering was performed via convolution with a 12 Hz resolution.

An important part of the rationale underlying the methods was related to upward spread of masking. This rationale was not of direct interest with regard to the normal-hearing listeners but is important from the standpoint of hearing impairment. Many hearing-impaired listeners are prone to greater than normal upward spread of masking (e.g., Tyler *et al.*, 1984; Gagné, 1988), and we wished to minimize the possibility that such masking could underlie differences between the results of the normal-hearing and hearing-impaired listeners. This objective was met with two types of stimulus manipulation. One manipulation was a level boost of the high-frequency speech band relative to the low-frequency band. Dubno *et al.* (2006) used a similar method in a previous study of speech perception in hearing-impaired listeners. The speech level was 85 dB SPL prior to bandpass filtering for the low band and 97 dB SPL prior to bandpass filtering for the high band. The other stimulus manipulation involved a condition where the low and high bands were presented to opposite ears (details in the following), preventing peripheral masking of the high band by the low band.

In the conditions meant to simulate hearing loss, a pink noise was presented at a level of approximately 37 dB/Hz SPL at 1 kHz. This noise resulted in masked thresholds of approximately 50–55 dB SPL at octave frequencies from 500 to 4000 Hz.

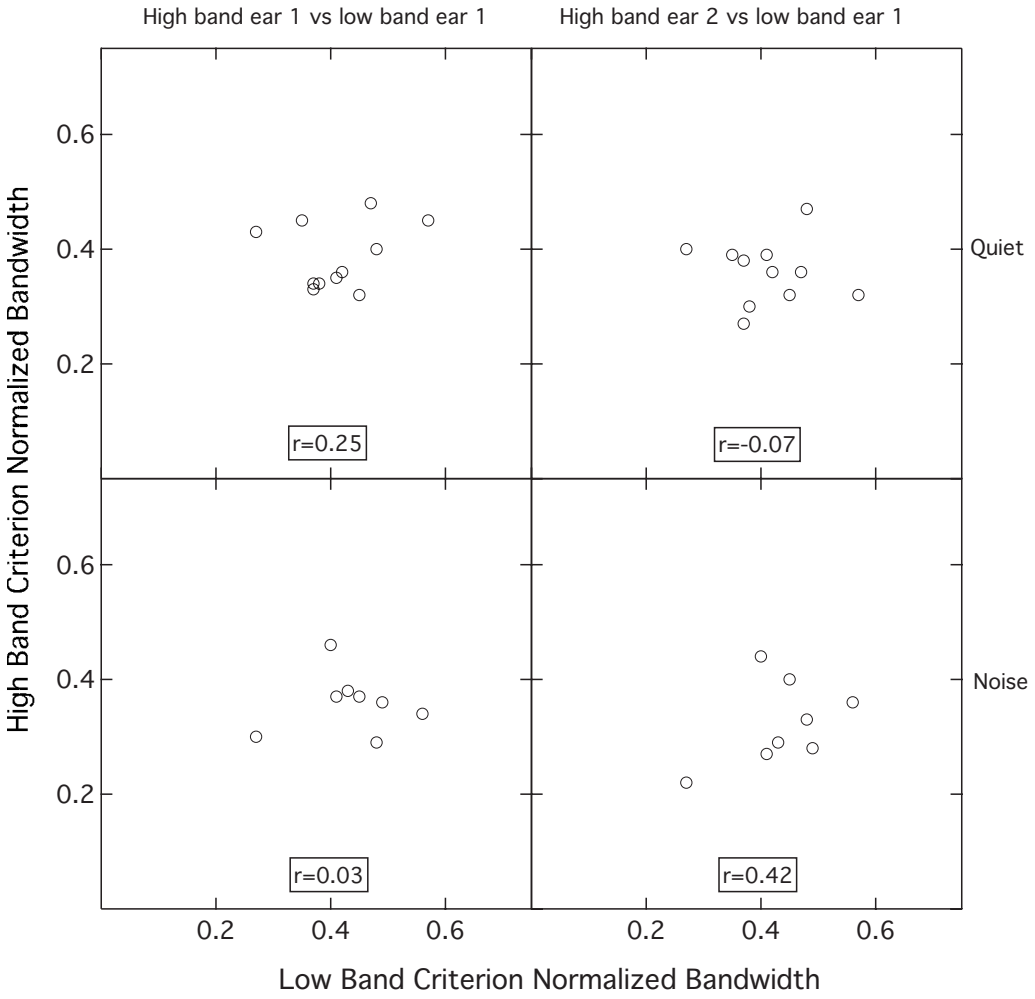


FIG. 1. Criterion normalized bandwidth for the high band vs criterion normalized bandwidth of the low band, with the correlation (r) shown in the box. Data are for normal-hearing listeners. The left-hand panels show data for ear 1 stimulation, and the right-hand panels show data for stimulation where the low band was presented to ear 1 and the high band was presented to ear 2. The upper and lower panels show data for listeners tested in quiet and noise, respectively.

3. Procedure

The listener sat in a double-walled sound booth and was instructed to repeat as many words as possible after each sentence was presented and to guess for words that were not intelligible. No feedback was provided. The experimenter was positioned in front of a visual display that showed the current sentence and monitored the listener's response via a talk-back loop. The experimenter recorded errors following each listener response. Stimuli were presented over Sennheiser HD 265 earphones.

A critical consideration was the specific bandwidth to which the speech was filtered at each center frequency, as the goal was to have the speech intelligibility fall within a relatively narrow range of poor performance (i.e., 15%–25% correct) in the single-band conditions. Whereas most previous approaches have used fixed bandwidths, the present approach instead allowed bandwidth to vary across listeners (Noordhoek *et al.*, 1999), partly because of the potential for intersubject variation in speech performance for a fixed bandwidth among hearing-impaired listeners. Another important reason for the adaptive approach in the first stage of testing was that it was desirable to home in on an appropriate speech bandwidth rapidly. Because of travel and other considerations, many of the hearing-impaired listeners were available for a relatively limited testing time and it was therefore critical to use efficient testing strategies.

The adaptive procedure was carried out separately for a band centered on 500 Hz, and for another band centered on 2500 Hz. At each center frequency the bandwidth was changed adaptively (by a factor of 1.21), with the bandwidth increasing following two sentences in a row where no key word was reported correctly, and with bandwidth decreasing following a sentence in which any key word was correctly reported. The run was stopped following eight reversals in bandwidth adjustment, and the threshold bandwidth was taken as the geometric mean of the bandwidths at the last six reversals. This threshold value will be referred to as the criterion speech bandwidth. Testing conducted prior to this study on ten normal-hearing listeners showed that the criterion speech bandwidth estimated from this stepping rule was associated with approximately 15%–25% correct (mean of 23.2% correct and standard deviation of 4.3% for the low band and mean of 20.3% correct and standard deviation of 4.7% for the high band) when listeners were retested with the bandwidth fixed at this criterion value.[1] There were three conditions evaluated in this stage of testing: (1) the low band

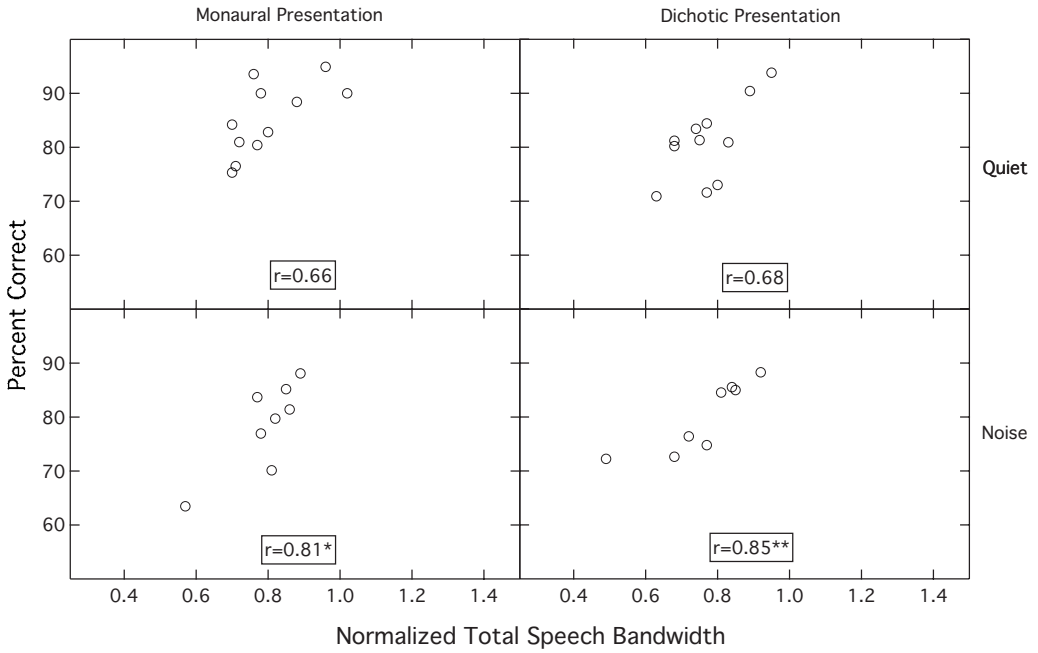


FIG. 2. Speech recognition percent correct vs total normalized bandwidth, with the correlation (r) shown in the box. Data are for normal-hearing listeners. The left-hand panels show results for monaural stimulation, and the right-hand panels show results for dichotic stimulation. The upper and lower panels show data for listeners tested in quiet and noise, respectively.

presented to ear 1; (2) the high band presented to ear 1; and (3) the high band presented to ear 2. For each of the normal-hearing listeners, ear 1 was randomly selected to be either the right or the left ear.

In the second stage of testing, fixed block trials were used to determine the percent correct speech identification obtained when the low and high bands were presented together. Each estimate of percent correct was determined using a single BKB list consisting of 16 sentences (one list), and three estimates were obtained in each condition. In this stage of testing, the bandwidth at each frequency was held constant at the criterion speech bandwidths that had been obtained in stage 1. There were two conditions in this stage of testing, each involving the simultaneous presentation of the low band and the high band: (1) both the low and high bands presented to ear 1 and (2) the low band presented to ear 1 and the high band presented to ear 2. The final estimate of percent correct in each condition was computed by applying an arcsine transformation (Studebaker, 1985) to the three replicate data points, averaging these values, and then converting the result back to percent correct.

B. Results and discussion

Figure 1 shows the relationships between the bandwidths estimated for the normal-hearing listeners at 500 and 2500 Hz for speech presented in both quiet and in noise. Data are expressed as the criterion speech bandwidth divided by the center frequency, a value referred to as the criterion normalized bandwidth. Panels on the left-hand side are associated with conditions where each band was presented to ear 1, and panels on the right-hand side are associated with conditions where the low band was presented to ear 1 and the high band was presented to ear 2. One finding apparent in Fig. 1 is that there was no obvious relation between the criterion normalized bandwidths at each center frequency: that is, listeners requiring a relatively wide criterion normalized bandwidth at 500 Hz did not necessarily require a wide bandwidth at 2500 Hz. The criterion normalized bandwidth ranged from 0.27 to 0.57 at 500 Hz, and from 0.22 to 0.48 at 2500 Hz. The criterion normalized bandwidth was relatively similar for the listeners tested in quiet and those tested in noise. This finding is consistent with an analysis that indicated that the signal-to-noise ratio was approximately 14 dB or greater in the presence of the pink noise masker, suggesting that speech audibility was not a limiting factor. A repeated measures analysis of variance was performed to examine this, with a within subjects factor of speech band condition (low band presented to ear 1, high band presented to ear 1, and high band presented to ear 2) and a between subjects factor of masker (quiet versus masking noise). In this analysis, the dependent variable was the criterion normalized bandwidth at each frequency. The analysis indicated a significant effect of speech band condition ($F_{2,34}=9.6$; $p=0.001$), but no effect of masking noise ($F_{2,34}=0.4$; $p=0.53$) and no interaction ($F_{2,34}=1.3$; $p=0.28$). Post hoc testing revealed that the significant effect of speech band was due to the criterion normalized bandwidth being slightly wider for the low band than for the high band.

Figure 2 plots the percent correct performance for the monaural and dichotic conditions where the low and high bands were presented simultaneously against a measure of the total speech bandwidth available in these conditions. The total speech bandwidth was defined as the sum of the criterion normalized bandwidths at 500 and 2500 Hz. For the monaural condition, the criterion normalized bandwidth for 2500 Hz was associated with ear 1, and for the dichotic condition, the criterion normalized bandwidth for 2500 Hz was associated with ear 2. It was of interest to plot the monaural

TABLE I. Correlation between percent correct obtained with the low- and high-frequency bands presented together either monaurally or dichotically and the criterion normalized bandwidth of the low band (low), the criterion normalized bandwidth of the high band (high), or the sum of these normalized bandwidths (both). Correlations that are significant at the 0.05 and 0.01 levels of probability are noted by an asterisk or double asterisk, respectively.

	Normal listeners								
	Quiet			Masking noise			Hearing-impaired listeners		
	Low	High	Both	Low	High	Both	Low	High	Both
Monaural	0.65*	0.35	0.66*	0.87**	0.16	0.81*	0.09	−0.34	0.53
Dichotic	0.45	0.51	0.68*	0.59	0.20	0.85**	0.23	−0.49	0.78*

and dichotic percent correct performance against these bandwidth metrics because of the possibility that listeners having relatively wide criterion bandwidths might perform relatively well in the case where the bands were presented simultaneously. This did indeed appear to be the case, with a trend for higher percent correct performance in cases where the total speech bandwidth was relatively wide (see Fig. 2). One interpretation of this effect is that some listeners are relatively poor in extracting information from a single band (and therefore require a large bandwidth for criterion performance), but that such listeners are better able to use the information when two bands are presented together. The columns of Table I dealing with normal-hearing listeners (left and middle) show associated correlations for the low and high bands alone and for these bands together. As can be seen, for the present listeners, there was some indication that the bandwidth of the low-frequency band may have contributed relatively more to the performance obtained when both bands were present, particularly for monaural presentation.

Another finding apparent in Fig. 2 is that the performance was relatively good regardless of whether stimulation was monaural or dichotic, or whether listeners were tested in quiet or in noise. The level of performance obtained across these conditions (approximately 64%–94% correct) was clearly greater than that obtained by additive combination of information present in the single-band conditions (which were associated with approximately 15%–25% correct). This result is consistent with previous demonstrations of speech band combination effects for frequency-separated bands (e.g., Grant and Braida, 1991; Warren *et al.*, 1995; Lippmann, 1996; Kasturi *et al.*, 2002). A repeated measures analysis of variance was performed with a within-subjects factor of mode of presentation (monaural versus dichotic presentation) and a between-subjects factor of masker (quiet versus masking noise). In this analysis, the dependent variable was the arcsine transformation of percent correct word identification. The analysis indicated no significant effect of presentation mode ($F_{1,17}=0.6$; $p=0.45$), no effect of masking noise ($F_{1,17}=1.9$; $p=0.18$), and no interaction ($F_{1,17}=2.25$; $p=0.15$). The fact that performance was relatively good in conditions where a background masking noise was present indicates that the speech band combination effect is relatively robust in listeners with normal hearing. This result suggests that for hearing-impaired listeners having thresholds similar to those simulated by the masking noise used here, little effect of audibility may be expected.

III. EXPERIMENT 2: HEARING-IMPAIRED LISTENERS

A. Methods

1. Listeners

There were nine hearing-impaired listeners, six females and three males. The listeners had an average age of 43.3 years with a standard deviation of 11.3 years. All listeners had mild-to-moderate sensorineural hearing losses as determined via air- and bone-conduction audiometry. Audiometric data and speech recognition scores (percent correct) for monosyllabic words presented in quiet for these listeners are shown in Table II. The assignment of ear 1 and ear 2 was random except in two cases (listeners 7 and 8). During audiometric testing, the responses of listener 7 to both speech and tones were relatively unreliable for right-ear presentation. This listener was therefore tested using the left ear only. Listener 8 had normal hearing in the right ear and so was tested using the left ear only. For this listener, a 40 dB HL speech-shaped noise was presented to the right ear during filtered speech testing in order to mask speech that may have crossed over from the left earphone.

TABLE II. Air-conduction audiograms (dB HL) and speech recognition scores (% correct) for monosyllabic words. Ear 1 (the ear tested with both the low and high bands presented the same ear) is identified in bold. "NT" indicates not tested.

		250	500	1000	2000	4000	8000	Recognition
HI1	**L**	**20**	**45**	**50**	**50**	**30**	**35**	**88**
	R	20	40	50	45	35	35	92
HI2	L	30	40	45	45	50	50	84
	R	**30**	**35**	**45**	**50**	**45**	**50**	**84**
HI3	L	55	70	75	50	65	75	84
	R	**45**	**55**	**55**	**50**	**60**	**70**	**76**
HI4	**L**	**25**	**25**	**50**	**50**	**45**	**40**	**88**
	R	25	25	45	45	45	35	88
HI5	L	35	35	25	30	40	60	100
	R	**25**	**30**	**30**	**30**	**45**	**65**	**100**
HI6	L	35	40	40	40	50	60	92
	R	**30**	**35**	**40**	**45**	**45**	**60**	**92**
HI7	**L**	**45**	**50**	**50**	**45**	**50**	**65**	**80**
	R	60	70	65	50	50	95	NT
HI8	**L**	**50**	**50**	**50**	**50**	**50**	**80**	**76**
	R	15	5	15	10	10	15	100
HI9	**L**	**50**	**50**	**45**	**40**	**40**	**40**	**84**
	R	60	60	65	60	65	75	52

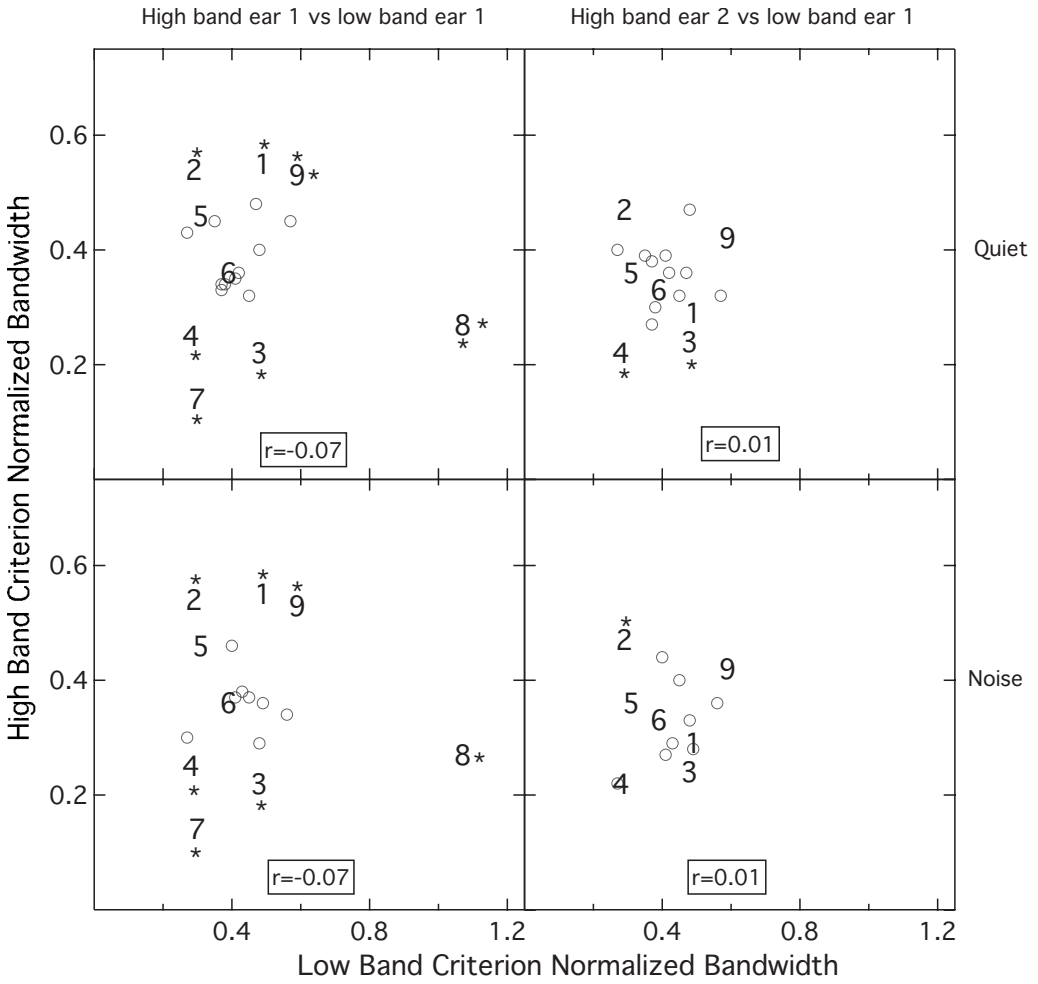


FIG. 3. Criterion normalized bandwidth for the high band vs criterion normalized bandwidth for the low band, with the correlation (r) shown in the box. Data for hearing-impaired listeners are depicted by listener number. The left-hand panels show data for ear 1 stimulation, and the right-hand panels show data for stimulation where the low band was presented to ear 1 and the high band was presented to ear 2. The upper panels allow comparison of the hearing-impaired listeners to the normal-hearing listeners who were tested in quiet, and the lower panels allow comparison of the hearing-impaired listeners to the normal-hearing listeners who were tested in noise (normal data are replotted from Fig. 1). Hearing-impaired listeners having criterion bandwidths outside the range of normal-hearing listeners are identified by an asterisk: Asterisks above and below the symbol signify high-band criterion bandwidths above and below the limits for normal-hearing listeners, respectively; asterisks to the right and to the left of the symbol signify low-band criterion bandwidths above and below the limits for normal-hearing listeners, respectively.

2. Methods and procedure

The methods and procedure were identical to those used in experiment 1, except that all speech was presented in quiet. Audibility of speech was judged at the outset as unlikely to limit performance of hearing-impaired listeners. As indicated in Table II, pure-tone thresholds at frequencies near the center of each speech band were 55 dB HL or better for all hearing-impaired listeners, similar to those of normal-hearing listeners tested in the presence of pink noise, as noted earlier. At threshold bandwidths estimated for noise-masked normal-hearing listeners, the signal-to-noise ratios at threshold were approximately 14 dB or greater, suggesting that speech was likely to be audible for the hearing-impaired listeners under these conditions.

B. Results and discussion

Figure 3 plots the criterion normalized bandwidth for the 500 Hz center frequency against the criterion normalized bandwidth for the 2500 Hz center frequency. The data for the normal-hearing listeners are replotted to aid comparison to the data of the hearing-impaired listeners, for whom data are identified by listener number. The upper panels are used to compare with normal-hearing listeners tested in quiet and the lower panels are used to compare with normal-hearing listeners tested in noise. As with the normal-hearing listeners, there was no apparent relation between the criterion normalized bandwidths at the two center frequencies. There were relatively large individual differences in the bandwidth necessary for criterion performance in the hearing-impaired listeners, with criterion normalized bandwidth ranging from approximately 0.28 to 1.06 Hz at 500 Hz, and from approximately 0.14 to 0.54 Hz at 2500 Hz. The criterion speech bandwidths obtained for the hearing-impaired listeners were broadly similar to those obtained by the normal-hearing listeners. Statistical analyses were performed to compare performance of the two groups. Because some of the hearing-impaired listeners were not tested in ear 2, repeated measures analyses of variance were performed to compare the normal-hearing and hearing-impaired listeners for the

two ear 1 conditions (low band and high band) and separate t-tests were performed to compare the normal-hearing and hearing-impaired listeners for the single ear 2 condition (high band). In all analyses, the dependent measure was the criterion normalized bandwidth.

Comparisons of the hearing-impaired listeners to the normal-hearing listeners tested in quiet will be considered first. The repeated measures analysis of variance had a within-subjects factor of condition (low band presented to ear 1 and high band presented to ear 1), and a between subjects factor of hearing loss (present or absent). The analysis indicated no significant effect of condition ($F_{1,18}=1.7$; $p=0.21$) or of hearing impairment ($F_{1,18}=0.15$; $p=0.70$) and no interaction ($F_{1,18}=0.55$; $p=0.47$). The t-test comparing the normal-hearing and hearing-impaired listeners for the high band presented to ear 2 also indicated no significant difference ($t_{16}=0.73$; $p=0.48$). The repeated measures analysis comparing the masked normal-hearing listeners to the hearing-impaired listeners indicated no significant effect of condition ($F_{1,15}=2.49$; $p=0.13$) or of hearing impairment ($F_{1,15}=0.16$; $p=0.69$) and no interaction ($F_{1,15}=0.04$; $p=0.84$). The t-test comparing the masked normal-hearing and the hearing-impaired listeners for the high band presented to ear 2 also indicated no significant difference ($t_{13}=0.23$; $p=0.82$).

At first glance, the result that the relative criterion bandwidth did not differ between normal and hearing-impaired listeners might appear to be in conflict with the results obtained by Noordhoek *et al.* (2000), where hearing-impaired listeners needed a wider bandwidth than normal to obtain 50% correct. There are at least three factors that should be considered in this regard:

(1) The listeners in the Noordhoek *et al.* study were tested using bandlimited speech presented in a complementary band-stop masking noise. It is possible that relatively poor frequency selectivity, common in sensorineural hearing loss (e.g., Tyler *et al.*, 1984; Stelmachowicz *et al.*, 1985; Leek and Summers, 1996), resulted in a greater masking effect for hearing-impaired than normal-hearing listeners, perhaps accounting for the need for a wider speech bandwidth in the hearing-impaired listeners of that study. The present finding that normal-hearing and hearing-impaired listeners required broadly similar speech bandwidth for criterion performance is consistent with the previous results of Grant *et al.* (2007), which indicated that, for the same narrow speech bandwidth, normal-hearing and hearing-impaired listeners obtained approximately the same, relatively low percent correct.
(2) The Noordhoek *et al.* study tracked 50% intelligibility and therefore required a larger bandwidth than the present study where a lower intelligibility was tracked. It is possible that the wider bandwidth tracked in the Noordhoek *et al.* study resulted in effects of hearing impairment related either to within-band masking effects or to a smaller than normal number of effectively independent frequency channels at the speech bandwidth associated with normal performance.
(3) Because the variability was relatively great among the hearing-impaired listeners in the present study, it is important to consider whether some of the results of the hearing-impaired listeners fell outside the normal range even though there was no overall group difference. This was addressed by assessing whether individual hearing-impaired listeners had criterion normalized bandwidths that were more than 2 s.d. above or below the normal mean. In Fig. 3, listeners having bandwidths above the normal limit are identified by an asterisk above the listener number for the high band and to the right of the listener number for the low band; listeners having bandwidths below the normal limit are identified by an asterisk below the listener number for the high band. No hearing-impaired listener had a bandwidth below the normal limit for the low band. With respect to the group of normal-hearing listeners tested in quiet, this analysis indicated the following. For the low band, none of the hearing-impaired listeners had criterion bandwidths narrower than the normal limit, and listeners 8 and 9 had bandwidths wider than the normal limit; for the high band presented to ear 1, listeners 3, 4, 7, and 8 had criterion bandwidths narrower than the normal limit, and listeners 1, 2, and 9 had criterion bandwidths wider than the normal limit. For the high band presented to ear 2, listeners 3 and 4 had criterion bandwidths narrower than the normal limit, and none of the listeners had criterion bandwidths wider than the normal limit. With respect to the group of normal-hearing listeners tested in masking noise, this analysis indicated the following. For the low band, none of the hearing-impaired listeners had criterion bandwidths narrower than the normal limit, and listener 8 had a bandwidth wider than the normal limit; for the high band presented to ear 1, listeners 3, 4, and 7 had criterion bandwidths narrower than the normal limit, and listeners 1, 2, and 9 had criterion bandwidths wider than the normal limit. For the high band presented to ear 2, none of the hearing-impaired listeners had criterion bandwidths wider or narrower than the normal limit. Overall, the results indicate that the criterion speech bandwidth was variable in the hearing-impaired group, with some listeners requiring narrower than normal values (for the high band) and others requiring wider than normal values (for the low and high bands).

Figure 4 plots the percent correct performance for the monaural and dichotic conditions where the low and high bands were presented simultaneously against the measure of the total normalized bandwidth available in these conditions. Figure 4 again replots the normal data in order to aid comparison. The right-most portion of Table I shows associated correlations for the low and high bands alone and for these bands together. As with the normal-hearing listeners, there was a trend for the hearing-impaired listeners with the largest total normalized bandwidth to have higher percent correct scores, although this was statistically significant only for the dichotic case.

The most notable finding was that the performance of the hearing-impaired listeners was generally quite good when both the low and high bands were present and was

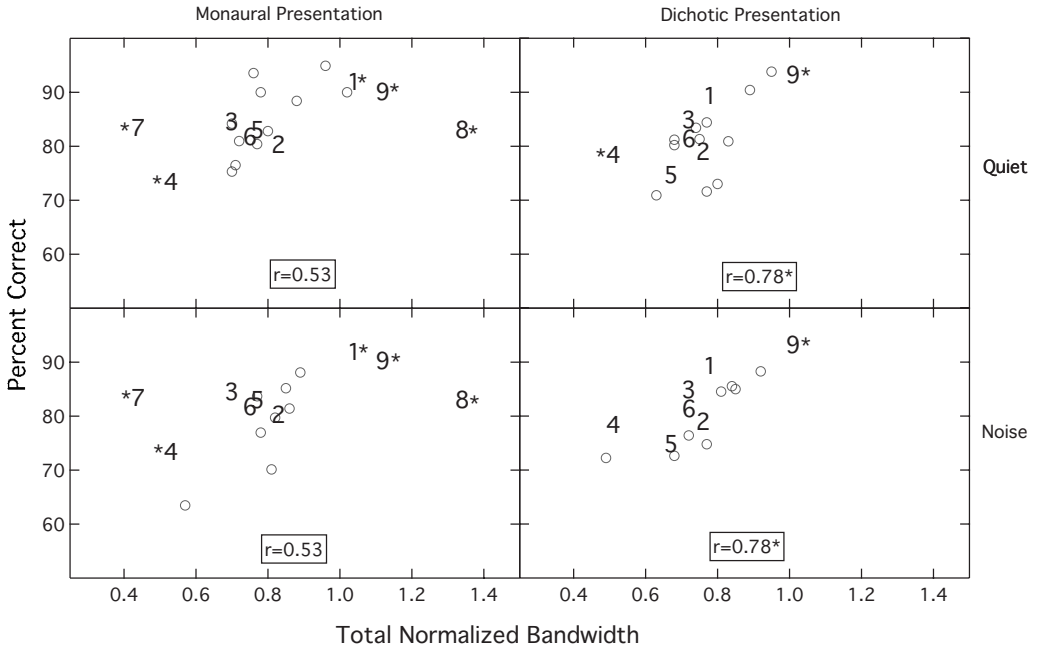


FIG. 4. Speech recognition percent correct vs total normalized bandwidth, with the correlation (r) shown in the box. Data for hearing-impaired listeners are depicted by listener number. The left-hand panels show data for monaural stimulation, and the right-hand panels show data for dichotic stimulation. The upper panels allow comparison of the hearing-impaired listeners to the normal-hearing listeners who were tested in quiet, and the lower panels allow comparison of the hearing-impaired listeners to the normal-hearing listeners who were tested in noise (normal data are replotted from Fig. 2). Hearing-impaired listeners having percent correct outside the range of normal-hearing listeners are identified by an asterisk: Asterisks above and below the symbol signify percent correct above and below the limits for normal-hearing listeners, respectively; asterisks to the right and to the left of the symbol signify total normalized bandwidths above and below the limits for normal-hearing listeners, respectively.

comparable to that for the normal-hearing listeners. t-tests were performed on the arcsine-transformed percent correct data to evaluate possible differences between the normal and hearing-impaired listeners. This testing indicated that the hearing-impaired listeners did not differ significantly from the normal-hearing listeners tested in quiet ($t_{18}=0.60$; $p=0.56$) or in noise ($t_{15}=1.49$; $p=0.16$). This finding also held for dichotic presentation both in quiet ($t_{16}=0.59$; $p=0.57$) and in noise ($t_{13}=0.91$; $p=0.38$). Because the performance of hearing-impaired listeners is often marked by high variability, it is also important to evaluate possible outliers within the impaired group. This was again assessed by determining whether hearing-impaired listeners fell more than 2 s.d. above or below the normal mean for either monaural or dichotic stimulation. Again, this was evaluated with respect to the arcsine-transformed percent correct data. The results were that none of the data of the hearing-impaired listeners fell outside the normal limit.

As noted in Sec. II, pilot data on ten normal-hearing listeners showed that the criterion speech bandwidth estimated from the adaptive testing was associated with approximately 15%–25% correct when listeners were retested in fixed blocks at this criterion bandwidth. Because no fixed-block testing was performed with hearing-impaired listeners for a single band at this criterion bandwidth, an additional analysis was done to evaluate the assumption that the initial, adaptive stage of testing converged on about the same percent correct for the normal-hearing and hearing-impaired listeners. The analysis was based upon the data from the adaptive tracks of both the normal-hearing and hearing-impaired listeners for either the low band or high band presented to ear 1. In the analysis, the bandwidths visited in each adaptive track were binned into equal log steps and psychometric functions were estimated with a linear fit (proportion correct plotted against bandwidth). Because fits based upon individual, raw data were relatively poor, data within each group were combined and normalized to the mean criterion bandwidth for each group. For example, if the mean criterion bandwidth (computed on log transform data) was 300 Hz for a listener but was 200 Hz for the group, the bandwidths for that individual were multiplied by a factor of 2/3. This procedure clustered the functions of all listeners around the centroid of the group without affecting the individual function slopes. The results of this procedure are shown in Fig. 5, with the size of the symbol reflecting the number of points contributing to the associated estimate. The line fitted to these data was used to estimate the proportion of correct responses associated with the mean criterion speech bandwidths obtained in the adaptive tracks; these values of proportion correct were approximately 0.14–0.17 for both the normal-hearing and hearing-impaired listeners. This analysis therefore supports the assumption that the adaptive threshold testing converged upon approximately the same level of speech recognition performance for the two groups of listeners.

IV. GENERAL DISCUSSION

A. Integration of spectrally separated speech information

The central question evaluated in this study was whether listeners with mild–moderate sensorineural hearing impair-

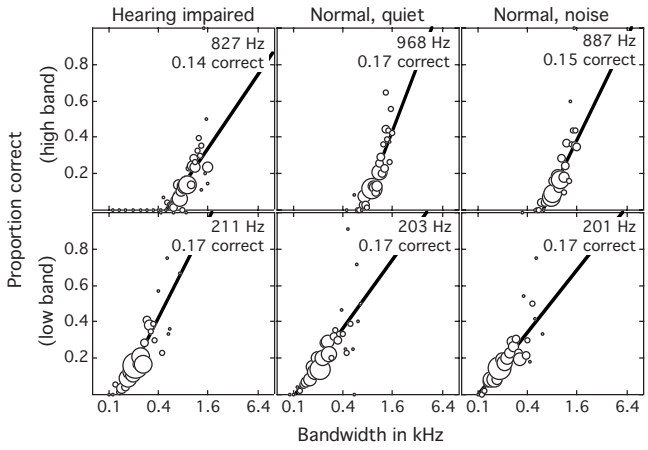


FIG. 5. Proportion correct plotted as a function of speech bandwidth (see the text for details). Functions are fitted to data from the adaptive track stage of testing for hearing-impaired listeners and normal-hearing listeners in quiet and in noise.

ment have an essential deficit in the ability to integrate information from simultaneous, frequency-separated narrow bands of filtered speech. In normal-hearing listeners, narrow bands of speech from widely spaced spectral regions can combine to produce a percent correct well above the sum of the percent correct values associated with each band separately. The results of the present study indicated that the hearing-impaired listeners showed a similar ability to combine speech information from frequency-separated bands. This occurred both for monaural and for dichotic presentation.

Previous vocoder-based speech results have suggested that sensorineural hearing loss may be associated with relatively poor across-frequency integration of speech information, and one interpretation considered by Grant *et al.* (2007) in their filtered speech study was also based upon poor across-frequency integration by hearing-impaired listeners. The results of the present study are not consistent with an interpretation that sensorineural hearing loss is associated with an essential deficit in the ability to integrate across-frequency speech information. However, the present results are not necessarily in conflict with those of the previous vocoder studies or with the results of Grant *et al.* (2007). For example, it is possible that the different pattern of results in the present study and the previous vocoder studies is related to differences in stimuli. Such stimulus differences could include those related to processing (vocoding versus filtering) and to level (high-frequency speech energy in the present study was boosted in level in order to avoid effects related to upward spread of masking). The issue of the level of high-frequency speech energy is also relevant to comparisons between the present study and the study of Grant *et al.* Although Grant *et al.* noted that a deficit in the ability to integrate across-frequency speech information could have been the basis for their filtered speech results, they also noted that other factors could have been at work, including increased upward spread of masking in the impaired listeners. Because the present stimuli had features designed to minimize effects of upward spread of masking, the present results should not be interpreted as being in conflict with those of Grant *et al.* (2007).

Although the present results on the effect of sensorineural hearing impairment on the ability to integrate across-frequency speech information were relatively straightforward, there are nevertheless reasons to interpret them with some caution. One reason is that the listeners of this study had mild–moderate hearing losses. It is possible that speech integration results would be different for listeners with more severe hearing loss. A potential difficulty in interpreting results obtained from listeners with severe hearing loss is that reduced speech perception abilities (even with the complete speech spectrum available) might put a ceiling on the magnitude of speech combination effects for frequency-separated bands. Another reason for caution in interpreting the present results is that the listeners of this study had relatively flat hearing loss configurations. It is possible that sloping hearing losses may be associated with different speech integration abilities.

B. Criterion speech bandwidths at 500 and 2500 Hz

Although the primary purpose of the present study was to examine the effect of hearing impairment on the ability to integrate speech information across frequency, the findings on the criterion speech bandwidth are also of interest. There was considerable variability among the normal-hearing listeners on the criterion bandwidth measure, and even greater variability among the hearing-impaired listeners. Perhaps most striking in this regard is that, for monaural presentation when both the low and high bands were present simultaneously, more hearing-impaired listeners fell outside the range for normal-hearing listeners (two listeners below and three above) than inside this range in terms of total normalized bandwidth (see Fig. 4). This result was not predicted and it is not readily accounted for. One possible interpretation of the wider bandwidth required by some hearing-impaired listeners is based on the fidelity with which speech information is encoded: if encoding of information is somehow impaired, then criterion performance would require more information via greater bandwidth. It is more challenging to account for the finding that some hearing-impaired listeners required a narrower than normal total speech bandwidth for criterion performance (a finding that appears to have been dominated by the bandwidth of the high band). One possibility is that some hearing-impaired listeners may adapt to the abnormal speech patterns available at the outputs of their relatively wide auditory filters. For example, whereas the temporal envelope of a speech stimulus at the output of a relatively wide auditory filter may be abnormal, it may nevertheless carry information that listeners can learn to use to differentiate among speech sounds. One specific possibility is that hearing-impaired listeners may learn to make use of relatively high-rate modulation cues related to the fundamental frequency, which may, in turn, provide information related to voicing and pitch (e.g., Arehart, 1994).

This study also yielded information about the relation between bandwidths required to support a criterion level of

performance at two separated frequency regions. While there was no experimental hypothesis about this relationship, one expectation that seems reasonable is that listeners who require a relatively narrow bandwidth at one region would also require a narrow bandwidth at another region. However, across the listeners tested here (normal-hearing listeners tested in quiet and in noise and hearing-impaired listeners) the correlation between the criterion speech bandwidths associated with the two frequency regions was consistently close to zero. This would imply that performance was not dominated by some general speech processing factor that applies across bandwidth in speech perception. Instead, it may point to the importance of processing factors that are specialized with respect to frequency region and vary independently across listeners.

V. CONCLUSIONS

(1) Listeners with sensorineural hearing impairment showed an ability to combine information from frequency-separated bands of speech that was similar to that demonstrated by normal-hearing listeners. This occurred both for monaural and dichotic stimulation. These results are consistent with an interpretation that listeners with mild–moderate sensorineural hearing loss do not have an essential deficit in the ability to combine across-frequency speech information.
(2) Neither normal-hearing nor hearing-impaired listeners showed a significant correlation between the criterion speech bandwidths at the two frequency regions examined here.
(3) Speech recognition performance when both the low and high bands were presented simultaneously tended to be better for listeners having relatively wide criterion bandwidths. This was true for both normal-hearing and hearing-impaired listeners.
(4) On average, listeners with sensorineural hearing impairment required criterion speech bandwidths that were similar to normal at center frequencies of 500 and 2500 Hz. However, there was relatively great variation in the criterion speech bandwidths of the hearing-impaired listeners. In some conditions, there were hearing-impaired listeners who had criterion speech bandwidths that were narrower than the normal limit and others who had criterion speech bandwidths that were wider than the normal limit.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant No. R01 DC00418. Professor Brian Moore provided several very helpful comments. Two anonymous reviewers also offered valuable suggestions that improved the quality of this manuscript.

[1]Although the percent correct converged upon in typical tracking procedures involving stimuli presented over independent trials can be calculated in a straightforward manner (Levitt, 1971), such a calculation is more complex in the current procedure, where a number of words are presented within a sentence and the probability of correctly identifying a particular key word may depend in part on the semantic context of the other words in the sentence. For this reason, the percent correct associated with the adaptively measured criterion bandwidth was determined empirically.

Arehart, K. H. (**1994**). "Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3574–3585.
Assmann, P. F., and Summerfield, A. Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York).
Bacon, S. P., and Gleitman, R. M. (**1992**). "Modulation detection in subjects with relatively flat hearing losses," J. Speech Hear. Res. **35**, 642–653.
Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.
Buss, E., Hall, J. W. III, and Grose, J. H. (**2004**). "Spectral integration of synchronous and asynchronous cues to consonant identification," J. Acoust. Soc. Am. **115**, 2278–2285.
Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.
Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2006**). "Spectral and threshold effects on recognition of speech at higher-than-normal levels," J. Acoust. Soc. Am. **120**, 310–320.
French, N., and Steinberg, J. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Gagné, J. P. (**1988**). "Excess masking among listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **83**, 2311–2321.
Grant, K. W., and Braida, L. D. (**1991**). "Evaluating the articulation index for auditory-visual input," J. Acoust. Soc. Am. **89**, 2952–2960.
Grant, K. W., Tufts, J. B., and Greenberg, S. (**2007**). "Integration efficiency for speech perception within and across sensory modalities by normal-hearing and hearing-impaired individuals," J. Acoust. Soc. Am. **121**, 1164–1176.
Hall, J. W., Buss, E., and Grose, J. H. (**2008**). "The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency," J. Acoust. Soc. Am. **123**, 955–962.
Healy, E. W., and Bacon, S. P. (**2002**). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1262–1275.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard noise'," J. Acoust. Soc. Am. **93**, 2915–2922.
Kasturi, K., Loizou, P. C., Dorman, M., and Spahr, T. (**2002**). "The intelligibility of speech with 'holes' in the spectrum," J. Acoust. Soc. Am. **112**, 1102–1111.
Leek, M. R., and Summers, V. (**1996**). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am. **100**, 1796–1806.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lippmann, R. (**1996**). "Accurate consonant perception without mid-frequency speech energy," IEEE Trans. Speech Audio Process. **4**, 66–69.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Moore, B. C., Shailer, M. J., and Schooneveldt, G. P. (**1992**). "Temporal modulation transfer functions for band-limited noise in subjects with cochlear hearing loss," Br. J. Audiol. **26**, 229–237.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners," J. Acoust. Soc. Am. **105**, 2895–2902.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2000**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," J. Acoust. Soc. Am. **107**, 1685–1696.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., and Mott, J. (**1985**). "Speech perception ability and psychophysical tuning curves in hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 620–627.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Turner, C. W., Chi, S. L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.
Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C. J., and Patterson, R. D. (**1984**). "Auditory filter asymmetry in the hearing impaired," J. Acoust. Soc. Am. **76**, 1363–1376.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (**1995**). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Percept. Psychophys. **57**, 175–182.

Predicting the path of a changing sound: Velocity tracking and auditory continuity

Poppy A. C. Crum and Ervin R. Hafter
University of California Berkeley, Berkeley, CA 94720

(Received 6 August 2007; revised 6 May 2008; accepted 16 May 2008)

Three studies demonstrate listeners' ability to use the rate of a sound's frequency change (velocity) to predict how the spectral path of the sound is likely to evolve, even in the event of an occlusion. Experiments 1 and 2 use a modified probe-signal method to measure attentional filters and demonstrate increased detection to sounds falling along implied paths of constant-linear velocity. Experiment 3 shows listeners perceive a suprathreshold tone as falling along a trajectory of constant velocity when the frequency is near to the region of greatest detection as measured in Experiments 1 and 2. Further, results show greater accuracy and decreased bias in the use of velocity information with increased exposure to a constant-velocity sound. As the duration of occlusion lengthens, results also show a downward shift (relative to a trajectory of constant velocity) in the frequency at which listeners' detection and experience of a continuous trajectory are greatest. A preliminary model of velocity processing is proposed to account for this downward shift. Results show listeners' use of velocity in extrapolating sounds with dynamically changing spectral and temporal properties and provide evidence for its role in perceptual auditory continuity within a noisy acoustic environment. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945117]

PACS number(s): 43.66.Fe, 43.66.Cb, 43.66.Lj [RYL] Pages: 1116–1129

I. INTRODUCTION

The perception of a visual source moving through space is influenced by predictions of the path of motion (Orban de Xivry *et al.*, 2006; Churchland *et al.*, 2003; Rushton and Wann, 1999; Assad and Maunsell, 1995; Yantis, 1995; Pavel and Stone, 1992). As a source travels along a path, it generates an expectation for that source's location at a future moment in time (Pavel and Stone, 1992). For instance, when a fixated observer sees a ball roll behind a screen, properties of the ball's motion such as trajectory and velocity, initially represented as a pattern of activity moving across the retina, provide the observer with a reasonably accurate estimate of when and where it will reappear. Here, we ask whether the auditory system uses properties such as the rate of frequency change, corresponding to a pattern of activity moving along the basilar membrane, to make similar predictions about the path of a sound changing over time in frequency space, i.e., when a sound disappears behind an auditory screen, are there expectations in the sensory system of where it should reemerge?

It is well known that interruptions in a sound, due to masking and occlusion by other more intense sounds, often go perceptually undetected (Miller and Licklider, 1950; Warren, 1999; Dannenbring, 1976). This is true even when a sound's frequency is changing across time [as during a frequency-modulated sweep (FM-sweep), Fig. 1(a)] and is truly discontinuous [Fig. 1(b)] but contains higher intensity noise during the discontinuity [Fig. 1(c)]. In this instance, energy present during the interruption that is sufficient to mask the absent components creates an illusory condition where the sound appears continuous [Fig. 1(d)] despite discontinuity of the original signal (Houtgast, 1972; Warren *et al.*, 1972). In a noisy environment, this ability is crucial. It enables a listener to successfully encode and interpret complex sounds where masking from more intense sounds has degraded the content of the original source and perceptual restoration of information is necessary to achieve a more useful experience of the sound's content. For hearing-impaired listeners this phenomenon may be even more critical. In conditions where the effective signal-to-noise ratio is lower the perceptual restoration of occluded content is likely more heavily relied upon to maintain a source's stability and contextual relevance in the acoustic scene.

Referred to as auditory continuity or auditory induction, this phenomenon has been demonstrated throughout studies of auditory processing (Miller and Licklider, 1950; Bregman, 1997; Warren, 1999), and in many conditions has led to a relevant perceptual representation of the missing or occluded information despite dynamic shifts in the sound's spectral content across time. In these instances, properties of the sound prior to an occlusion may be used to induce the probable path of change during an interruption.

This tendency of sounds interrupted by noise to be perceptually continuous (Warren *et al.*, 1972; Warren 1999; Ciocca and Bregman, 1987; Kluender and Jenison, 1992; Bregman, 1994) has historically been regarded as a problem resolved by interpolation between audible regions of a sound across an occlusion (Warren, 1999; Grossberg *et al.*, 2003; Husain *et al.*, 2005). In this scenario, perceptual restoration is accomplished through a best-fit approximation between information occurring before and after the interruption. In contrast, perceptual restoration during an interruption could occur through extrapolation as previously described in the visual analogy. This scenario would, instead, depend on prediction of the spectral and temporal path of the changing sound as indicated from the portion of the sound that preceded the interruption. Use of information such as the rate of

 0001-4966/2008/124(2)/1116/14/$23.00

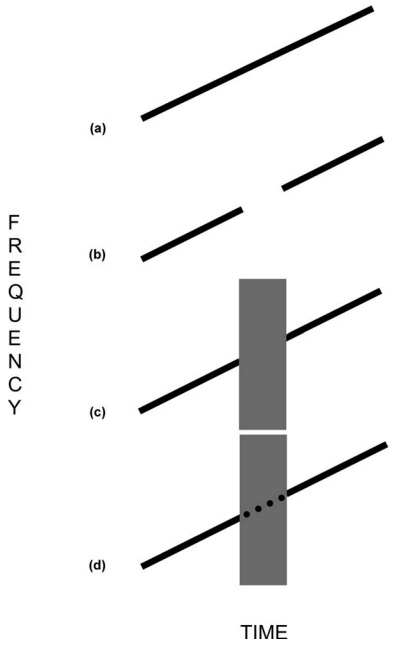


FIG. 1. A cartoon representation of perceptual continuity for sounds changing in frequency across time. (a) A continuous stimulus. (b) Two discrete sections interrupted by a silence. (c) Two discrete sections interrupted by an N-occluder. (d) The perceptual restoration that occurs during conditions of (c).

frequency change to extrapolate the path of a dynamically changing sound would suggest a computationally more efficient system for resolving occluded information than previously discussed. In an extrapolative scenario, listeners' narrow and position their listening region in expectation of the continuation of an occluded sound rather than monitoring multiple frequency regions (necessary in an interpolative scenario) for a continuation and interpolating across the interruption only after energy has been detected in a subsequent band. There are similarities between the predictive processes which occur while observing an object moving in time through visual space (Orban de Xivry *et al.*, 2006; Churchland *et al.*, 2003; Pavel and Stone, 1992; Assad and Maunsell 1995; Yantis, 1995) and a sound changing in time through frequency space. Understanding these similarities should elevate the role of extrapolation within current views of auditory processing.

To address the influence of continuously changing spectral and temporal information on listeners' perception, we measured whether listeners had an expectation of where an auditory trajectory should reappear when it disappears behind an occluder. Previously, Kluender and Jenison (1993) allowed listeners to adjust the starting frequency of the second of two disjoint FM-sweeps across a noise-occluder (N-occluder) to produce the strongest perception of continuity. Listeners' most often heard two sweeps separated by a N-occluder as optimally continuous when the sweeps were aligned along a trajectory of constant rate of change. While Kluender and Jenison (1992) discussed both the influence of extrapolation and interpolation on the perception of auditory continuity, their method did not allow extraction of the independent contribution of extrapolation without the influence of interpolation. Both FM-sweeps were audible, and, consequently, the possibility that listeners were interpolating between audible portions of each sweep cannot be discounted. In order to isolate the role of extrapolation, we measured changes in listeners' ability to detect frequencies falling on and in close proximity to a path predicted by a constant rate of frequency change induced by an FM-sweep/N-occluder combination. (A constant rate of frequency change will hereafter be referred to as frequency velocity with the implication that a sound increasing or decreasing in frequency across time invokes a spatially shifting pattern of response across the basilar membrane and subsequent areas of the auditory pathway that can be interpreted as having a velocity.) A modified probe-signal method (Schlauch and Hafter, 1991) was used to measure changes in listeners' detection of low-level pure-tone signals of variable frequencies following an FM-sweep/N-occluder cue. The probe-signal method measured change in listeners' threshold detectability to frequencies falling at, and surrounding, frequencies indicated by a preceding auditory stimulus. This enabled use of simple detection tasks to observe both the frequency locations and breadth (typically bandpassed in shape) of change in listeners' thresholds induced by the cue.

The indirect nature of the probe-signal method was crucial. It allowed measurement of which frequency region a listener was attending to along an implied path without relying on a listener's answer regarding where an estimated signal should appear. Rather, it implied a sensory change in the expectation for the location of the signal without direct input from the listener (Schlauch and Hafter, 1991; Hafter *et al.*, 1993; Wright and Dai, 1993). Therefore, unlike studies of interpolation between two disjoint FM-sweeps (Kluender and Jenison, 1992; Ciocca and Bregman, 1987), a listener here would have had to rely solely on extrapolation from a single sweep. While these two listening strategies need not be mutually exclusive, the purpose of the present work was to determine whether, in the frequency domain, listeners used frequency velocity to track how sounds change and whether this information could be used to extrapolate predictions of a sound's projected path, even in the presence of an occluder. It is well known that higher-order processes such as a listener's familiarity with the semantics of speech can influence the perceptual restoration of occluded content (Bashford and Warren, 1987b; Sivonen *et al.*, 2006). However, the aim of our studies was to see whether listeners showed a relative increase in detection of a low-level stimulus (hereafter referred to as sensitivity) along a projected path. This would imply the use of velocity tracking and emphasize the importance of a lower-level process where, potentially, changes in lower sensory processing contributed to the restoration of occluded content.

Three experiments were completed. Experiments 1 and 2 used a modified probe-signal method to measure detection of sounds falling along an implied path of constant-linear-velocity. Experiment 1 asked whether listeners were able to use velocity information to predict the frequency location of a sound following an N-occluder and further whether the duration of constant-velocity information preceding the interruption affected listeners' ability to do this task. Experiment 2 asked whether the duration across which listeners were

required to extrapolate affected their prediction, and, finally, Experiment 3 used suprathreshold tones to determine listeners' perceptions of optimally continuous trajectories in order to address the influence of extrapolation on the perceptual phenomenon of auditory continuity.

II. METHODS OF EXPERIMENTS 1 AND 2

A modified probe-signal method (Schlauch and Hafter, 1991) was used to measure listeners' ability to detect a low-level (90% detection) probe tone of varying frequency following an above-threshold FM-sweep/N-occluder combination. Studies using a modified probe-signal method have measured the detection of low-level pure-tones in background noise following cues of the same frequency or frequencies in close relational proximity (i.e., $\pm 10\% - 20\%$) (MacMillan and Schwartz, 1975; Wright and Dai, 1993; Schlauch and Hafter, 1991). Previously, a majority of experiments using probe-signal methods used pure-tone "cues" that were either the same as or bore a constant ratio to the expected frequency (Scharf *et al.*, 1987; Schlauch and Hafter, 1991; Hafter *et al.*, 1993; Wright and Dai, 1993). Results from these studies have demonstrated a significant increase in listeners' performance for frequencies near to the implied frequency of the cue relative to frequencies that were not. A paucity of experiments used cues that changed dynamically across time (Ebata *et al.*, 2001; Wright and Dai, 1998). However, the spectra of cues in these conditions also contained energy that directly stimulated the frequency region of the probe signal. In contrast, the present studies used cues that only contained spectral energy outside the frequency region of the probe signal and, instead, implied the frequency region of the probe solely through extrapolation of dynamic properties of the cue.

A. Stimulus and experimental design

1. General procedure

On each trial, listeners heard an FM-sweep followed by an N-occluder (exemplar conditions used in Experiment 1 are shown in Fig. 2) where the duration and starting frequency of the FM-sweep varied from trial to trial. This combination will be referred to as the FM-sweep/N-occluder cue. On half of the trials, the cue was followed immediately by a low-level 100-ms pure-tone. This combination will be referred to as a signal trial. The pure-tone was absent for the remaining 50% of the trials. This combination will be referred to as a noise trial. All signal and noise stimuli were presented in a continuous low-level background noise masker that was bandpass filtered between 125 and 5125 Hz with a spectrum level of 20 dB/Hz with regard to 20 μPa. Sixty-percent of the time the signal fell at the expected frequency, as implied by the combination of the velocity of the FM-sweep and the duration of the N-occluder (Fig. 2). For the remaining 40% of the trials, the signal fell half of the time at either +10% or −10% of the expected frequency and half of the time at either +20% or −20% of the expected frequency. The level of the signal varied as a function of frequency and was determined by the listener's 90% equal-detectability function (EDF, see Appendix A for description

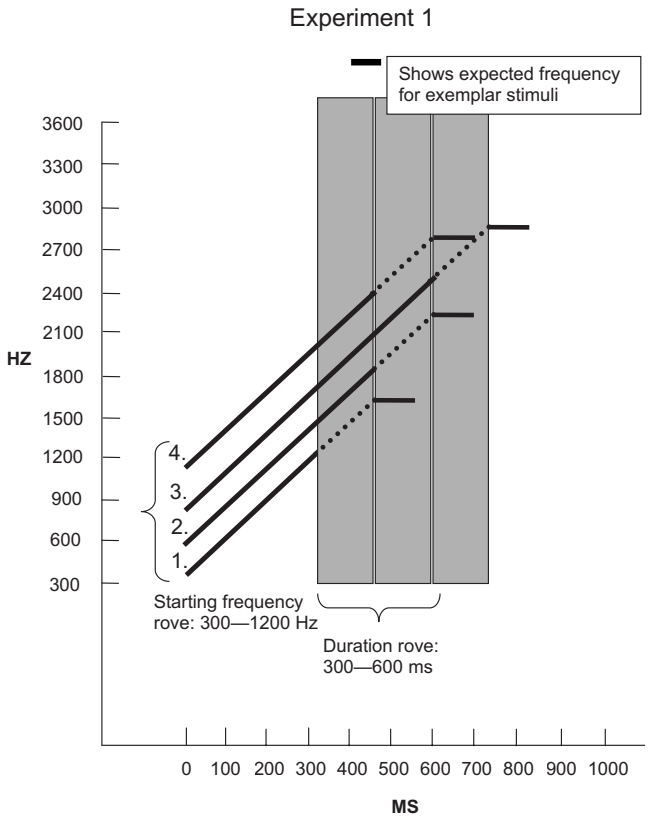


FIG. 2. Stimulus design for Experiment 1. Solid lines represent cartoon exemplar stimuli used in Experiment 1. Stimuli 1–4 demonstrate both a rove on start frequency and initial-sweep duration.

of methods used to establish listeners' EDFs) (Green *et al.*, 1959) for detection of a pure-tone in low-level background noise (having spectral characteristics as previously described).

In a single-interval yes-no task, listeners responded "yes" if they detected a signal following the N-occluder and "no" if they did not detect a signal. Response was indicated through a mouse click on a graphical user interface. After making a response, listeners received immediate visual feedback indicating whether they were correct or incorrect. Listeners were given unlimited time to respond, and the subsequent trial began approximately 1 s following a response.

2. Trial presentation

A testing block consisted of 60 trials, and a listener heard between 40 and 80 blocks. For each stimulus, the cue duration and starting frequency were unique. Cue durations were roved in equal intervals between 300 and 600 ms so that the final distribution the listener heard across all blocks was uniform in linear space. Starting frequency was roved in logarithmically spaced equal intervals between 300 and 1200 Hz so that the final distribution the listener heard across all blocks was uniform in log space. Thirty rove levels on each parameter were used in a single block so that every trial contained a unique stimulus (each parameter value was presented as both a signal and noise stimulus creating 60 unique trials per block). However, for each block a different set of randomly sampled parameter values was drawn from uniform distributions spanning the rove ranges. For every

listener, at least 2400 unique stimuli were used in testing that, when analyzed across multiple blocks, enabled a dense sampling of each parameter across the total rove range and prohibited the listener from learning any properties of the cue other than the intended velocity.

3. Specifications of the FM sweep

Hafter *et al.* (1993) demonstrated that listeners could use the relationship between a cue and a signal of a musical fifth to increase detection of randomly presented tones in noise. While many varieties of cues have been shown to enhance detection, this study was one of few that measured a change in detection to a frequency region outside the range of the cue. For this reason, FM-sweeps used in the current study were linear rather than logarithmic to avoid listeners' use of a musical relationship between starting frequencies of the sweep and the frequency of the signal. While attention was taken to limit the potential correlation between starting frequency and signal frequency by roving the starting frequency, the duration of the FM-sweep, and the duration of the N-occluder (Experiment 2), the decision to use only upward FM-sweeps (rather than mixing upward and downward sweeps) necessitated some degree of correlation between the two variables. Nonetheless, use of a linear sweep prohibited listeners from using a musical relationship to make direct estimates of the signal frequency from the starting frequency. The overall range for the initial-sweep duration of 300–600 ms was chosen to be predominantly outside the temporal limits of energy integration (Green *et al.*, 1957; Stephens, 1973). This was done to avoid change in listeners' performance as a function of duration occurring as a result of absolute stimulus detection rather than as a change in the perception of velocity. The velocity of the FM-sweep ($\Delta f/\Delta t$, where f=frequency and t=duration) was always 3000 Hz/s. Furthermore, linear sweeps were judged to produce a strong experience of perceptual continuity in pilot listening by the first author and members of the laboratory.

4. Specifications of the noise occluder

The duration of the N-occluder in Experiment 1 was fixed at 125 ms. In Experiment 2, the duration of the N-occluder was roved across a range of 50–200 ms with occluder duration densely sampled on a trial-by-trial basis from a linear uniform distribution across this range. Figure 3 demonstrates exemplar conditions used in Experiment 2. For all experiments, Gaussian noise was filtered to have a flat spectrum between 200 Hz and 5 kHz, and the start and end of the N-occluder overlapped the cue and signal by 3.34 ms, respectively, to maximize perceptual continuity (Warren, 1999).

5. Establishment of stimulus levels

As previously described, the level of the signal was derived from the listener's 90% EDF (typically ranging between 36 and 39 dB SPL). The level of the FM-sweep cue was set to be 10 dB higher. The level of the N-occluder was held constant for each listener during a specified experiment. This level was set to be 15 dB higher than the level of the

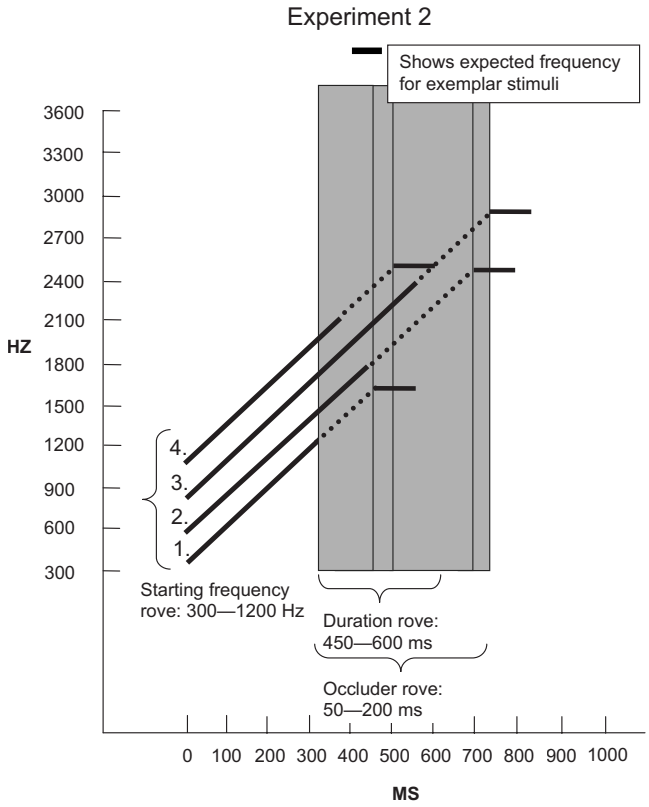


FIG. 3. Stimulus design for Experiment 2. Solid lines represent cartoon exemplar stimuli used in Experiment 2. Stimuli 1–4 demonstrate a rove on start frequency, initial-sweep duration, and the duration of the N-occluder.

signal for the median expected frequency in the range of frequency roves (approximately 5 dB higher than the FM-sweep). In Experiment 1, the lowest potential expected frequency was for a starting frequency of 300 Hz and a sweep duration of 300 ms, and the highest potential expected frequency was for a starting frequency of 1200 Hz and a sweep duration of 600 ms. Thus, the minimum and maximum expected frequencies were 1745 and 3665, respectively, and the median expected frequency was 2705 Hz. Therefore, the level of the N-occluder was 15 dB higher than the 90% threshold level for a pure-tone of 2705 Hz for each listener. Listeners' EDFs were computed (due to the use of the continuous masker) to account for a well-known frequency dependence of signal-to-noise ratios under conditions of simultaneous masking. In contrast, while some degree of forward masking was anticipated, the relative effects of frequency-dependent forward masking across stimuli were expected to be small.

After establishing the relative levels of the cue, N-occluder, and signal for each stimulus, RMS averages of the stimuli were computed before presentation. The 90% threshold levels used to establish the relative stimulus levels were derived from a listener's detection of a pure-tone signal in background noise. Clearly, the N-occluder introduced some amount of forward masking to the detection of the signal so that listeners' performance was not expected to be 90%. In testing, it was acceptable if listeners' overall performance in the experimental blocks was near to 80%. If the listener consistently performed well below 80%, the overall

stimulus level was increased by 1 dB. This increase meant that all ratios of the FM-sweep, N-occluder, and signal were maintained, but the entire rms averaged stimulus was increased.

6. General apparatus

All stimuli were generated digitally with a sampling rate of 50 kHz. Stimulus presentation was through a locally made 16 bit D/A converter, Crown D-75 amplifier, and Stax (SR5) electrostatic headphones. The spectrum of the headphones was flat across the frequency range used in the described experiments. The background noise was produced through a locally made random-noise generator. Its spectrum was bandpass filtered through a Kemo Dual Variable Filter Type VBF8.

B. Analysis methods

1. Establishment of comparison parameters

To establish a quantitative measure for estimating the location and width of the frequency region where listeners were optimally sensitive, all data from Experiments 1 and 2 were fitted with a form of rounded-exponential function. Previously, Dai *et al.* (1991) suggested that changes in listeners' sensitivity to signal frequency locations during a probe-signal task could be represented as changes in the relative attenuation of the probe frequencies to the expected frequencies. This assumption has been used and extended to interpret shifts in detectability in terms of a listener's attentional filter or measured "listening band" (Schlauch and Hafter, 1991; Hafter *et al.*, 1993; Moore *et al.*, 1996). Relative change in the shape of the best-fitting function to performance data collected through the probe-signal method was converted to decibels and approximated with the following equation:

$$W(x) = A\left(1 + \frac{p(x-g)}{g}\right)e^{p(x-g)/g}, \tag{1}$$

where W was the transfer function of the listening band, p was the bandwidth, g was the relative center of the band, and A was the non-normalized performance gain of the listener. Historically, studies of the modified probe-signal method have had reason to assume that the listening band would be centered on a single specified frequency and assumed that maximum sensitivity would be to signals at, and closely surrounding, the cued frequency. However, in the current studies, no assumptions were made regarding the predicted center of the listening band, and, instead, the estimated center frequency of the band was derived from the measured probe-signal data. In the visual system, an underestimation in the predicted path of an object's motion has been shown following increasingly longer periods of extrapolation (Peterken *et al.*, 1991). If a similar underestimation exists in prediction of a sound's frequency trajectory across time, the location of the estimated center frequency cannot be assumed. Thus, the parameter g in the symmetric exponential function was treated as a free parameter. A third free parameter was included in the model to allow a measure of the relative gain of the system (performance level) across conditions. Therefore,

TABLE I. The range and resolution used to estimate the values of parameters p, g, and A in the least-squares optimization that was performed are shown in Table I. An initial coarse search spanning a broader range was computed to determine the preliminary location of the local optimum. A second cost function using the values in Table I was then computed at a high sampling rate to span the location of the preliminary estimation of the local optimum.

	Lower bound	Upper bound	Sampling unit
p	3	15	0.1
g	0.8	1.2	0.01
A	1	5	0.01

Eq. (1) was used in the present analyses to fit a least-squares approximation to a symmetrical rounded-exponential function in the following manner:

For each listener, values of d' (Green and Swets, 1966) measured in the single-interval probe-signal task were converted to decibels using the listener's psychometric functions for detecting a pure-tone in background noise. The psychometric functions were approximated from data collected during a two-interval-forced-choice (2IFC) task (described in Appendix A). In order to compare the d' values from the single-interval task to the corresponding decibel values of the psychometric function, all scores recorded in the probe-signal conditions were converted to percent correct using the following equations. Equation (2) converts the d' value obtained in the single-interval task used in the probe-signal studies to a comparable d' value for a 2IFC task (Wickens, 2002, p. 104).

$$d'_{2\mathrm{IFC}} = d'\sqrt{2}. \tag{2}$$

Equation (3) converts the rescaled d' value into a comparable percent-correct score (Wickens, 2002, p. 97)

$$P_C = \Phi(d'_{2\mathrm{IFC}}/\sqrt{2}), \tag{3}$$

where Φ represents the cumulative distribution function of a Gaussian normal $N(0,1)$.

Once percent-correct scores for detection were obtained, listeners' psychometric functions were used to convert from percent correct into decibel values relative to a listener's 80% detection level. This was done for each value of d' for each listener prior to any averaging. Individual scores in decibels were then averaged across all listeners.

For each condition, the averaged data were fitted with Eq. (1) to estimate the three free parameters: p, g, and A, and a least-squares optimization was performed to compute a cost function on each parameter for values within the following ranges and resolutions (Table I).

The rounded-exponential function was used only as a rough estimate of the listening band that enabled measurement of changes in center frequency and width of the listening region across conditions. Further, introduction of roving parameters in the current studies was expected to increase the variance in listeners' estimation (Drennan and Watson, 2001). This inclusion was absolutely necessary to exclude listeners' potential use of cues besides velocity to optimally detect the location of the signal; however, an increase in the

variance of a listener's listening region was expected that, in turn, would have produced an overestimation of the parameter p.

2. Significance testing for individual parameters

A Monte Carlo randomization test (Manley, 1997) was performed to establish a measure of relative change between testing conditions and to test the null hypothesis that observed condition differences in the estimated parameters: p, g, and A were due to the sampling error rather than true differences in the underlying populations. Details of the randomization analysis used are described in Appendix B.

III. EXPERIMENT 1

Experiment 1 measured listeners' ability to use velocity information implied by a constant-velocity FM-sweep to extrapolate the frequency of a signal following an interrupting N-occluder. In Experiment 1, both the starting frequency and the duration of the FM-sweep were roved (as described in Sec. II). The duration of the N-occluder was always a constant 125 ms.

A. Subjects

Four normal-hearing listeners participated in the experiment. All were undergraduate university students between the ages of 18 and 24 who received monetary compensation for their listening time.

B. Results

Following an N-occluder, listeners show increased sensitivity to frequencies that fall along an implied path of constant-linear velocity. Further, listeners more accurately use velocity information to extrapolate the trajectory, showing decreased variance and bias in the width and position of the listening band, when presented with longer duration FM cues (corresponding to longer samples of constant velocity).

Figure 4(a) shows averaged d' values of four listeners for two ranges of cue duration. Figure 4(b) shows least-squares approximations to the averaged data of the four listeners. Each listener's d' values are transformed into the corresponding decibel values through their psychometric functions for detecting a pure-tone in noise. The estimated center (g) for the longer durations (450–600 ms, $g=0.97$) is significantly closer to the expected frequency value (1) than the estimated center (g) for the shorter durations (300–450 ms, $g=0.93$), $p<0.001$. The width of the estimated function is significantly narrower for the longer duration sweep range than for the shorter duration sweep range, $p=9.5$ and $p=6.8$, respectively, $p<0.05$ (where a higher value of p corresponds to a narrower estimated band). A narrower width corresponds to decreased variance and increased accuracy in use of velocity information to predict the frequency location of the signal. Additionally, listeners show a shift toward a more conservative criterion value in conditions with longer initial-sweep durations [$\lambda=1.1119$ (300–450 ms) and $\lambda=0.8455$ (450–600 ms)].

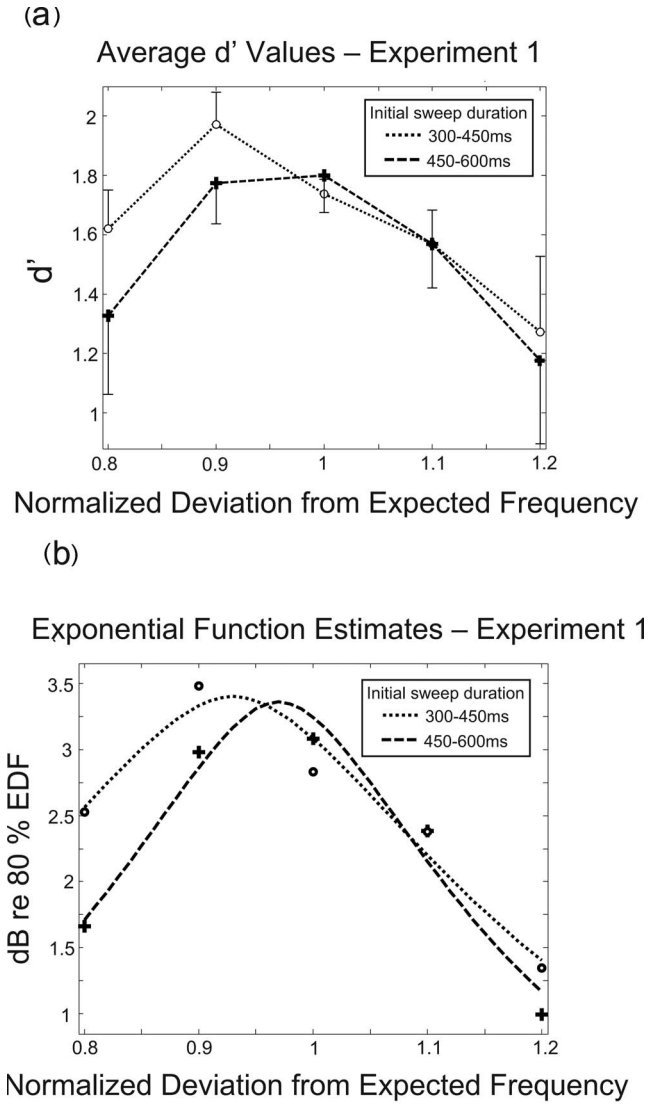


FIG. 4. (a) d' values for averaged data of four listeners as a function of normalized deviation from the expected frequency. The dotted line and dashed lines represent data for initial-sweep durations ranging between 300–450 ms and 450–600 ms, respectively. (b) Fits to a rounded-exponential function for the same initial-sweep durations. Actual data used to generate the fit are indicated for the 300–450 ms condition by open circles, and for the 450–600 ms condition by crosses. All dB levels are across listener averages taken from each listener's 80% EDF as determined by their psychometric function, and all between subject error is accounted for in the statistical analyses that were performed.

IV. EXPERIMENT 2

Results of Experiment 1 show that velocity information can be used to alter sensitivity along a projected path. Accuracy in using this information increased as the duration of the cue increased. However, it is unclear whether accuracy in use of velocity information is dependent on the duration across which a prediction is made. As discussed earlier, it is possible that listeners' use of velocity information may not produce a linear extrapolation of frequency across an indefinite time interval (previously observed in studies of visual and auditory temporal predictions, Peterken *et al.*, 1991; Szelag *et al.*, 2002). Experiment 2 measured listeners' ability to use velocity information across increasingly longer periods of extrapolation. In addition to roving the initial start frequency

and FM-sweep duration (450–600 ms), the duration of the N-occluder also was roved (50–200 ms) per trial to introduce variability in the period listeners were required to extrapolate. (Exemplary conditions are shown in Fig. 3.) Normalized center frequency, bandwidth, and non-normalized performance gain of the measured listening band were assessed to determine dependence of the listening region on the duration of extrapolation.

A. Subjects

Three normal-hearing listeners participated in the experiment. All three participated in Experiment 1.

B. Results

Across increased occluder duration (and corresponding periods of extrapolation), listeners continue to use velocity information to predict the frequency of a sound following the occluder. However, listeners show a downward shift in the location of the listening region relative to a trajectory of constant-linear velocity as the duration of the occluder increases. In contrast to results of Experiment 1, this shift is not coupled with an increase in the width of the listening band suggesting that the effective use of velocity information does not change but, instead, points listeners to a less accurate position.

Figure 5(a) shows averaged d' values of three listeners. Figure 5(b) shows least-squares approximations to the averaged data for each range. The estimated values for each parameter for each range of occluder duration are shown in Table II. The estimated-center-frequency of the listening band decreases as the duration of the occluder increases, $g=1.01$, $g=0.95$, and $g=0.92$ for the 50–100 ms (S), 100–150 ms (M), and 150–200 ms (L) ranges, respectively. A comparison between the S and M conditions is significant with $p<0.01$, whereas a comparison between the M and L conditions is not ($p=0.08$). Nonetheless, a downward trend in the estimated-center-frequency with increased occluder duration is evident.

The estimated bandwidths of the optimal fits for all three ranges of occlusion are similar. The change in bandwidth is not significant for the S and M conditions, $p=(7.2)$ and $p=(6.4)$, respectively, $p>0.05$. However, the change in bandwidth for the M and L conditions is marginally significant ($p=8.7$ for L), $p<0.05$. It is important to note that the measured listening band in the L condition is not significantly wider than either the M or S conditions (larger values of p corresponding to narrower estimated bandwidths).

Differences in measured detectability across the three (S, M, and L) conditions are marked. Performance is significantly higher for conditions containing longer occluder durations [L in comparison to M, $A_L=(4.31)$ and $A_M=(3.42)$, $p<0.01$; M in comparison to S, $A_S=(2.84)$, $p<0.01$]. This result is addressed in Sec. III.

V. EXPERIMENT 3

Experiments 1 and 2 measured changes in detection of pure-tone signals falling on or within 20% of a frequency implied by a trajectory of constant-linear velocity. The results of both experiments suggested that listeners were able to use this information to alter sensitivity along a predicted path. Further, results of Experiment 2 showed a downward shift in the estimated-center-frequency as a function of N-occluder length. This presented the question of whether listeners limited their listening range by avoiding listening

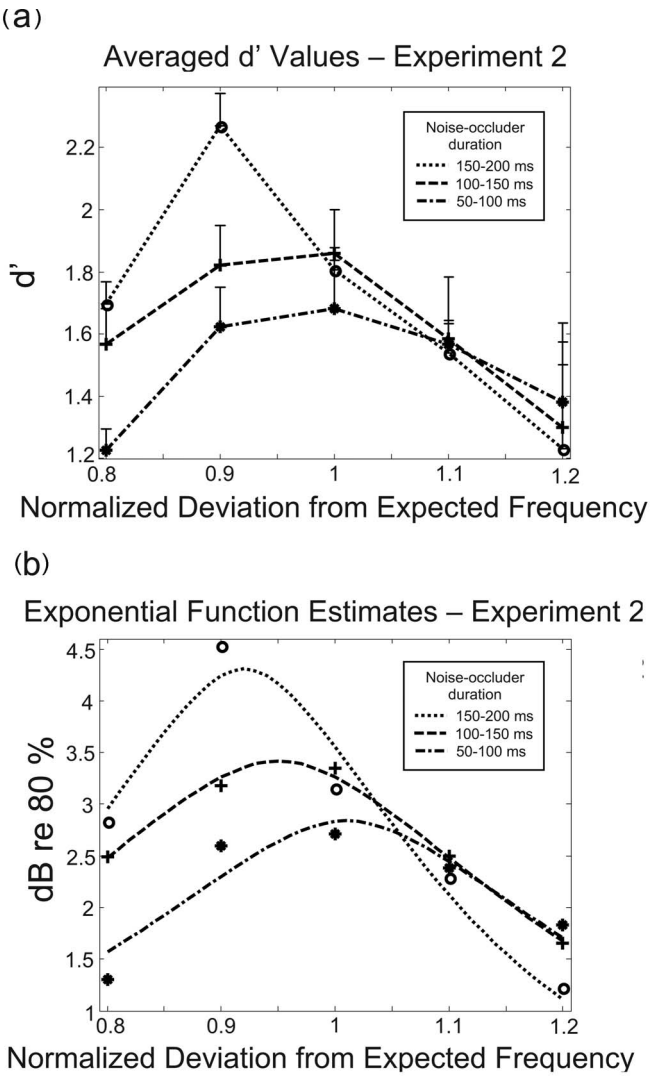


FIG. 5. (a) Averaged d' values for the data of three listeners as a function of normalized deviation from the expected frequency. The dash-dot, dashed, and dotted lines represent data for N-occluder durations ranging between 50–100 ms, 100–150 ms, and 150–200 ms, respectively. (b) Fits to a rounded-exponential function for the same N-occluder durations. Actual data are indicated for the 150–200 ms range by open circles, for the 100–150 ms condition by crosses, and for the 50–100 ms range by asterisks. All dB levels are across listener averages taken from each listener's 80% EDF as determined by their psychometric function, and all between subject error is accounted for in the statistical analyses that were performed.

TABLE II. Estimated values of parameters p, g, and A for Experiment 2 as determined using the rounded exponential function defined by Eq. (1).

	S 50–100 ms	M 100–150 ms	L 150–200 ms
p	7.2	6.4	8.7
g	1.01	0.95	0.92
A	2.8	3.4	4.3

for frequencies near the upper edge of the testing range or truly experienced a perceptual downward shift as the duration of extrapolation increased. To address this, in Experiment 3, listeners were asked to judge whether suprathreshold tones were above or below the frequency implied by the linear trajectory. This study allowed for both the effects of extrapolation and interpolation and enabled comparison between estimates of the center of the listening bands found through the detection studies of Experiments 1 and 2 to listeners' perceptual experience of a constant trajectory across a given occluder duration. If the previously observed downward shift in relative frequency was also observed in a subjective, suprathreshold testing condition, it would imply that the downward shift seen in Experiment 2 was not a product of listeners' listening strategies biasing them to listen away from the upper edge of the range. Rather, it would suggest that the observed shift in performance indicated a change in listeners' sensitivity for certain frequencies along a trajectory that, in turn, affected the conditions under which listeners' perceived the trajectory to be continuous.

A. Subjects

Three normal-hearing listeners participated in the experiment. All three listeners participated in Experiment 1. Two of the three listeners participated in Experiment 2.

B. Methods

1. Stimulus design

All stimulus design was identical to the descriptions in Sec. II except for the following changes. To reduce any possibility that the observed downward shift was specific to the previously tested velocity, the velocity of the initial FM-sweep was changed to 3400 Hz/s. The duration of all initial sweeps was a constant 500 ms, and the level of the pure-tone signal following the N-occluder was set to be equal to that of the initial FM-sweep. The frequency of the pure-tone signal was set to occur at the following percentages above and below the expected frequency: −25, −20, −15, −10, −5, +5, +10, +15, +20, +25. Unlike Experiments 1 and 2 which used a distribution of signal frequencies weighted toward the expected frequency, the distribution used in Experiment 3 was uniform around the expected frequency. No signals were included at the expected frequency. The starting frequency of the FM-sweep was roved, as described in Experiment 2. Three durations of N-occluder were presented randomly within a block: 50, 125, and 200 ms.

2. Procedure

In a single-interval task, listeners heard an FM-sweep followed by an N-occluder and a pure-tone. All were told that the pure-tone would never fall directly on the trajectory, but, rather, would be slightly above or below it. On each trial, listeners were required to assign a label of high or low to the tone following the occluder, even if the tone appeared to fall along the trajectory. Listeners did not receive feedback informing them of the correctness of their response. Prior to beginning testing, listeners performed training sessions using the most extreme percentages above and below the expected frequency. Training continued until successful completion of a block of 20 trials with 100% correct identification.

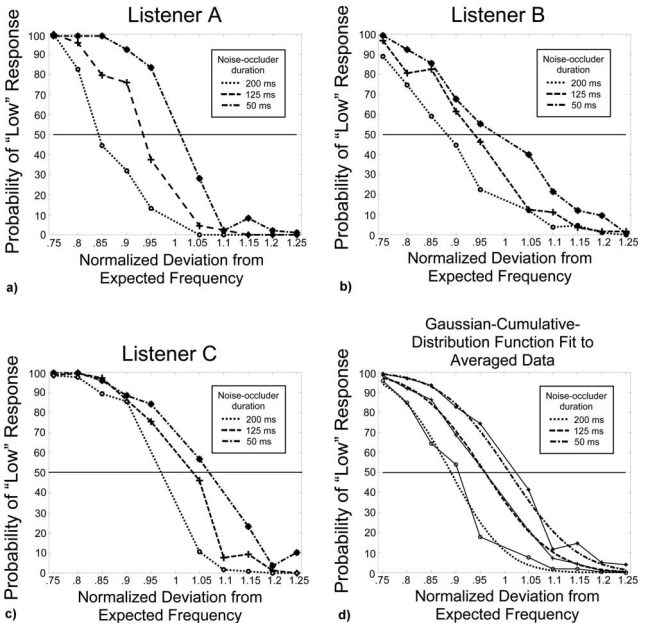


FIG. 6. [(a)–(c)] For individual listeners, the probability of responding that a tone was below a continuous trajectory as a function of normalized deviation from the expected frequency. The dotted, dashed, and dash-dot curves plot data for the 200 ms, 125 ms, and 50 ms N-occluder durations, respectively. (d) Least-squares estimates of Gaussian-cumulative-distribution functions for the average of the data of listeners A, B, and C. The smooth curves indicate least-squares fits to the averaged data. Solid thin lines show actual averaged data to which the functions were fitted.

C. Results

As the length of the occluder increases, listeners consistently perceive lower frequencies relative to a trajectory of constant-linear velocity as falling along a path of constant rate of frequency change. Figures 6(a)–6(c) plot individual responses for three listeners. Each curve demonstrates the probability a listener responded that the tone was lower than a trajectory of constant velocity as a function of normalized frequency. Figure 6(d) shows least-squares estimates of a psychometric function fit to the averaged data of the three listeners. A Gaussian-cumulative-distribution function was used to compute the estimates. The solid thin lines overlaid on each curve represent the data used to derive the optimal fit.

A downward shift in the point-of-subjective-equality, at which listeners respond "low" 50% of the time, is consistent across all listeners for increasing N-occluder durations. The crosses in Fig. 7 plot, for the averaged data, the estimated normalized frequency at which listeners respond low 50% of the time as a function of N-occluder duration. This point-of-subjective-equality is compared to the solid circles in Fig. 7 which represent the estimated-center-frequencies observed in Experiment 2. (Additionally, while listener C shows a clear upward bias [50 ms curve shown in Fig. 6(c)] with respect to listeners A and B, a downward trend with increased occluder duration is evident.) The similarity between the points-of-subjective-equality measured in Experiment 3 and the estimated-center-frequencies measured in Experiment 2 (Fig.

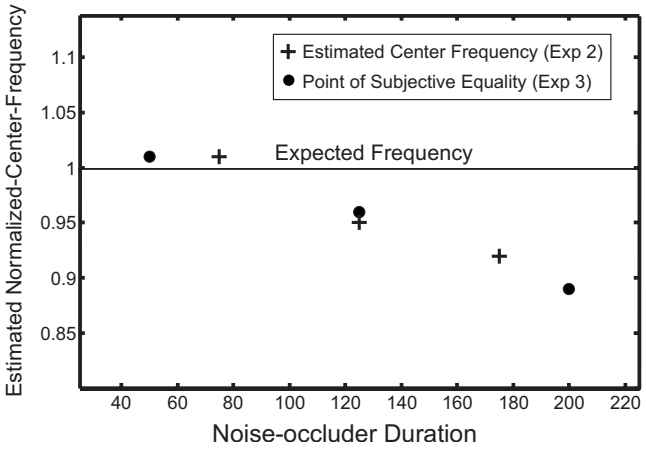


FIG. 7. Closed circles indicate points at which listeners respond 50% of the time that the signal tone falls below a perceptually continuous trajectory for levels of increasingly longer durations of N-occluder duration. Data are extracted from the estimated Gaussian-cumulative-distribution functions fit to the average data of three subjects, as shown in Fig. 6(d). The point-of-subjective-equality shifts downward from the expected frequency (defined as a trajectory of constant-linear increase in frequency equal to a normalized value of 1 and indicated by the solid line) as the duration of the N-occluder increases. Crosses indicate the frequency regions of maximal sensitivity (estimated-center-frequencies) as a function of the mean N-occluder duration for the ranges measured in Experiment 2 (i.e., 75 ms for an occluder range of 50–100 ms).

7) suggests that changes in listeners' sensitivity to frequencies along an implied path may be directly reflected in their perceptual experience of a continuous trajectory.

VI. GENERAL DISCUSSION

The present experiments demonstrate listeners' use of velocity information to increase sensitivity along an extrapolated path. Consistent with studies of visual perception (Peterken *et al.*, 1991), listeners show increased accuracy in the use of velocity information with extended exposure to a constant-velocity cue (Experiment 1). That is, as the duration of constant-velocity information increases, the center frequency of the listening band shifts closer to the expected frequency and the width of the listening band narrows.

Additionally, the accuracy of listeners' use of velocity information to extrapolate the path of the changing sound is dependent on the duration of extrapolation (Experiment 2). During conditions of shorter occlusion, listeners' extrapolated estimates closely match the expected frequency (as predicted by extrapolation of a trajectory of constant-linear-velocity). In comparison, during conditions of longer occlusion, listeners' normalized estimates of the center frequency of the listening region shift downward relative to the expected frequency. This downward shift across longer occluder durations is not coupled with an increase in the bandwidth of the listening region indicating that listeners are not simply listening across a larger frequency region as the duration, and corresponding range of frequency extrapolation (Δf), increases. Rather, it indicates that listeners are, instead, listening across a comparable frequency region centered at a lower frequency than the expected frequency would predict.

Finally, results of Experiment 3 show that the downward shift in estimated center frequency corresponds to listeners' perception of the frequency at which they optimally experience a continuous trajectory rather than the result of biasing listening strategies to optimize performance. While previous studies have seen a downward shift (Kluender and Jenison, 1992; Pollack, 1977), results of Experiments 1 and 2 are the first to show that a downward shift in the perception of an optimally aligned trajectory does not depend on what follows the interruption, but, rather, arises from changes in threshold sensitivity of the system during the period of extrapolation. This change leads listeners to expect a sound to reappear at a lower frequency than extrapolation of constant-linear-velocity would predict, and it may represent a direct influence of changes in the sensory system on the perceptual representation of dynamic sounds across an interruption.

A. Discussion: Effects of noise-occluder duration on performance gain (Experiment 2)

Although not directly relevant to the primary conclusions and discussion of the present studies, the counterintuitive increase in performance for longer occluder durations seen in Experiment 2 [Fig. 5(b)] warrants discussion. While a direct answer for this difference in performance is not obvious, one explanation may be the result of effects of forward masking from the initial FM-sweep on information contained within a theoretical temporal integration window with respect to the signal. All signals were 100 ms long. The effects of forward masking due to the N-occluder are expected to increase as the duration of the occluder increases (Kidd and Feth, 1982). This did not happen, and, instead, listeners' performance improved in conditions with longer occluder durations, suggesting a property of the stimulus other than the noise occluder was influencing the loudness of the signal. Rather, it may be the case that signal detection in Experiment 2 is affected by forward masking introduced by the ending portion of the initial FM-sweep falling within the integrated window of the signal. In studies of the effects of the duration of a forward masker on listeners' increment detection, results of Schlauch *et al.* (1997) were also inconsistent with previous studies of forward masking. They found that shorter duration maskers produced more masking than longer duration maskers. Further, their results, as well as those seen by Oxenham and Plack (2000), suggest that the degree of masking is dependent on the similarity of the masker to the signal, with greater similarity producing conditions of greater masking. In the current experiments, trials with shorter occluder durations created conditions where both the N-occluder as well as the ending portion of the initial FM-sweep would likely be integrated within the same temporal window as the signal. This led the ending portion of the initial sweep to contribute to the overall masking of the signal in some conditions and not in others. For example, a trial containing the shortest N-occluder duration (50 ms) and beginning at the highest initial-sweep start frequency (1200 Hz) has an expected frequency of 3150 Hz. The estimated equal-rectangular-bandwidth [ERB, as computed using Eq. (4)] for this frequency is 365 Hz.

$$\mathrm{ERB}_N = 24.7(4.37F + 1). \quad (4)$$

Therefore, for a condition with an FM-sweep rate of 3000 Hz/s, 72 ms of the initial FM-sweep, traversing a 215 Hz region, would fall within the ERB of the signal frequency and likely contribute to the effective loudness of the signal. This is in addition to any forward masking effects introduced by the 50-ms N-occluder. While for all shorter duration occluders it is true that some portion of the initial FM-sweep falls within the ERB of the expected frequency (and is assumed to contribute to its effective loudness), it would not be true for signals following longer duration occluders. Therefore, the amount the initial FM-sweep contributes to the effective loudness of any signal is a function of both the initial start frequency and the occluder duration, with the shortest occluder durations most affected—having both the longest portion of initial FM-sweep included in their estimated temporal window and the frequency content of the initial FM-sweep most near (and potentially most similar) to the expected signal frequency. Assuming a temporal integration window of approximately 200 ms, both the S and M conditions would have integrated effects of the end of the initial FM-sweep on the signal (with the greatest effects present in the S condition). In contrast, the entire signal duration in L conditions would be free of the influence of the initial FM-sweep, leaving only the effects of the N-occluder to alter the effective loudness of the signal. While this is only a proposed explanation for the change in performance as a function of N-occluder duration, it is an intrinsic property of the stimulus set that may have affected overall performance (A). However, all three parameters, p, g, and A, are estimated independently and the estimates of p and g would not be influenced by the observed change in performance.

B. Proposed explanation for downward shift in estimated center frequency

A simple process is offered to explain the observed downward shift in the estimated-center-frequency as a function of the N-occluder duration. A continuous representation of velocity information would require computation of the instantaneous time derivative of the frequency (df/dt) across the duration of the changing sound. Results of Experiment 1 demonstrate that use of velocity information continues to increase in accuracy (narrowing of the listening band and a decrease in error of the estimated-center-frequency away from the expected frequency) beyond a duration of 300 ms. This suggests that the instantaneous velocity estimate is a function of past velocities where information is integrated or averaged across a longer temporal window.

Figure 8 shows the estimated-center-frequency as a function of time for a trajectory of constant-linear velocity. Plotted with this is the predicted-center-frequency following application of a finite impulse response (FIR) exponentially-weighted moving average (windowed at 365 ms) to the instantaneous velocity. A 365-ms window was applied as an estimate of the window width across which velocity information is averaged based on the results of Experiment 1 and optimal prediction of the results of Experiment 2. (Further analysis of data from Experiment 1 using initial-sweep durations ranging between 375 and 525 ms showed minimal change in convergence in comparison to 450–600 ms conditions, suggesting that the window width is approximately between 350 and 400 ms.) The FIR filter is defined by

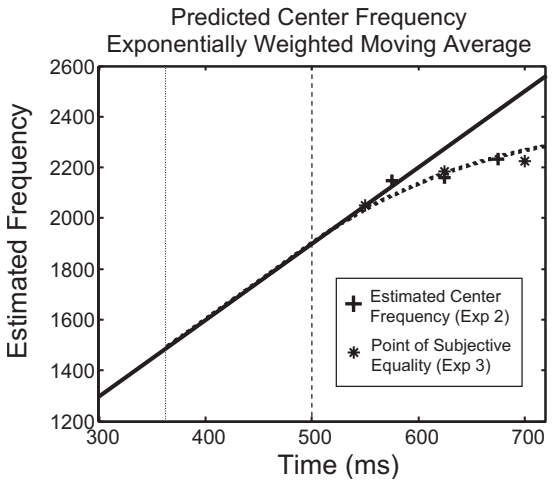


FIG. 8. The solid line indicates the expected center frequency for exact linear extrapolation of velocity information of a 500-ms FM-sweep with a start frequency of 400 Hz and a velocity of 3000 Hz/s through a broad-band N-occluder. The dotted line represents the predicted-center-frequency with an FIR exponentially-weighted moving average with a window width of 365 ms applied to the instantaneous values of df/dt. The dotted vertical line indicates the onset of predicted values, and the dashed vertical line indicates the onset of the N-occluder. Crosses indicate estimated-center-frequencies obtained in Experiment 2, and asterisks indicate points-of-subjective-equality obtained in Experiment 3.

$$\begin{aligned} &0, \quad t < 0, \\ &ce^{-0.35t}, \quad 0 < t < 365 \text{ ms}, \\ &0, \quad t > 365 \text{ ms}, \end{aligned} \quad (5)$$

where c is a constant of normalization. It has been applied under the assumption that the instantaneous velocity becomes zero at the onset of the N-occluder. While it is clear that the instantaneous df/dt for a broad-band noise is not zero, when averaged across a temporal window the average df/dt quickly converges to zero. Crosses shown in Fig. 8 indicate estimated-center-frequencies obtained in Experiment 2, and asterisks indicate points-of-subjective-equality obtained in Experiment 3.

The process used to produce the estimates shown in Fig. 8 is a simple representation of how the auditory system may be transforming information. Additional studies manipulating the local time derivative of the frequency around the onset of the N-occluder as well as further manipulations of the initial-sweep duration in order to observe a more accurate value of the proposed window size and weighting of information across the duration of the window are needed to enable a more accurate model of this process. Nonetheless, it is striking how closely the low-pass filtered estimates are to the measured data.

It is also necessary to consider that the auditory system may represent velocity logarithmically or linearly in units of ERBs (Bark scale). If this is true, an upward linear sweep would be interpreted as decelerating. A representation of a

decelerating signal would predict a lower estimated-center-frequency across longer duration noise occlusions. However, previous studies considering the optimal representation of a signal of changing frequency suggest that this is not the case. Results show that a logarithmic or Bark scale representation of frequency change is not a better predictor than a linear scale of the optimal frequency that continues a trajectory of constant change (Pollack, 1977; Kluender and Jenison, 1992). Rather, downward shifts in the optimal frequency were observed for signals changing at constant rates within all three scales (linear, logarithmic, and Bark). Pollack found a similar ratio for the downward shift in the optimal frequency relative to the estimated frequency for both linear and logarithmic trajectories. This would represent a considerably greater underestimation in absolute frequency for the logarithmic stimulus relative to the linear stimulus. The consistent shift across all scales implicates a more general process leading to the downward shift for increasing trajectories and would be predicted by the weighted moving average proposed above. Such a process would predict an underestimation of the expected frequency across longer periods of prediction regardless of the scale of transformation. In addition, while all frequency trajectories used in the present studies were increasing across time, the proposed model also predicts an underestimation in prediction for frequency trajectories decreasing across time. In such conditions, listeners would show an increase rather than a decrease in the measured optimal frequency of continuation. Data from Pollack (1977) using series of pulsed tones support this prediction.

Further, it is important to consider that the well-known logarithmic representation of frequency in the auditory system (Moore, 1997) does not require that velocity also be represented logarithmically. Carre (2004) demonstrated that first formant transitions of natural speech signals are more or less linear in an absolute-frequency scale rather than a log-frequency scale. The presence of roughly linear acoustic changes in prominent natural sounds such as speech implies an important role for linear representations of frequency change in the auditory pathway.

C. Prediction of frequency change in the absence of noise

Results from previous studies (Kluender and Jenison, 1992; Pollack, 1977) suggest that the observed downward shift in center frequency with increased duration of the N-occluder may occur in prediction regardless of whether masking energy is present during the period of interruption. Therefore, it is conceivable that the observed changes in sensitivity for tones falling along the projected path, as shown in Experiments 1 and 2, would still arise if an intervening silence replaced the occluding noise. If, within the auditory system, df/dt is averaged or integrated across a time window greater than 300 ms, one would predict similar trends in the estimated-center-frequency regardless of whether a silence or noise interval is present. However, it is difficult to ignore the marked difference in the perception of a sound with an interrupting silence in comparison to an interrupting N-occluder. One possibility is that prediction does not require stimulation, but when sufficient masking energy is present, sensitivity to frequencies falling along the predicted path is elevated. A conceivable scenario could be that global stimulation in the auditory periphery introduced by the N-occluder is unevenly represented at higher auditory areas such that frequencies falling along the projected path are amplified relative to the background, in turn, enhancing and shaping the perception of continuity.

D. Neural correlates of velocity coding and temporal induction

Physiological studies of the visual system have observed cells in the primate parietal cortex that show greater response to inferred motion during temporary occlusion of a moving stimulus (Assad and Maunsell, 1995). In comparison, it is currently unclear whether auditory continuity for dynamically changing sounds is similarly encoded in the auditory cortex at the level of a single neuron. Studies of continuity (using a constant-frequency tone) in the primary auditory cortex (AI) have found cells showing similar response patterns to both a truly continuous sound and a perceptually continuous sound (Petkov *et al.*, 2007; Sugita, 1997). While these studies have independently demonstrated sensitivity to conditions of continuity and discontinuity, the neural correlates underlying the robust perception of continuity for a dynamically changing stimulus may be quite different. In fact, encoding of velocity for predictive use as suggested by the current studies may lead to a very different change at the level of a single cell independent of additional processing necessary for the experience of perceptual continuity. For example, study of the encoding of spatial velocity in the optic tectum of the barn owl finds, for cells responsive to interaural timing differences, context-dependent receptive field shifts suggesting predictive anticipation of future stimulus locations that are dependent on the speed and direction of a moving source (Witten *et al.*, 2006). It is probable that cells sensitive to rate-of-change in frequency space (presently referred to as velocity) may show a similar predictive change in receptive field at some level of the auditory system. Continued convergence and increased predictive accuracy of velocity information beyond a duration of 300 ms (as demonstrated in Experiment 1) imply that physiological circuits involved in this process would likely be located high in the auditory pathway (i.e., beyond AI).

E. Summary

The present experiments demonstrate that listeners do use velocity information to track the change of a dynamic sound, and further that increased sensitivity along an extrapolated path significantly influences listeners' predicted and perceived trajectories of constant change through an occluder. These findings suggest that use of velocity information in the extrapolation of the path of a changing sound may be a fundamental aspect of auditory processing in the presence of spectrally and temporally dynamic sounds such as speech and music. In particular, these findings offer strong implications for the underlying physiology that enables processing of both speech and music in a noisy environment. At

this time, further studies of both the psychophysical and physiological correlates of the phenomenon are warranted.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Thomas Wickens for discussion and guidance in appropriate quantitative techniques, and Dr. Frederic Theunissen, Dr. David Wessel, Dr. Pierre Divenyi, Dr. Anne-Marie Bonnel, and anonymous reviewers for thoughtful opinions and editorial comments of the present research. P.A.C.C. is now employed by the Johns Hopkins School of Medicine.

APPENDIX A: ESTABLISHING EQUAL DETECTABILITY

During the typical single-interval task used in testing, levels of all signals were set to yield equal percentages of correct detection as obtained from a 2IFC task for detecting a pure-tone in background noise. In order to equate the signal tones in detectability prior to probe-signal sessions, listeners' thresholds were measured for detecting a pure-tone of varying frequency in a low-level background noise as previously described. A 2IFC, three-down one-up, tracking method (Levitt, 1971) was used to establish listeners' 79.4% threshold for detecting a 100-ms signal of a specified frequency. Each listener completed five blocks of the tracking 2IFC task at 400, 800, 1000, 1500, 2000, 2500, 3000, and 4000 Hz.

For each subject at each frequency, the average level of the five blocks was considered their base line testing level. Following tracking sessions, listeners completed three to five blocks of 100 trials in a 2IFC detection task at each of three different levels. The levels for each listener were as follows: base line (as obtained from the average of their tracking block), +1.5 or +2 dB from their base line, and −1.5 or −2 dB from their base line. Whether they completed blocks at ±1.5 or ±2 dB depended on their initial performance at ±2 dB. Desired performance was near 90% at +2 dB. If performance at this level was generally above 95%, the level was lowered by 0.5 dB in order to measure points where the slope of the listener's psychometric function was expected to be steeper. Similarly, the level was raised following performance below 65% at −2 dB. This set was completed for all eight frequencies.

Performance at these levels was then used to fit a least-squares optimization of a psychometric function (using a Gaussian-cumulative-distribution function) to a listener's performance as a function of stimulus intensity for each frequency. The slope of the psychometric function for each listener was assumed to be constant across frequency. All fits were made with a single free parameter for each listener's set, the decibel level of the function's mean (relative offset in decibels). Following this, a least-squares regression line was approximated for each listener as a function of decibel level versus linear frequency (Green *et al.*, 1959). This approximation will be referred to as the listener's EDF. Throughout all experiments, for each trial, the signal level of the expected frequency was imputed from the point on the regression line of the listener's 90% EDF corresponding to the specified testing frequency.

APPENDIX B: RANDOMIZATION TEST FOR SIGNIFICANCE

The following Monte Carlo randomization test was performed to establish a measure of relative change between testing conditions and to test the null hypothesis that observed condition differences in the estimated parameters: p, g, and A were due to sampling error rather than true differences in the underlying populations.

For each comparison, all raw data files for each listener were combined on a trial-by-trial basis across the assessed conditions and stratified by percentage away from the expected frequency. In this way, all data collected at a fixed percentage of the expected frequency were combined only with data from the same percentage. During a single iteration, rows of data from the combined file for each listener were randomly permuted and a sample equal to one-half the size of the combined conditions was drawn without replacement from the combined file. A d' value was calculated for each stratified percentage level relative to the expected frequency: −20%, −10%, 0%, +10%, and +20% and will be referred to as a d' set. This process was repeated 1000 times to create a matrix of size 1000 by 5 containing 1000 sets of d' values computed by randomly drawing samples from the combined file.

The new sets of d' values were used to estimate the exponential function [Eq. (1)] in the following manner. On a single iteration, a single set of d' values was drawn from a permutation of the listener's set of d' values. This was done for each listener without replacement across the series of iterations. For each listener, data were converted to decibels using an estimate of the listener's psychometric function for detecting a pure-tone in noise. The decibel values were averaged for the group of listeners at each percentage value from the expected frequency. The averaged values were used to approximate a least-squares estimate of the exponential function with three free parameters: width, center frequency, and gain. One iteration yielded a single estimate of each parameter. This process was completed 1000 times to generate a file of size 1000 by 3 containing 1000 estimates of each parameter.

Each file of parameter estimates was then randomly divided into two groups. For each parameter, an estimated difference value, referred to as $d\hat{\nu}$, was computed by randomly drawing an estimated parameter value from each group and taking the absolute value of the difference. For example, if two center frequency estimates were drawn, 0.97 and 0.985, $d\hat{\nu}$ would be 0.015. This process was repeated for each parameter 1000 times to create a file of size 1000 by 3 consisting of estimates of $d\hat{\nu}$ for each parameter. These sets were then used to establish a measure of variability in the experimental data. For each condition being compared, we computed the difference value of the uncombined (and unpermuted) measured parameter values following fitting to the exponential function (shown in Table II and now referred to as $d\bar{\nu}$). For each parameter, the number of times that $d\hat{\nu}$ exceeded $d\bar{\nu}$ was recorded. This number provided a probability measure of the chance that the observed differences in the actual uncombined, unpermuted estimates were due to the sampling error rather than representing a true difference in

TABLE III. Individual listeners estimated value of d'—Experiments 1 and 2.

	Experiment 1: Normalized deviation from expected frequency				
	0.8	0.9	1	1.1	1.2
A_{300_450}	1.5874	2.2718	1.5921	1.2846	0.5230
A_{450_600}	1.5468	1.8819	1.8041	1.1709	0.4080
B_{300_450}	1.2632	1.7456	1.7510	1.5217	1.5434
B_{450_600}	0.5743	1.3789	1.6113	1.5417	1.2782
C_{300_450}	1.8314	1.9348	1.8229	1.8314	1.4049
C_{450_600}	1.7905	1.9984	2.1500	1.8683	1.2747
D_{300_450}	1.7978	1.9346	1.7848	1.6362	1.6206
D_{450_600}	1.3987	1.8345	1.6313	1.6994	1.7475
	Experiment 2: Normalized deviation from expected frequency				
	0.8	0.9	1	1.1	1.2
A_{50_100}	1.2069	1.8347	2.0710	1.5256	1.1701
A_{100_150}	1.5135	2.0437	1.9843	1.6460	0.7799
A_{150_200}	1.5573	2.3066	1.8135	1.1228	0.4798
C_{50_100}	1.1229	1.6413	1.5221	1.4768	1.3891
C_{100_150}	1.7879	1.8193	2.0161	1.6421	1.7151
C_{150_200}	1.7671	2.4272	1.8570	1.5314	1.3536
D_{50_100}	1.2945	1.4815	1.4551	1.6832	1.5462
D_{100_150}	1.4097	1.6822	1.5655	1.4056	1.4602
D_{150_200}	1.8395	1.9900	1.7319	1.9730	1.8358

the population. For example, in Experiment 1, the $d\bar{v}$ value for center frequency as a function of initial-sweep duration was 0.04 (the difference between the actual values 0.97 and 0.93). After combining data across initial-sweep duration, we found that the probability of observing $d\hat{v}$ of 0.04 or greater when the conditions were combined was 9/1000. It may then be said that the observed difference in center frequency as a function of initial-sweep duration was significant, $p<0.01$.

APPENDIX C: INDIVIDUAL DATA: EXPERIMENTS 1 AND 2

Table III shows the data of Experiments 1 and 2.

Assad, J. A., and Maunsell, J. H. R. (**1995**). "Neuronal correlates of inferred motion in primate posterior parietal cortex," Nature (London) **373**, 518–521.

Bashford, J. A., Jr., and Warren, R. M. (**1987b**). "Multiple phonemic restorations follow the rules for auditory induction," Percept. Psychophys. **42**, 114–121.

Bregman, A. S. (**1994**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Carre, R. (**2004**). "From an acoustic tube to speech production," Speech Commun. **42**, 227–240.

Churchland, M. M., Chou, I., and Lisberger, S. G. (**2003**). "Evidence for object permanence in the smooth-pursuit eye movements of monkeys," J. Neurophysiol. **90**, 2205–2218.

Ciocca, V., and Bregman, A. S. (**1987**). "Perceived continuity of gliding and steady-state tones through interrupting noise," Percept. Psychophys. **46**, 39–48.

Dai, H., Scharf, B., and Buus, S. (**1991**). "Effective attenuation of signals in noise under focused attention," J. Acoust. Soc. Am. **89**, 2837–2842.

Dannenbring, G. L. (**1976**). "Perceived auditory continuity with alternately rising and falling frequency transitions," Can. J. Psychol. **30**, 99–114.

Drennan, W. R., and Watson, C. S. (**2001**). "Sources of variation in profile analysis. II. Component spacing, dynamic changes, and roving level," J. Acoust. Soc. Am. **110**, 2498–2504.

Ebata, M., Miyazono, H., Kumamaru, K., and Chisaki, Y. (**2001**). "The formation of attentional filters for a missing-fundamental complex tone and frequency-gliding tones," Acoust. Sci. & Tech. **22**, 401–406.

Green, D. M., Birdsall, T. G., and Tanner, W. P. (**1957**). "Signal detection as a function of signal intensity and duration," J. Acoust. Soc. Am. **29**, 523–531.

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. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hafter, E. R., Schlauch, R. S., and Tang, J. (**1993**). "Attending to auditory filters that were not stimulated directly," J. Acoust. Soc. Am. **94**, 743–747.

Houtgast, T. (**1972**). "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am. **51**, 1885–1894.

Husain, F. T., Lozito, T. P., Ulloa, A., and Horwitz, B. (**2005**). "Investigating the neural basis of the auditory continuity illusion," J. Cogn Neurosci. **17**, 1275–1292.

Kluender, K. R., and Jenison, R. L. (**1992**). "Effects of glide slope, noise intensity, and noise duration on the extrapolation of FM glides through noise," Percept. Psychophys. **51**, 231–238.

Levitt, H. (**1971**). "Transformed up-down methods in psychophysics," J. Acoust. Soc. Am. **49**(2), 467–477.

Macmillan, N. A., and Schwartz, M. (**1975**). "A probe-signal investigation of uncertain-frequency detection," J. Acoust. Soc. Am. **58**, 1051–1075.

Manley, B. F. J. (**1997**). *Randomization, Bootstrap, and Monte Carlo Methods in Biology*, 2nd ed. (Chapman & Hall, New York/CRC, Boca Raton, FL).

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

Moore, B. C. J., Hafter, E. R., and Glasberg, B. R. (**1996**). "The probe-signal method and auditory-filter shape: Results from normal and hearing-impaired listeners," J. Acoust. Soc. Am. **99**, 542–552.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, London).

Orban de Xivry, J., Bennet, S. J., Lefevre, P., and Barnes, G. R. (**2006**). "Evidence for synergy between saccades and smooth pursuit during transient target disappearance," J. Neurophysiol. **95**, 418–427.

Oxenham, J., and Plack, C. J. (**2000**). "Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity," Hear. Res. **150**, 258–266.

Pavel, M. C. H. and Stone, V. (**1992**). "Extrapolation of linear motion," Vision Res. **32**, 2177–2186.

Peterken, C., Brown, B., and Bowman, K. (**1991**). "Predicting the future position of a moving target," Percept. Psychophys. **20**, 5–16.

Petkov, C. I., O'Connor, K. N., and Sutter, M. L. (**2007**). "Encoding of illusory continuity in primary auditory cortex," Neuron **54**, 153–165.

Pollack, I. (**1977**). "Continuation of auditory frequency gradients across temporal breaks: The auditory Poggendorff," Percept. Psychophys. **21**, 563–568.

Rushton, S. K., and Wann, J. P. (**1999**). "Weighted combination of size and disparity: A computational model for timing a ball catch," Nat. Neurosci. **2**, 186–190.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (**1987**). "Focused auditory attention and frequency selectivity," Percept. Psychophys. **42**, 215–223.

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.

Sivonen, P., Maess, B., and Friederici, A. D. (**2006**). "Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context," Neurosci. Lett. **408**, 220–225.

Stephens, S. D. G. (**1973**). "Auditory temporal integration as a function of intensity," J. Sound Vib. **37**, 235–246.

Sugita, Y. (**1997**). "Neuronal correlates of auditory induction in the cat cortex," NeuroReport **8**, 1155–1159.

Szelag, E., Kowalska, J., Rymarczyk, K., and Poppel, E. (**2002**). "Duration processing in children as determined by time reproduction: implications for a few seconds temporal window," Acta Psychol. **110**, 1–19.

Warren, R. M., Obusek, C. J., and Ackroff, J. M. (**1972**). "Auditory induction: Perceptual synthesis of absent sounds," Science **176**, 1149–1151.

Warren, R. M. (**1999**). *Auditory Perception: A New Analysis and Synthesis*, 1st ed. (Cambridge University Press, Cambridge).

Wickens, T. D. (**2002**). *Elementary Signal Detection Theory* (Oxford University Press, New York).

Witten, I. B., Bergan, J. F., and Knudsen, E. I. (**2006**). "Dynamic shifts of the owl's auditory space map predict moving sound localization," Nat. Neurosci. **9**, 1439–1445.
Wright, B. A., and Dai, H. (**1993**). "Detection of unexpected tones with short and long durations," J. Acoust. Soc. Am. **95**, 931–938.
Wright, B. A., and Dai, H. (**1998**). "Detection of sinusoidal amplitude modulation at unexpected frequencies," J. Acoust. Soc. Am. **104**, 2991–2996.
Yantis, S. (**1995**). "Perceived continuity of occluded visual objects," Psychol. Sci. **6**, 182–186.

Sound segregation based on temporal envelope structure and binaural cues

Othmar Schimmel
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Steven van de Par[a] and Jeroen Breebaart
Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

Armin Kohlrausch
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands
and Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

(Received 29 January 2007; revised 18 April 2008; accepted 21 May 2008)

The ability to segregate two spectrally and temporally overlapping signals based on differences in temporal envelope structure and binaural cues was investigated. Signals were a harmonic tone complex (HTC) with 20 Hz fundamental frequency and a bandpass noise (BPN). Both signals had interaural differences of the same absolute value, but with opposite signs to establish lateralization to different sides of the medial plane, such that their combination yielded two different spatial configurations. As an indication for segregation ability, threshold interaural time and level differences were measured for discrimination between these spatial configurations. Discrimination based on interaural level differences was good, although absolute thresholds depended on signal bandwidth and center frequency. Discrimination based on interaural time differences required the signals' temporal envelope structures to be sufficiently different. Long-term interaural cross-correlation patterns or long-term averaged patterns after equalization-cancellation of the combined signals did not provide information for the discrimination. The binaural system must, therefore, have been capable of processing changes in interaural time differences within the period of the harmonic tone complex, suggesting that monaural information from the temporal envelopes influences the use of binaural information in the perceptual organization of signal components.
 [DOI: 10.1121/1.2945159]

PACS number(s): 43.66.Pn, 43.66.Mk, 43.66.Rq [RYL] Pages: 1130–1145

I. INTRODUCTION

The perceptual organization of sound sources within an auditory scene is based on grouping the signal components that are likely to come from the same source into one auditory object. The grouping of signal components into auditory objects is established by similarities in spectral, temporal, and/or spatial cues, whereas dissimilarities in these cues cause auditory objects to segregate from each other (cf. Bregman, 1990). The process by which a composite signal is analyzed to identify its constituent components, and the interactions between various grouping cues that contribute to segregation, are, however, not yet fully understood.

For example, when the spectra of two simultaneously presented auditory objects are different, the binaural information within each left-right pair of auditory filters is dominated by the spatial cues of the auditory object that has the most energy in the frequency range of those auditory filters. Although, in principle, this could provide sufficient cues for parsing individual auditory objects, two concurrent tones that are harmonically related cannot be segregated by their differences in interaural timing (Buell and Hafter, 1991). This suggests that signal components with a harmonic relationship are grouped into one auditory object, regardless of their spatial cues. Furthermore, grouping of competing frequency bands to form vowel-like sounds cannot be performed based solely on shared interaural time differences (Culling and Summerfield, 1995). Thus, in addition to different spectral cues and interaural time differences, additional evidence supporting the presence of another auditory object is required for segregation to occur. Such evidence could consist of a repeating tone, which can capture one harmonic of a vowel into a separate auditory object (Hukin and Darwin, 1995; Darwin and Hukin, 1997). Alternatively, correlated dynamic variation in frequency and/or amplitude, such as in speech, is sufficient to segregate two vowel-like sounds based on interaural time differences (Stern *et al.*, 2006) or increases speech intelligibility in the presence of spatially separated competing speech or noise signals (Noble and Perrett, 2002).

The above studies indicate that segregation of concurrent auditory objects depends on the specific spectral, temporal, and spatial cues of the signal components. In contrast to spectral and temporal grouping cues, spatial cues alone do not appear to be sufficient to *cause* segregation (cf. Culling and Summerfield, 1995). Spatial cues do, however, *add to* segregation that is established on the basis of monaural spectral or temporal grouping cues (Shackleton and Meddis, 1992; Darwin and Hukin, 1998). Subjects were shown to use the continuity of interaural time differences as an important

[a]Electronic mail: steven.van.de.par@philips.com

cue for sequential organization of signal components, i.e., tracking an auditory object across time (Hukin and Darwin, 1995; Darwin and Hukin, 1997, 1999; Shinn-Cunninham, 2005). It seems that monaural grouping cues are essential for segregation of spatially separated auditory objects to occur, and that these monaural cues influence the contribution of binaural cues to the segregation.

Based on these results from literature, the question arises how the monaural grouping cues influence the ability to use binaural grouping cues in the perceptual organization of concurrent signals. The current study focused on the relationship between temporal and binaural cues. Experiments that explored whether and to what extent subjects can discriminate between the different spatial configurations of two concurrent and spectrally overlapping signals that differed in their temporal envelope structures are described. Each of the two signals is easily recognized when listened to in isolation. By presenting the signals simultaneously, with equal spectral excitation patterns and interaural differences of the same absolute value but with opposite signs, binaural cues from both signals are, on average, equally represented in each frequency band of the signals. This way, only temporal cues from the signals' different temporal envelope structures and binaural cues are available in the composite signal for discrimination between its different spatial configurations. It is of interest to investigate whether subjects can analyze the temporal envelope cues and the interaural differences present in the composite signal to distinguish between the different spatial configurations, and to what extent they are able to associate the binaural information with the underlying signal components.

II. GENERAL METHOD

As an indication for segregation ability, the experiments explored the ability of subjects to discriminate between the spatial configurations of two simultaneously presented, spectrally overlapping signals with different temporal envelopes. The two signals were a harmonic tone complex (HTC) with a 20 Hz fundamental frequency and a bandpass noise (BPN), which occupied the same spectral range. Both signals had interaural differences of the same absolute value, but with opposite signs, to establish lateralization to different sides of the medial plane. Their combination yielded two different spatial configurations, i.e., a configuration with the HTC on the one side and the BPN on the other side, and the reverse configuration.

All experiments used the same method to measure thresholds for various signal parameters. A three-interval, three-alternative, forced-choice procedure with an adaptive parameter adjustment was used in a within-subject experimental design with counterbalanced block randomization. The subjects' task was to identify the interval that was different from the other two intervals. Feedback was provided after each trial. The adaptive parameter [interaural time difference (ITD) or interaural level difference (ILD)] was adjusted according to a two-down one-up rule, to track the 70.7%-correct response level (Levitt, 1971), by multiplying or dividing it with a certain factor. Initially, this factor was 2.51 ($=10^{8/20}$). After each second reversal, the factor was reduced by taking its square root until the minimum factor of 1.12 ($=10^{1/20}$) was reached. Another eight reversals were measured at this minimum factor, and the median of these eight values was used to estimate the threshold.

For each condition, at least four attempts were made by each subject to measure a threshold. However, when the adaptive interaural difference exceeded a limit of 2 ms ITD or 96 dB ILD, the tracking procedure was terminated and no threshold was registered. For conditions where incidently no threshold was registered, the measurements were repeated to obtain a total of at least four threshold values. Because of possible lateralization ambiguities in the ITD conditions, resulting from phase shifts beyond π for the highest-frequency components of the HTC due to the 1 ms ITD starting value, the conditions that did not yield threshold values were repeated with a 100 μs ITD starting value. For conditions that did not consistently yield threshold values, measurements were stopped and occasionally obtained threshold values were discarded. The measured thresholds for each condition were pooled and checked for severe outliers (thresholds that deviated from the average more than three times the interquartile range of the pooled data for each condition), which were then removed from the data set. Five male subjects without any reported hearing problem, including the four authors, participated in the experiments.

The experiments were conducted in an acoustically isolated listening room at the Philips Research Laboratories. A computer running MATLAB software generated the stimuli and automated the experiments and data collection. Digital stimuli were converted to analog signals by a Marantz CDA-94 two-channel 16 bit digital-to-analog converter at a sampling rate of 44.1 kHz and presented to the subjects over Beyer Dynamic DT990 Pro headphones.

In Sec. III, the ability to discriminate between the spatial configurations of two spectrally and temporally overlapping signals with different temporal envelope structures is explored. In Secs. IV and V, it is checked whether the results of Sec. III may have been obtained by judging an overall lateralization of the composite signal or by monaural listening. In Sec. VI, various bandwidth conditions were explored to further investigate the observed influence of bandwidth on discrimination performance in Sec. III. In Sec. VII, various temporal envelope structures were applied to the individual signals to investigate their effect on the ability to discriminate between the spatial configurations of the signals.

III. EXPERIMENT 1

The first experiment investigated the extent to which subjects could discriminate between the spatial configurations of two spectrally and temporally overlapping signals with different temporal envelope structures. As a reference for discrimination between the spatial configurations of the signals, the just-noticeable differences in lateralization for each of the individual signals were also measured.

A. Stimuli

The stimuli consisted of two signals, for which the choice of parameters was inspired by a related earlier study (van de Par *et al.*, 2005): a HTC with a 20 Hz component spacing and a BPN. The HTC was defined as a complex of sinusoidal components at multiples of 20 Hz, with zero starting phase as follows

$$x(n) = \sum_{k=i}^{j} A \sin\left(2\pi k f_0 \frac{n}{f_s}\right), \quad (1)$$

where n is the sample number, f_s is the sample frequency, and i and j are integers that indicate the harmonic number of the lowest and highest components in the HTC. The noise was generated in the time domain by creating a 3000 ms buffer containing a broadband noise with a Gaussian probability distribution. This buffer was transformed to the frequency domain using a fast fourier transform (FFT), and the appropriate samples were set to zero to obtain the required bandwidth and center frequency. Using an inverse FFT, the buffer was transformed back to the time domain. For each threshold measurement, one bandpass filtered buffer was generated, and for each trial, three independent 400 ms excerpts were selected from the buffer by drawing random starting positions from a uniform distribution. Both signals had the same spectral range with a flat spectral envelope and the same overall level [65 dB sound pressure level (SPL)], but differed in their temporal envelope structures. Whereas the BPN had an irregular temporal envelope, the envelope of the HTC was highly modulated and had a period of 50 ms.

For measuring the discrimination between their spatial configurations, the two signals were presented concurrently, resulting in an overall stimulus level of 68 dB SPL. The target interval of the forced-choice procedure had interaural time or level differences for the two signals such that the HTC was lateralized to the right and the BPN to the left, using interaural time or level differences of the same absolute value, but with opposite signs. In the two reference intervals, the interaural differences of both signals were reversed. The subjects' task was to identify the target interval that differed from the other two intervals, i.e., the interval in which the HTC was lateralized to the right and the BPN to the left. These conditions are referred to as *composite signal* conditions.

For measuring the just-noticeable differences in lateralization for the individual signals, the same HTC and BPN were presented in isolation. The target interval of the forced-choice procedure had an interaural time or level difference such that the signal was lateralized to the right. The reference intervals had an identical opposite binaural cue such that the signal was lateralized to the left. Again, the subjects' task was to identify the target interval that differed from the other two intervals, i.e., the interval in which the signal was lateralized to the right. These conditions are referred to as *single signal* conditions. Please note that the interaural differences as reported hereafter are relative to the medial plane, and that, due to the lateralizations to opposite sides of the medial plane, the total interaural difference cue between the signals is, in fact, twice the size of these interaural differences.

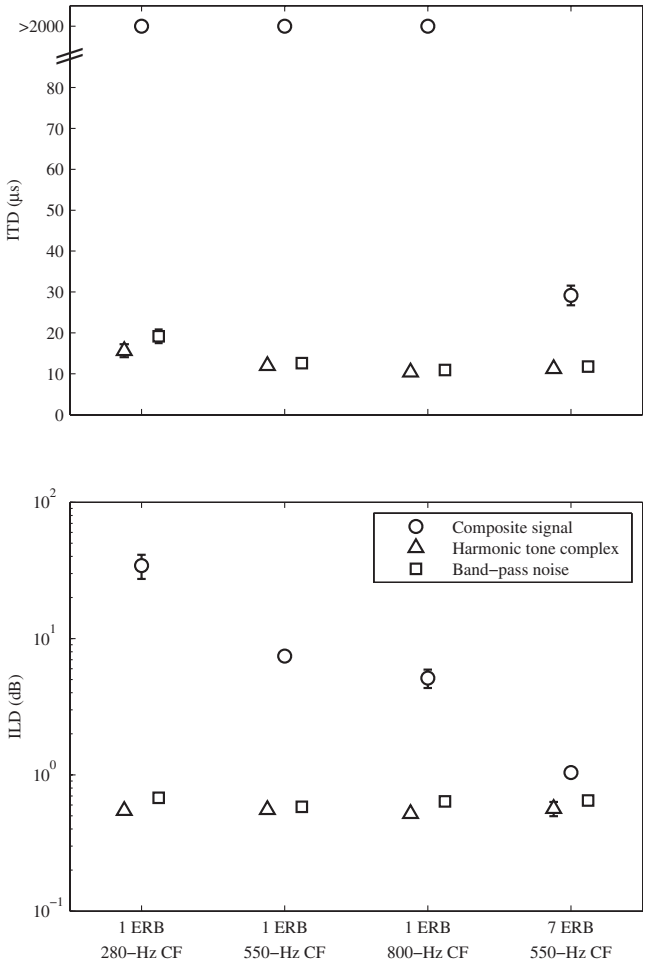


FIG. 1. Mean ITD (top panel) and ILD (bottom panel) thresholds for the composite and single signal conditions from Exp. 1. ITD thresholds that could not be measured are indicated by a symbol at a threshold of >2000 μs. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

The two signals were of various equal bandwidths. In the narrowband conditions, the signals were the width of one auditory filter [1 Equivalent Rectangular Bandwith (ERB)] and centered at 280 Hz (bandwidth of 60 Hz), 550 Hz (bandwidth of 80 Hz), or 800 Hz (bandwidth of 100 Hz). In the wideband condition, the signals were 600 Hz wide (7 ERB) and centered at 550 Hz. The intervals had a duration of 400 ms, including 30 ms raised-cosine onset and offset ramps to avoid spectral splatter, and were separated by 300 ms of silence. For the ILD conditions, the level changes were such that the level of the combined signals remained constant at 68 dB SPL.

B. Results

Figure 1 displays the mean thresholds and the standard errors of the mean of the pooled data of the five subjects for the composite and single signal conditions. The top panel shows the data for the ITD conditions and the bottom panel the data for the ILD conditions. The abscissa indicates the four signal bandwidth conditions. The ordinate indicates the size of the thresholds in microseconds (ITD) or decibels

(ILD). The symbols represent the mean thresholds for the composite signal conditions (circles), and for the HTC (triangles) and BPN (squares) in the single signal conditions. The ITD conditions for which no threshold could be obtained, because the adjusted value exceeded the procedural limit of 2 ms, are indicated by a symbol at a threshold of >2000.

For the ITD conditions, all thresholds for the single signal conditions were about 10–20 μs, and decreased slightly for both signals with an increase in center frequency. The thresholds for the wideband condition agreed with those for the narrowband condition with the highest center frequency. These results indicate similar lateralization performance for both signals when presented in isolation, independent of the signals' detailed spectral and temporal features. For the composite signal conditions, no thresholds could be obtained for the narrowband conditions, indicating that discrimination between the two signals' spatial configurations based on interaural time differences was not possible when their spectral energy was limited to one auditory filter. For the wideband condition, the threshold was 29 μs, indicating that such a discrimination was possible when the spectral energy of the two signals was present across multiple auditory filters.

For the ILD conditions, all thresholds for the single signal conditions were about 0.6 dB, independent of the center frequency and bandwidth. For the composite signal conditions, the thresholds were in the range 1–32 dB. For the narrowband conditions, thresholds decreased from 32 to 5 dB with increasing center frequency. The 32 dB threshold for the narrowband condition at 280 Hz center frequency is, however, beyond a plausible range for naturally occurring interaural level differences, and reflects the difficulty of assigning these binaural cues to either one of the two signals. For the wideband condition, the threshold of 1 dB was close to the corresponding thresholds in the single signal condition. These results indicate that, in contrast to the narrowband ITD conditions, discrimination between the two signals' spatial configurations based on interaural level differences was possible within a single auditory filter, although the thresholds were considerably higher than for the single signal conditions. Analogous to the wideband ITD conditions, performance in the composite signal conditions was best when the signals' spectrum covered multiple auditory filters.

C. Discussion

When asked, all subjects reported that in the composite signal the two constituent signals, and predominantly the HTC, could easily be recognized and, especially in the ILD conditions, lateralized when they had a sufficiently large interaural difference cue. At an interaural difference cue close to the threshold, their individual lateralizations would become much harder to discern, although the presence of the HTC, as the focus for identifying the target interval, could still be recognized in the composite signal. This recognition may, however, be due to the extended exposure to, and focus on the HTC during the procedure, and it is possible that both signals would otherwise have been merged into one auditory object based on their common properties. It seems that in these conditions, which are (close to) being perceived as diotic, the known characteristics of the specific temporal envelope of the HTC still allowed its segregation from the composite signal.

The fact that subjects were able to discriminate between the spatial configurations of the target and reference intervals in the composite signal condition suggests that subjects were able to segregate the two signals and discern the lateralization of at least one of them. A possible objection to this interpretation of the obtained results may be that successful identification of the target and reference intervals also could have been performed otherwise, for instance, based on an asymmetry in the overall spatial image of the composite signal.

The stimuli were physically constructed in such a way that, due to the two constituent signals' interaural differences of the same absolute value but with opposite signs, the long-term interaural cross-correlation patterns or long-term patterns after equalization-cancellation for the composite signal were essentially identical for target and reference intervals. Figure 2 shows an example of the long-term interaural cross-correlation patterns for the ITD condition of the individual wideband signals (dotted and dashed lines) and their composite signal (solid line) for one interval, computed from the output of the auditory filter centered at 550 Hz of a gammatone filterbank. In this example, the individual signals have an opposite interaural time difference of 100 μs, which is reflected in slightly different positions of the maxima of the cross-correlation patterns. The cross-correlation pattern of the composite signal has its maximum at 0 μs, and is highly symmetrical, as can be seen by comparing the pattern (the solid line) with its mirrored version (the circles on the solid line). Therefore, the long-term interaural cross correlation of the composite signal is not expected to provide directional information for distinguishing between target and reference intervals.

Similarly, the spatial perception of a single auditory object consisting of multiple merged signals is based on a weighted average of the various directional cues of the constituent signals, which fuse into a single intracranial image (Stellmack and Lutfi, 1996). Given their common temporal onset, spectral cues, and identical but opposite interaural differences, an unsuccessful segregation of the two signals would be expected to result in lateralization in the center and, therefore, inhibit discrimination between the different spatial configurations of the target and reference intervals.

However, this reasoning is based on linear processing in the auditory system, while the contributions of the two different signals to the internal long-term interaural cross correlation may be different. For instance, peripheral compression may have reduced the peaks in the temporal envelope of the HTC stronger than those in the temporal envelope of the BPN. If so, the combination of the two signals could have been perceived as a single auditory object with an averaged lateralization different from the center. Then, the intracranial image would be lateralized to different sides for the target and reference intervals, and subjects could have used this change in overall lateralization of the composite signal to

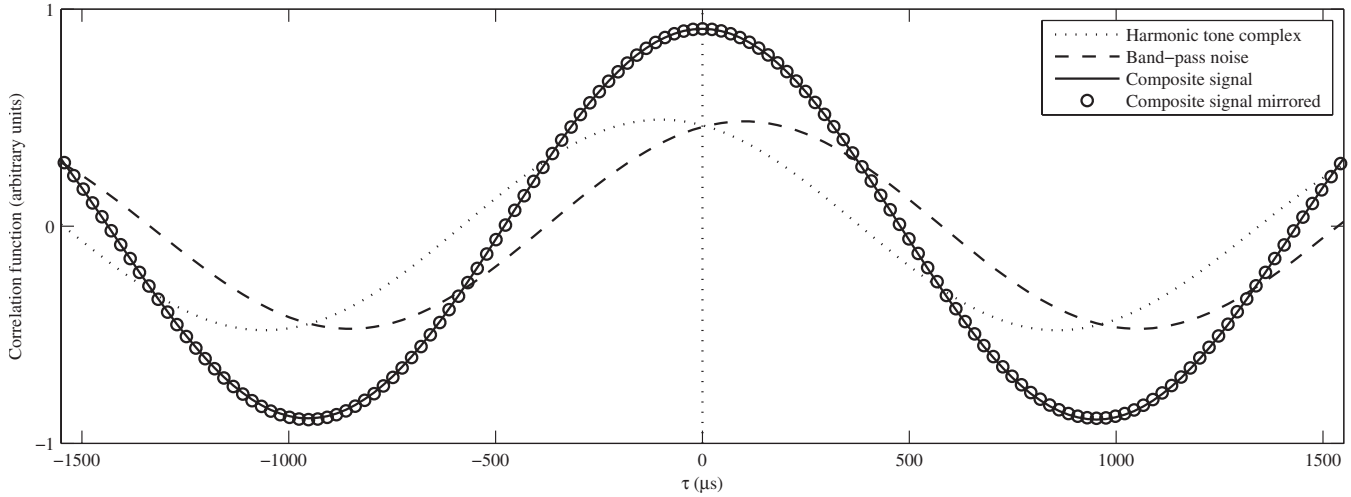


FIG. 2. Long-term interaural cross correlations of the signals at the output of the auditory filter centered at 550 Hz. The dash-dotted and dashed lines represent the interaural cross correlations of the individual signals, showing opposite lateralizations (peaks centered at $\tau = \pm 100$ μs). The solid line represents the interaural cross correlation of the composite signal, showing the effective cancellation of the individual signals' lateralizations (peak centered at $\tau = 0$ μs). The circles represent the cross correlation of the composite signal when plotted mirrored. Note that their pattern is nearly identical to the solid line.

identify the intervals, without the need to discriminate between the spatial configurations of the two signals.

Another possible criticism to the current interpretation of the results is that, in the ILD conditions, differences in the relative levels of the constituent stimulus components could have allowed monaural listening. The discrimination between the spatial configurations of the signals could then have been performed by listening to one ear and selecting the target interval based on overall monaural level cues. Because the interpretation of the results is only valid if the identification of the target and reference intervals could not have been performed other than through segregation and lateralization of at least one of the individual signals, Exps. 2A and 2B in the next section were designed to test the two mentioned alternative accounts for the results of Exp. 1.

IV. EXPERIMENT 2A

This first control experiment investigated whether subjects could have performed the identification of the target and reference intervals for the stimuli with interaural time differences of Exp. 1 by judging a possible overall lateralization of the composite signal, without the need to discriminate between the spatial configurations of the two signals. To effectively inhibit an overall lateralization as a cue for performing the identification, in each interval the actual interaural time difference, as adjusted by the adaptive procedure, was changed for both signals by the same amount using a random offset (rove).

A. Stimuli

The measurements were limited to the wideband ITD condition of Exp. 1, using the same HTC and BPN in the composite signal condition. The amount of roving was random for each interval and equally distributed between zero and the size of the currently adapted interaural time difference, in order to enable a maximum rove while keeping the signals on the opposite sides of the medial plane. It was applied in the *same direction* to both signals to keep the absolute difference in directional cues between them constant. This way, a possible overall lateralization would change randomly in each interval within a trial, while the interaural difference cue between the signals was maintained.

B. Results

The mean threshold for the nonroving condition from Exp. 1 was 29 μs, and for the roving condition from the current experiment it was 34 μs. Both values had the same standard error of the mean of 2 μs. Analysis by a t test showed that these results were statistically not significantly different ($p = 0.109$).

C. Discussion

The 5 μs difference in mean thresholds between the nonroving and roving conditions was well below the 10–20 μs thresholds from the single signal conditions, indicating that the difference between the nonroving and roving conditions is too small to be considered perceptually relevant. If an overall lateralization of the composite signal was the main cue for identifying the target and reference intervals, a significant increase in mean threshold was expected. Finding no significant or perceptually relevant effect of roving the interaural differences on the mean threshold leads to the conclusion that the results from Exp. 1 cannot be explained by an overall lateralization of the composite signal.

V. EXPERIMENT 2B

This second control experiment investigated whether identification of the target and reference intervals for the stimuli with interaural level differences from Exp. 1 could have been performed by monaural listening, or whether it required an analysis of binaural cues. The contributions of monaural and binaural listening were addressed separately.

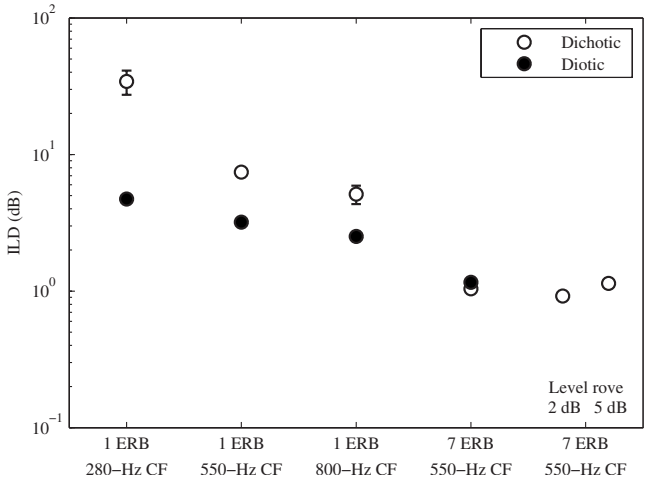


FIG. 3. Mean ILD thresholds for the dichotic signal conditions from Exp. 1 and the diotic and dichotic signal conditions from Exp. 2B. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

A. Stimuli

To investigate the potential of monaural listening, the measurements were repeated for all ILD conditions from Exp. 1, with the modification that only the right-ear signal was presented. It was presented diotically, to establish that all binaural cues were excluded and only the monaural level cues within one ear, resulting from the different temporal envelopes of the HTC and the BPN, were available for identifying the target and reference intervals.

To investigate the influence of binaural listening, the measurements were repeated for the wideband ILD condition from Exp. 1. For each interval within a trial, a random level rove, equally distributed between −2 and 2 dB or between −5 and 5 dB and sufficiently larger than the thresholds for the wideband composite and single signal conditions, was applied to the HTC. This level rove was followed by an independent level normalization of the composite left- and right-ear signals to 68 dB SPL each. Due to the resulting variability in the relative levels of the HTC and the BPN, the monaural level cues were unreliable and only the binaural cues were available for identifying the target and reference intervals.

B. Results

Figure 3 displays the mean thresholds and standard errors of the mean of the pooled data of the five subjects for the dichotic signal conditions from Exp. 1 and the diotic and dichotic signal conditions from the current experiment. The abscissa indicates the three narrowband conditions, the wideband condition, and the level-rove wideband condition. The ordinate indicates the size of the thresholds in decibels. The symbols represent the mean thresholds for the dichotic signal conditions (open markers) and the diotic signal conditions (closed markers).

Monaural (diotic) thresholds in the narrowband conditions decreased with an increase in center frequency, much like the corresponding binaural (dichotic) thresholds from Exp. 1. However, the monaural thresholds were considerably lower than the binaural thresholds, indicating better performance in the diotic conditions. The monaural and binaural thresholds for the wideband condition were similar, indicating equivalent performance. Analysis of variance on the means of the dichotic versus diotic conditions showed significant effects of signal bandwidth ($F_{(3,153)}=152.88$, $p<0.001$) and interaural condition ($F_{(1,153)}=82.72$, $p<0.001$), and a significant interaction between these two parameters ($F_{(3,153)}=23.82$, $p<0.001$). Tukey *post hoc* analysis revealed that, for all three narrowband conditions, the diotic conditions were significantly different from the dichotic conditions. For the wideband condition, the diotic and dichotic conditions were statistically identical.

The thresholds for the two dichotic conditions in which the level was randomly roved to inhibit overall level cues for identifying the target and reference intervals were also similar to the threshold of the wideband condition without a level rove, showing equivalent performance as well. Analysis of variance on the means of the wideband nonroving and the two roving conditions showed that these three conditions were statistically not significantly different from each other ($F_{(2,63)}=2.02$, $p=0.142$).

C. Discussion

The difference between diotic and dichotic listening in the narrowband conditions shows that better performance could have been achieved if subjects were able to listen to the signals at one ear only, which apparently they could not. Because the main difference between the diotic and dichotic listening conditions is the presence of binaural cues, this finding reveals that binaural cues actually *reduce* the performance for narrowband stimuli, while for wideband stimuli, performance is similar. This result suggests that, for narrowband stimuli, binaural processing cannot be ignored. Although this may seem surprising, it was shown before that subjects are sometimes unable to ignore information presented to one ear while performing a task that could be solved with information simultaneously presented to the other ear (Heller and Trahiotis, 1995; Gallun *et al.*, 2007). For the wideband conditions, the similarity between the results for diotic and dichotic listening indicates that the target and reference intervals could be identified equally well using either monaural or binaural cues, making it impossible to distinguish between monaural and binaural listening in the wideband ILD condition.

To investigate the contribution of binaural listening in the wideband ILD condition, the relative levels of the HTC and the BPN were randomly roved. Due to the random level roves, the overall monaural level cues were not reliable and identification of the target and reference intervals must have resulted from binaural processing. The results for these level-roving conditions were, however, similar to the corresponding nonroving condition of Exp. 1. This result suggests that, for the wideband ILD conditions, identification of the target

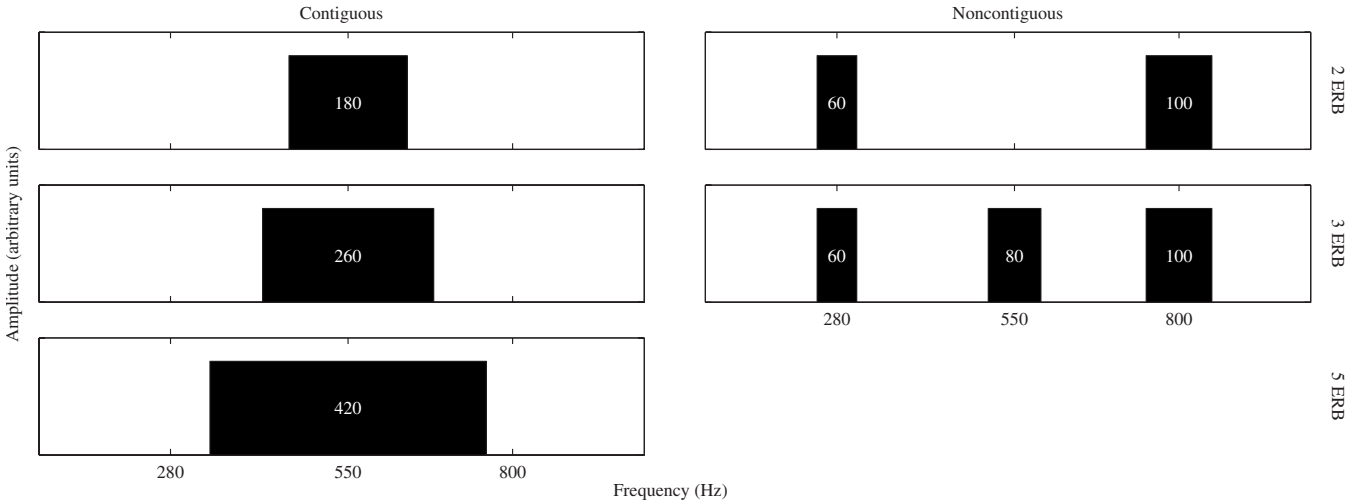


FIG. 4. Schematic illustration of the conditions of Exp. 3: Signal center frequency and bandwidth for the HTC and the BPN (in Hz), with the spectral energy distributed over contiguous or noncontiguous auditory filters.

and reference intervals could have been mediated by binaural cues. It remains, therefore, impossible to distinguish between monaural and binaural listening in these conditions.

VI. EXPERIMENT 3

The results of Exp. 1 showed that signal bandwidth has a pronounced effect on the ability to discriminate between the spatial configurations of the two signals. To investigate whether the observed increase in performance with an increase in bandwidth is due to across-channel or within-channel cues, various bandwidth conditions were explored, with the spectral energy of both signals distributed over either contiguous or noncontiguous auditory filters.

A. Stimuli

Both the HTC and the BPN were presented with bandwidths of 2, 3, and 5 ERB, all centered at 550 Hz. In the case of 2 and 3 ERB bandwidths, both signals were presented in contiguous as well as in noncontiguous auditory filters, i.e., the spectral energy of the signals was distributed over frequency bands of 1 ERB wide that were not adjacent. These noncontiguous conditions involved the same bands as the narrowband conditions from Exp. 1. For the 2 ERB condition, the bands were centered at 280 and 800 Hz; for the 3 ERB condition, the bands were centered at 280, 550, and 800 Hz. Figure 4 gives a schematic overview of the spectral conditions.

B. Results

Figure 5 displays the mean thresholds and the standard errors of the mean of the pooled data of the five subjects for the various signal bandwidth conditions. The top panel shows the data for the ITD conditions and the bottom panel shows the data for the ILD conditions. The abscissa indicates the five signal bandwidth conditions, including the narrowband centered at 550 Hz (1 ERB) and wideband (7 ERB) conditions from Exp. 1. The ordinate indicates the size of the thresholds in microseconds (ITD) or decibels (ILD). The symbols represent the mean thresholds for the contiguous auditory filter conditions (open markers) and the noncontiguous auditory filter conditions (closed markers).

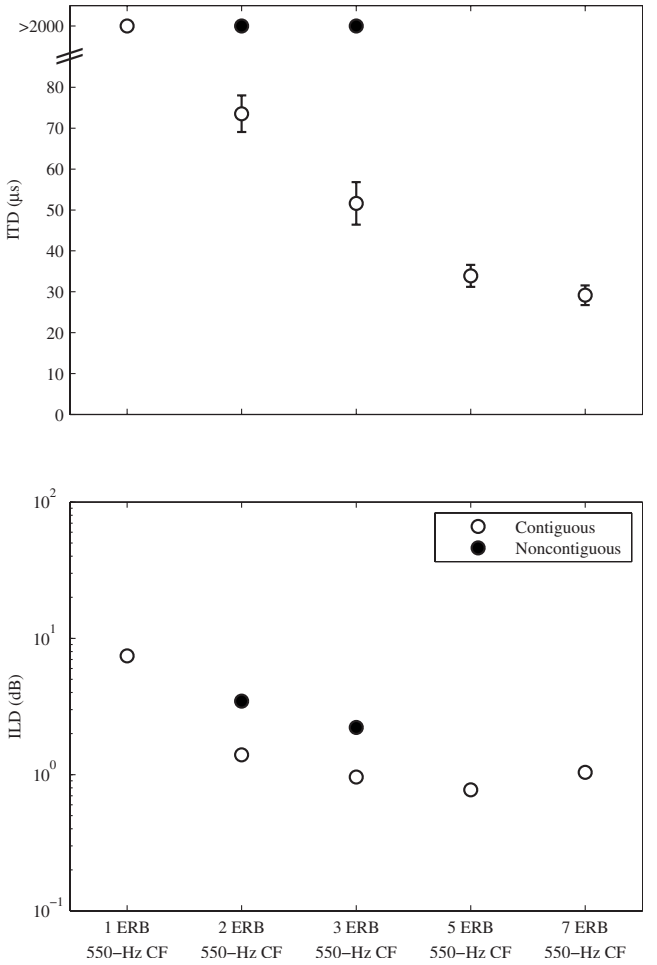


FIG. 5. Mean ITD (top panel) and ILD (bottom panel) thresholds for the signal bandwidth conditions from Exp. 3. Included for reference are the results for the narrowband (1 ERB) and wideband (7 ERB) conditions from Exp. 1. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

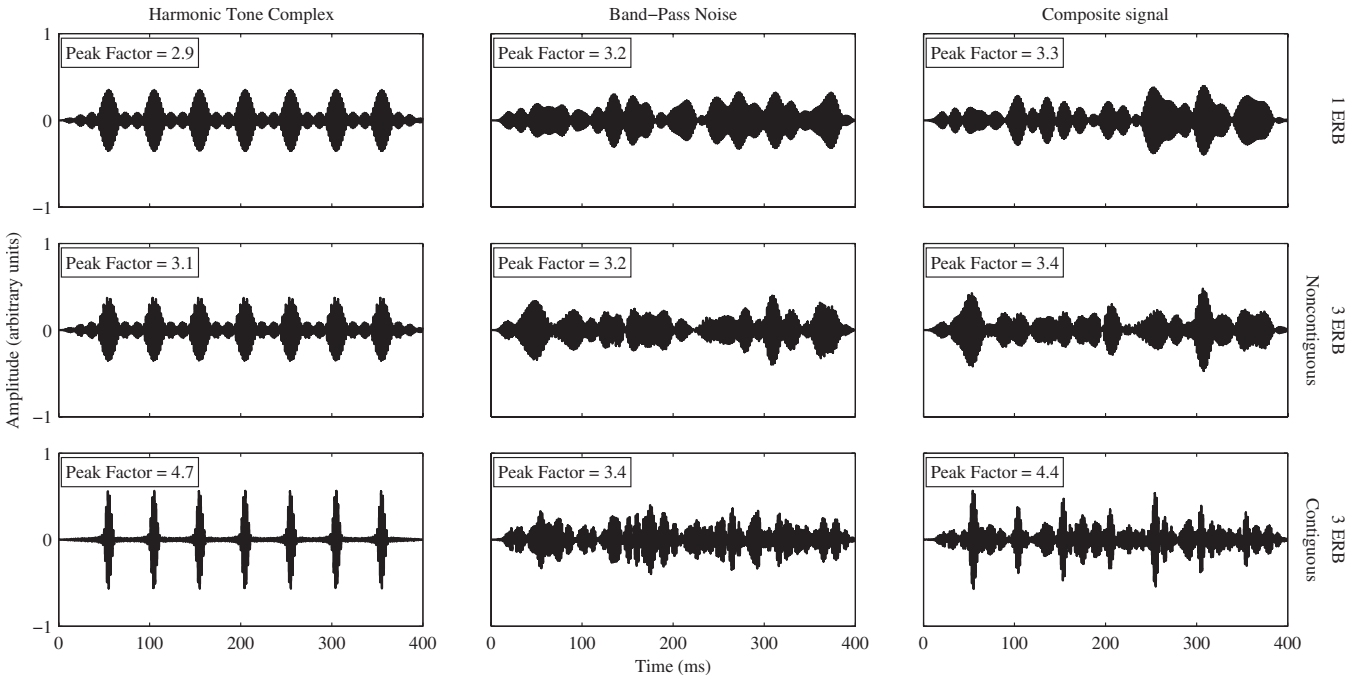


FIG. 6. Output of the auditory filter centered at 550 Hz for the HTC (left column), the BPN (middle column), and the composite signal (right column) at signal bandwidths of 1 ERB (top row) and 3 ERB in both noncontiguous (middle row) and contiguous auditory filters (bottom row). The insets show the peak or crest factor, i.e., the signal's peak amplitude divided by its rms value.

For the ITD conditions, thresholds of 74 and 52 μs were measured for the 2 and 3 ERB bandwidths, respectively, provided that the signals' spectral energy was in contiguous auditory filters. No thresholds could be obtained for these bandwidths when the signals' spectral energy was in noncontiguous auditory filters of 1 ERB width. These results indicate that the combination of the information across auditory filters *per se* did not improve the performance. The threshold of 34 μs for the 5 ERB condition was already similar to the previously established threshold (29 μs) for the 7 ERB wideband condition. These data indicate that discrimination between the signals' spatial configurations based on interaural time differences improves when their spectral energy covers an increasing number of contiguous auditory filters.

For the ILD conditions, the threshold for signals with a bandwidth of 2 ERB in contiguous auditory filters was 1.4 dB, and already very close to the thresholds of 1 dB for all larger signal bandwidths in contiguous auditory filters. For the signal bandwidths with the spectral energy in two and three noncontiguous auditory filters, thresholds were 3.5 and 2.2 dB, respectively 2.1 and 1.2 dB higher than for the corresponding bandwidth conditions with contiguous auditory filters. These data indicate that, again in contrast to the corresponding ITD conditions, discrimination of the signals' spatial configurations based on interaural level differences was possible for narrow bandwidths, even when the spectral energy was in noncontiguous auditory filters. When the bandwidth was increased by distributing the spectral energy over two or three noncontiguous auditory filters, thresholds decreased compared to the condition with the spectral energy in only one auditory filter, showing an ability to combine information from the auditory filters and improve the performance.

C. Discussion

The obtained data, in particular those for the ITD conditions, indicate a specific effect of having the signal in a number of contiguous auditory filters. The ability to discriminate between the spatial configurations of the HTC and the BPN increases with increasing signal bandwidth, which may be caused by the peakedness of the temporal envelopes of the presented stimuli. With increasing bandwidth, the temporal envelope of the BPN does not change, except for a relative increase in higher envelope frequencies. The temporal envelope of the HTC, which is similar to the temporal envelope of the BPN when the bandwidth is small, becomes increasingly peaked with an increase in bandwidth, and, therefore, progressively different from the temporal envelope of the BPN.

Figure 6 shows the output for an auditory filter, centered at 550 Hz, computed using a gammachirp filterbank (Irino and Patterson, 2006), for the HTC (left column), the BPN (middle column), and the composite signal (right column). Shown are the results for signal bandwidths of 1 ERB (top row), 3 ERB in noncontiguous auditory filters (middle row), and 3 ERB in contiguous auditory filters (bottom row). For each signal, the insets indicate the peak or crest factor, i.e., the signal's peak amplitude divided by its rms value. The peak factor of the BPN is about the same for all three spectral conditions. The peak factor of the HTC increases with the presence of signal energy in adjacent auditory filters (compare top and bottom panels), but is not affected by adding components at remote spectral regions (compare top and middle panels). A similar change in peak factor is observed for the composite signal. The availability of these distinguishable peaks in the composite signal's envelope appar-

ently enhances the ability to process binaural cues of the signal components at these moments in time, similar to an onset response that enables the processing of signal components at a specific spectrotemporal position to be dominant for a brief period of time (Bregman *et al.*, 1994).

In the next experiment, further evidence is provided for the idea that the ability to discriminate between the spatial configurations of two signals is influenced by their temporal envelope structures. From the previous discussion, this ability may be expected to break down when their temporal envelope structures are more similar. To explore a possible breakdown of discrimination ability, the experiments were partially repeated for changed temporal envelope structures of either the HTC or the BPN.

VII. EXPERIMENT 4

This experiment investigated the effect of temporal envelope structures on the ability to discriminate between the spatial configurations of two spectrally and temporally overlapping signals. Various temporal envelope structure manipulations were applied to either the HTC or to the BPN. The experiment was limited to measuring thresholds for the wideband signal conditions, because these measurements led to the lowest thresholds in the previous experiments.

A. Stimuli

The temporal envelope structure of the HTC was manipulated in three different ways to yield a temporal envelope more similar to that of the BPN. First, a random starting phase was applied to each of the HTC components, which results in a temporal envelope with a low peak or crest factor. The HTC was defined as

$$x(n) = \sum_{k=i}^{j} A \sin\left(2\pi k f_0 \frac{n}{f_s} + \Phi_r\right), \tag{2}$$

where Φ_r is a random number uniformly distributed in the range $[0, 2\pi)$.

Second, both positive and negative Schroeder phases were applied to the HTC components, i.e., a downward and upward frequency sweep of the components, resulting in a flat temporal envelope of the HTC (Schroeder, 1970). The HTC was now defined as

$$x(n) = \sum_{k=i}^{j} A \sin\left(2\pi k f_0 \frac{n}{f_s} + \Phi_S(k-i)\right), \tag{3}$$

with

$$\Phi_S(m) = \frac{\frac{1}{2}\pi + c\pi(m+1)m}{j-i+1}, \tag{4}$$

where $c=1$ for a positive Schroeder phase and $c=-1$ for a negative Schroeder phase. Due to the linear frequency modulation, the excitation peak in the different auditory filters was not synchronized as it was for the zero-starting-phase signal.

Figure 7 displays the waveforms of the four phase conditions as applied to the wideband HTC. Each phase condition results in a different temporal envelope structure. The first panel shows the zero-phase HTC as used in Exp. 1, with its regularly peaked pattern. The second panel shows that the periodic pattern of strong peaks for the zero-phase HTC completely disappears when a random phase is given to each component. The third and fourth panels show the positive and negative Schroeder-phase HTCs, respectively, for which the application of the Schroeder phases to the components results in a periodic signal with a flat temporal envelope.

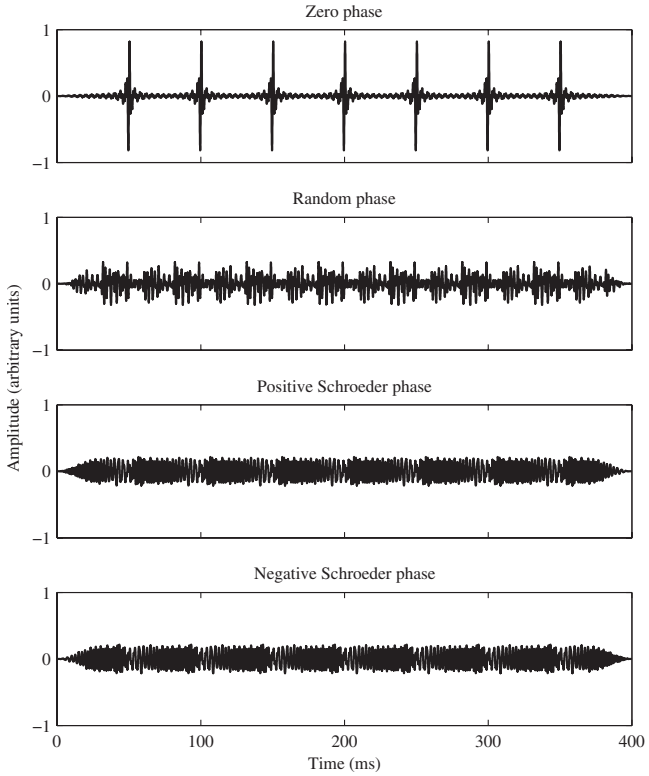


FIG. 7. Four phase conditions of the wideband HTC, resulting in different temporal envelopes. The phase of the components was manipulated to yield temporal envelopes of the HTC more similar to the temporal envelope of the BPN.

Third, the component spacing of the HTC was increased to 40 and 80 Hz, which results in an increase in the number of peaks in the temporal envelope. This manipulation, however, reduces the peak factor after filtering on the basilar membrane, because fewer components fall within the bandwidth of a single auditory filter. The HTC was defined as in Eq. (1), for $f_0=40$ Hz and $f_0=80$ Hz. To preserve the harmonicity of the HTC at multiples of 40 and 80 Hz, the spectral features of both signals had to be slightly changed and were set to a 560 Hz center frequency with a 600 Hz bandwidth for $f_0=40$ Hz, and a 600 Hz center frequency with a 640 Hz bandwidth for $f_0=80$ Hz.

The temporal envelope of the BPN was manipulated by applying a 20 Hz sinusoidal amplitude modulation to yield a temporal envelope with regular peaks at the same period as those of the zero-phase HTC. The amplitude-modulated BPN was defined as

TABLE I. Conditions of Exp. 4: Temporal envelope structure manipulations for the harmonic tone complex (HTC) or the bandpass noise (BPN), including their spectral properties. The manipulation was only applied to the temporal envelope of the signal mentioned; the temporal envelope of the other signal remained the same as in Exp. 1.

Signal	Temporal envelope	CF	BW
HTC	Random phase	550	600
HTC	Positive Schroeder phase	550	600
HTC	Negative Schroeder phase	550	600
HTC	40 Hz fundamental	*560*	600
HTC	80 Hz fundamental	*600*	*640*
BPN	Amplitude modulation (ϕ=0)	550	600
BPN	Amplitude modulation ($\phi=\pi$)	550	600

$$y(n) = \sqrt{\frac{2}{3}}N(n)\left(1 + \cos\left(2\pi f_0 \frac{n}{f_s} + \Phi_{\mathrm{AM}}\right)\right), \quad (5)$$

where N is the BPN, and the multiplication with $\sqrt{\frac{2}{3}}$ is applied to preserve the energy of the noise. In combination with the amplitude-modulated BPN, the HTC was presented with a zero starting phase and a 20 Hz fundamental frequency. The HTC and amplitude-modulated BPN were presented either temporally in phase (Φ_{AM}=0), such that the envelope maxima of both signals coincided, or out of phase ($\Phi_{\mathrm{AM}}=\pi$), such that the envelope maxima of one signal coincided with the envelope minima of the other signal. Table I gives an overview of the signals' temporal envelope structure conditions, including their spectral properties.

B. Results

Figure 8 displays the mean thresholds and the standard errors of the mean of the pooled data of the five subjects for the various temporal envelope structure conditions. The top panel shows the data for the ITD conditions and the bottom panel the data for the ILD conditions. The abscissa indicates the eight temporal envelope structures, including the combination of the wideband zero-phase HTC and the unmodulated BPN from Exp. 1 to the left in both panels. The ordinate indicates the size of the thresholds in microseconds (ITD) or decibels (ILD). The symbols represent the mean thresholds for the manipulations of the temporal envelope of the wideband HTC (open markers) and of the BPN (closed markers).

For the condition with a random starting phase on each component of the HTC, discrimination of the spatial configurations based on interaural time differences was not possible. Discrimination based on interaural level differences was seriously degraded compared to the reference zero-phase condition, and the threshold ILD of 29 dB was beyond the value needed for maximal lateralization.

When positive and negative Schroeder phases were applied to the HTC components, discrimination based on interaural time differences was seriously degraded compared to the 29 μs threshold of the zero-phase condition, with thresholds of 132 and 86 μs, respectively. Discrimination based on interaural level differences was only slightly degraded compared to the 1 dB threshold of the zero-phase condition, with thresholds 1.2 and 0.6 dB higher for the positive and negative Schroeder-phase conditions.

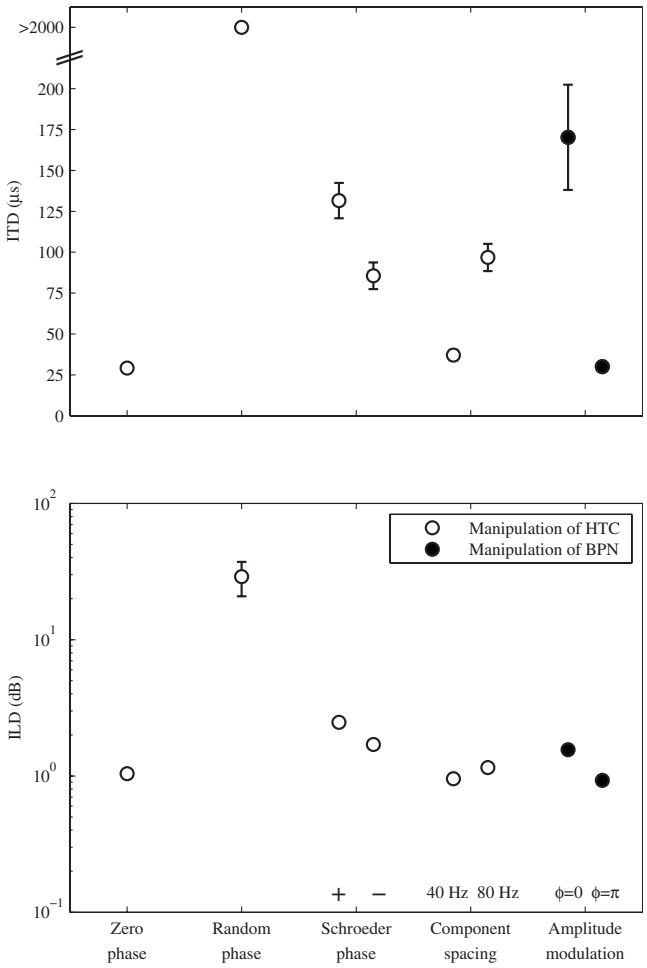


FIG. 8. Mean ITD (top panel) and ILD (bottom panel) thresholds for the temporal envelope structure conditions from Exp. 4. Included for reference is the result for the wideband zero-phase condition from Exp. 1. Error bars represent the standard errors of the mean. Error bars smaller than the symbol size are omitted. Please remember that the total interaural difference cue between the two signals is, in fact, twice the size of these interaural differences.

For the condition with 40 Hz component spacing of the HTC, discrimination based on interaural time differences was, with a threshold of 37 μs, similar to the performance for the reference 20 Hz component spacing condition. For the condition with 80 Hz component spacing, with a threshold of 97 μs, discrimination was seriously degraded. For these two component spacing conditions, discrimination based on interaural level differences was equal to the performance for the zero-phase condition, with thresholds 0.0 and 0.2 dB higher for the 40 and 80 Hz component spacing conditions, respectively. Thus, halving the time between the regular maxima in the temporal envelope of the HTC to 25 ms caused only little change in binaural sensitivity. By again halving the time between regular maxima to 12.5 ms, binaural processing of interaural time differences was made much more difficult. Again, no change in binaural sensitivity to interaural level differences was found, indicating that small opposite interaural level differences for each signal,

combined with different temporal envelopes, sufficed to allow discrimination between the spatial configurations of the two signals.

For the condition with the HTC and the amplitude-modulated BPN presented in phase ($\phi_{AM}=0$), discrimination between the spatial configurations of the two signals based on interaural time differences was heavily degraded, with a threshold of 170 μs. As can be seen in Fig. 8, top panel, the performance in this condition differed substantially across subjects. For the condition with the two signals presented out of phase ($\phi_{AM}=\pi$), discrimination ability was equal to the reference zero-phase condition, with a threshold only 1 μs higher. For these two conditions, discrimination based on interaural level differences was similar to the performance for the zero-phase condition, with thresholds 0.6 dB higher and 0.1 dB lower for the in-phase and out-of-phase conditions, respectively. Thus, when the maxima of both signals coincided temporally, the monaural envelopes were apparently more similar, resulting in a substantially higher threshold for discrimination based on interaural time differences. Small level differences between the signals provided sufficient monaural or binaural information to discriminate between their spatial configurations. When the maxima of the signals alternated temporally, the binaural processing of these signals was comparable to that for the reference zero-phase condition, with comparable performance in discrimination between the spatial configurations of the two signals.

C. Discussion

Because the long-term interaural cross-correlation patterns or long-term patterns after equalization-cancellation for the composite signals in the ITD conditions were indistinguishable between the target and the reference intervals (see Fig. 2), short-term binaural processing must have played a role in discriminating between the spatial configurations of the two signals. One interpretation is that in order to achieve the observed lateralization of the individual signals', the binaural system would require sufficient information about their temporal envelopes to distinguish between them and assign the spatial cues to the correct signal. Then, the monaural temporal envelope cues somehow would have to support the selection of temporal intervals at which the one or the other signal was dominant, and facilitate the organization of temporally varying interaural time differences within the period of the HTC.

In line with this view, it follows that for discrimination between the spatial configurations of two spectrally and temporally overlapping signals, the temporal envelopes of the signals need to be sufficiently different, either in terms of the degree of modulation or the relative timing of the envelope maxima, in order to be able to link the pattern of temporally changing interaural time differences to the envelope maxima of the signals. For the selection of interaural time differences within a composite signal, these interaural differences could be emphasized during the periods when a single source is dominant in one or more auditory filters.

A recent approach to selecting these time instants by analyzing the interaural coherence is the directional cue selection model of Faller and Merimaa (2004). In this model, the time instants at which a single sound is dominating in an acoustic scene are identified as moments where the interaural coherence is high. Here, a perceptually-motivated adaptation of their model, using a gammachirp filterbank (Irino and Patterson, 2006) instead of a gammatone filterbank and peripheral compression stages, is used to determine the short-term cross-correlation patterns for the stimuli used in these experiments.

Figure 9 shows the results for the wideband condition from Exp. 1, in which the HTC had its components in zero starting phase, and the noise signal was unmodulated. The three top panels show the individual signals and the composite signal for one ear, while the fourth panel shows the output of a single auditory filter centered at 550 Hz from the gammachirp filterbank, followed by a hair cell model. The fifth panel shows the cross-correlation pattern between the left- and right-ear signals, computed with an exponentially decaying window with a time constant of 10 ms (Faller and Merimaa, 2004). High correlation values are indicated by darker colors in the figure, and the maximum of the cross-correlation function at each temporal instant is traced by the white curve. The absolute interaural time difference of the signals in this and the following examples was 100 μs, a value much larger than the experimental thresholds observed. The maximum of the cross-correlation pattern varies periodically between the values of ± 100 μs (indicated by the dashed lines) in close synchrony with the monaural envelope of the composite signal, enabling the determination of the individual signal components' lateralizations.

When analyzing the conditions in which ITD discrimination thresholds were successfully obtained, i.e., the contiguous bandwidth conditions of Exp. 3, the various component spacings of the HTC, and the amplitude-modulated BPN of the current experiment, similar observations are made. In all these analyses, the cross-correlation pattern is found to follow the monaural envelope of the composite signal in close synchrony. For example, Fig. 10 shows the temporal envelopes and interaural cross-correlation pattern for the zero-phase HTC and the amplitude-modulated BPN presented with their envelope maxima temporally out of phase.

In contrast to these conditions, Fig. 11 shows the corresponding patterns for the combination of a random-phase HTC and the unmodulated BPN. Changing the temporal envelope structure of the HTC reduced its peak factor and made it similar to the temporal envelope of the BPN, instead of regularly peaked as in the zero-phase condition. Changing the temporal envelope structure also influenced the regular cross-correlation pattern of interaural differences. Here, neither the monaural envelope nor the cross-correlation pattern allow determination of time instants at which one of the two constituent signals is dominant. As a consequence, no information is available for linking specific moments in the internal representation of interaural delay to those of the monaural temporal envelopes, and thus to identify the target and reference intervals based on the available spatial information.

Similar observations can be made for the conditions in which ITD discrimination thresholds could not be measured,

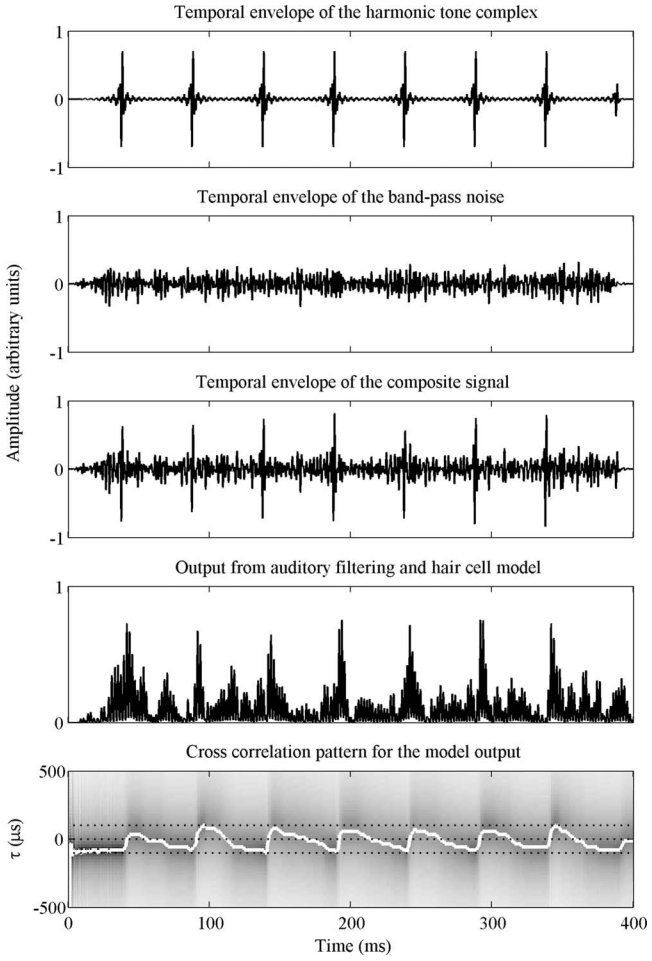


FIG. 9. The relation between interaural cross-correlation pattern and the temporal envelope, for the wideband condition with the components of the HTC in zero phase from Exp. 1. The three top panels show the temporal envelopes of the individual wideband signals and their composite signal. The HTC and the BPN in this example have 100 μs interaural time differences with opposite signs. The composite left- and right-ear signals were used as input to the directional cue selection model of Faller and Merimaa (2004). The two lower panels show the results from the model, i.e., the output of the auditory filter centered at 550 Hz and hair cell model, and the interaural cross-correlation pattern computed from this output, respectively. The dashed lines in the lower panel represent the 0 and $\pm$100 μs interaural time differences.

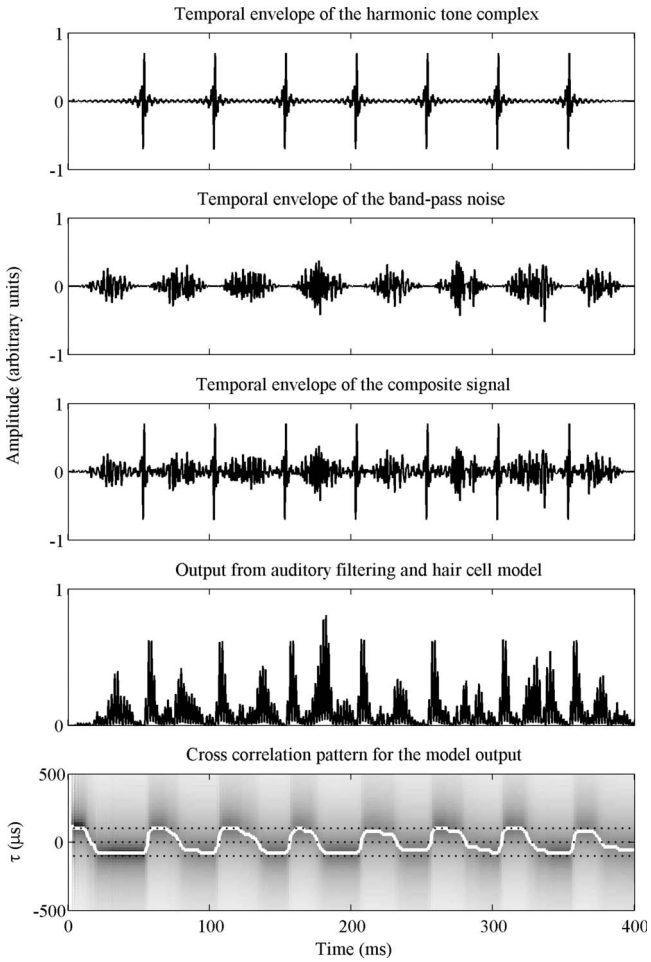


FIG. 10. The relation between interaural cross-correlation pattern and the temporal envelope, for the condition with the zero-phase HTC and the amplitude-modulated BPN presented with their envelope maxima temporally out of phase. Similar to Fig. 9, the panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and $\pm$100 μs interaural time differences.

as, for example, the narrowband condition of 1 ERB centered at 550 Hz from Exp. 1, as shown in Fig. 12. In this narrowband condition, the peaks of the HTC, although regularly spaced in time, are insufficiently peaked to dominate the temporal envelope of the composite signal in the same manner as seen in the wideband condition. As in the random-phase condition, the information from the internal representation of the interaural delay could not be linked to the information from the monaural temporal envelopes.

From the regular short-term interaural cross-correlation pattern, as displayed in the bottom panel of Figs. 9 and 10, it may be argued that identification of the target and reference intervals could have been established by distinguishing between the regularly changing pattern of the cross-correlation function in the target interval and its interaural inverse in the reference intervals. The pattern of dynamically changing interaural time differences itself may then have provided sufficient information to the binaural system to perform the discrimination task. For the random-phase and narrowband conditions, in which no threshold could be obtained, the cross-correlation pattern does not exhibit a systematic asymmetrical pattern, as can be seen in Figs. 11 and 12. Using the asymmetrical cross-correlation patterns that are mirrored for target and reference intervals would require subjects to be able to track fast chances in binaural cues, such as observed in the fifth panel synchronous to the peaks of the auditory filter output in the fourth panel. This requirement, however, contrasts to earlier findings on the inability to track fast changes in binaural cues (Grantham and Wightman, 1978). In combination with the ability to lateralize the individual signals, as reported by the subjects (see discussion of Exp. 1), it is considered unlikely that subjects only used the interaural cross-correlation pattern to identify target and reference intervals. In addition, when the same analysis was applied to the condition with the zero-phase HTC and the amplitude-modulated BPN presented with their envelope maxima temporally in phase, see Fig. 13, the cross-correlation pattern

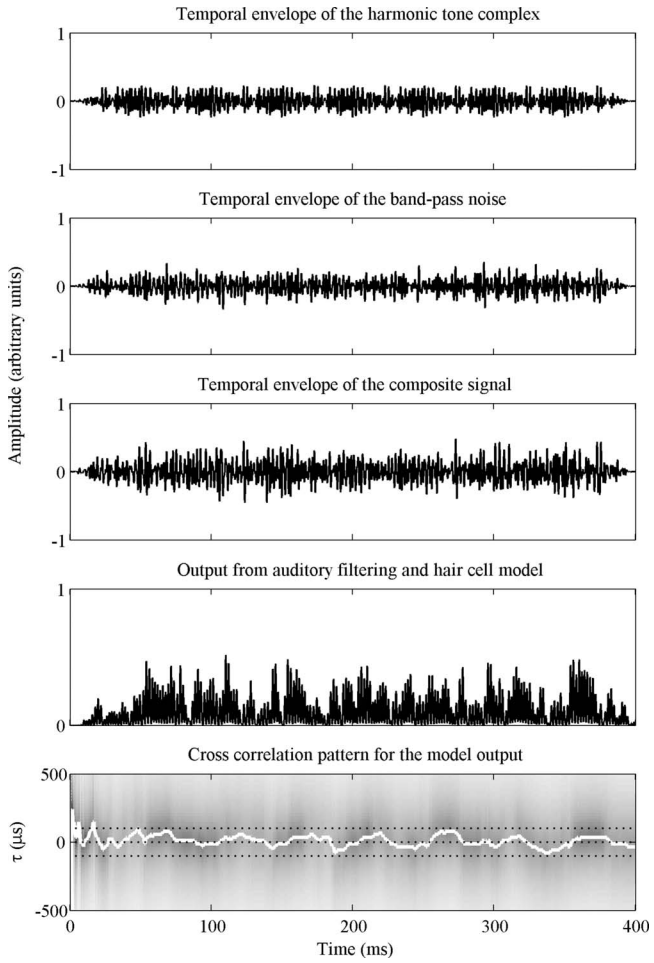


FIG. 11. The relation between interaural cross-correlation pattern and the temporal envelope, for the condition with the components of the HTC in random phase. Again, the panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences.

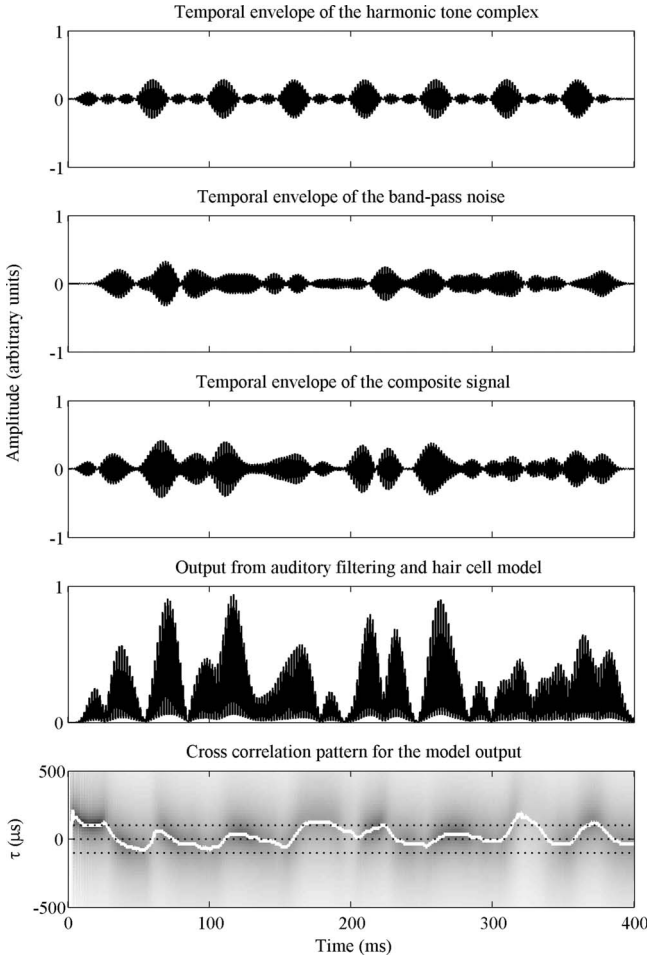


FIG. 12. The relation between interaural cross-correlation pattern and the temporal envelope, for the narrowband condition of 1 ERB centered at 550 Hz from Exp. 1. The panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ±100 μs interaural time differences.

exhibits similar fast changes in the positions of the maxima as displayed in Figs. 9 and 10. If such fast changes, with different orientations for the target and reference intervals, were the dominant cue for identification of these intervals, it is hard to understand why patterns as in Figs. 9 and 10 yield thresholds of about 30 μs, while a pattern as in Fig. 13 the threshold was increased to about 170 μs. From these considerations, it is concluded that synchrony between the temporal envelopes and the pattern of changing interaural time differences seems to be necessary, but not sufficient, to identify target and reference intervals.

The principle of determining a time instants' dominant signal and processing the momentary spatial cues would also explain the increased difficulty to discriminate between the spatial configurations for several ITD conditions in the current experiment. For the Schroeder-phase conditions, the effect of using Schroeder phases for the HTC is twofold. Due to the frequency-dependent group delay, maxima in the temporal envelope are asynchronous across frequency. Therefore, no single moment in time can be defined where the HTC dominates across all frequencies synchronously. Furthermore, positive Schroeder phases compensate for the phase dispersion on the basilar membrane, resulting in a larger peak factor at the output of the basilar membrane, while negative Schroeder phases result in a lower peak factor (Kohlrausch and Sander, 1995). Apparently, synchrony across frequency is not essential, although it influences the ability to discriminate between the target and reference intervals. For the 80 Hz component spacing condition, the reduced temporal interval between peaks may have reduced the peak factor of the temporal envelopes after filtering on the basilar membrane such that dominant signal components were more difficult to be distinguished monaurally. In addition, here the temporal resolution of the binaural system may have been insufficient to accurately follow the faster switching between the opposite binaural cues. For the HTC presented in phase with the amplitude-modulated BPN, the determination of the dominant signal at a certain time instant may have been inhibited by the similarity of the signals' temporal envelopes and the temporal coincidence of both envelope maxima and minima.

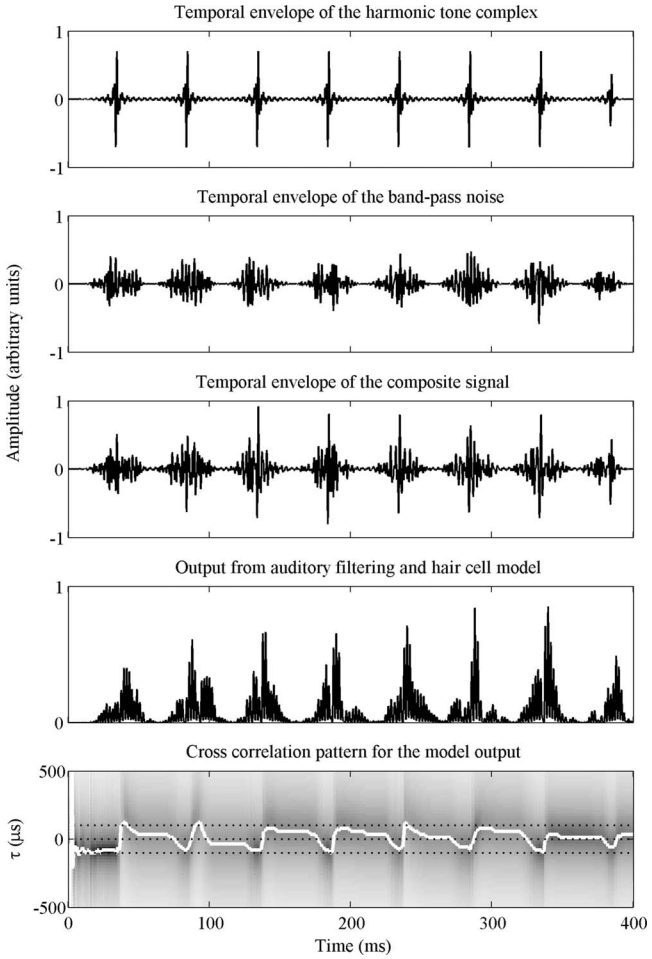


FIG. 13. The relation between interaural cross-correlation pattern and the temporal envelope, for the zero-phase HTC and the amplitude-modulated BPN with their envelope maxima temporally in phase. The panels show the temporal envelopes of the individual signals, their composite signal, the output of the auditory filter and hair cell model, and the interaural cross-correlation pattern of this output. The dashed lines in the lower panel represent the 0 and ± 100 μs interaural time differences. In contrast to the conditions shown in Figs. 9 and 10, subjects had more difficulty to identify target and reference intervals in this condition.

VIII. GENERAL DISCUSSION

In the current study, the ability to segregate two spectrally and temporally overlapping signals based on differences in temporal envelope structure and binaural cues was investigated. As an indication for segregation ability, threshold interaural time and level differences for discriminating between the signals' spatial configurations were measured for various signal bandwidths, center frequencies, and temporal envelopes. The interpretation of these measures is only valid if the target and reference intervals of the three-alternative forced-choice experiment cannot be identified otherwise than by segregating and lateralizing the constituents of the composite signal in each interval. It was shown that the application of a rove on the interaural time difference for the composite signal did not degrade the discrimination of the signals' spatial configurations. It seems unlikely that the results can be explained by the subjects' perception of a change of the overall lateralizations of the composite signal between the target and reference intervals. Alternatively, the identification of the intervals by distinguishing between their mirrored patterns in the short-term cross-correlation functions may explain why for narrowband signals and the random-phase HTC discrimination based on interaural time differences was not possible. The corresponding dynamic patterns of interaural time differences have no systematic asymmetry that would allow discrimination between the intervals. However, it was shown that similar asymmetrical fast changes in the maxima of the cross-correlation function can be found for the ITD conditions that yielded the lowest and the highest threshold in discrimination between the target and reference intervals. In addition, the use of only the mirrored cross-correlation patterns does not match the subjects' report on the use of the lateral position of the HTC for identifying target and reference intervals.

For the ITD conditions, the spatial configurations of the signals could only be discriminated when the spectral energy was in multiple contiguous auditory filters, with an increase in performance with an increase in bandwidth. This bandwidth dependency may be related to across-frequency integration of the coherence between temporal envelopes in the stimulated auditory filters (cf. Trahiotis and Stern, 1994). However, it is unclear how the across-frequency integration could account for the inability to discriminate between the spatial configurations of signals that had their energy in noncontiguous bands. This inability may be attributed to the observed difference in peak factor of contiguous-band versus noncontiguous-band signals (see Fig. 6). For the contiguous-band HTC, the signal components in adjacent auditory filters contributed to the within-channel peak factor, which did not occur for the noncontiguous-band HTC and the noncontiguous-band and contiguous-band BPNs. A similar contribution of signal components in adjacent auditory filters to within-channel perception has also been shown in the context of comodulation masking release (Verhey *et al.*, 1999).

The obtained results show that the differences in temporal envelope structures of the two signals influence the discriminability between their spatial configurations. The results are consistent with the hypothesis that the analysis of the monaural information from the temporal envelopes at each ear supports the attribution of binaural information to signal components. For the ITD conditions, discrimination of the signals' spatial configurations was found to be inhibited for conditions in which their temporal envelopes were similar or the maxima and minima of the temporal envelopes coincided. For instance, for the narrowband or noncontiguous-band signal conditions, the temporal envelopes of the two signals were much more similar and discrimination was not possible. Also, for the wideband random-phase HTC and the in-phase amplitude-modulated BPN conditions, it was more difficult to distinguish between the temporal envelopes of the two signals, resulting in either the complete inability or the increased difficulty to discriminate between the signals' spatial configurations. For the contiguous wideband signal conditions, the HTC had a considerably higher peak factor than the BPN and discrimination was possible. This reasoning may explain the lack of effect from interaural time differences as a grouping cue in the experiments of Culling and Summerfield (1995). By using various combinations of fil-

tered bands of continuous noise, their vowel-like stimuli may have been composed in such a way that the temporal envelopes of the constituent signals were too similar to facilitate attribution of the binaural information to either of them.

For the ILD conditions, the signals' spatial configurations could be discriminated within a single auditory filter, as long as the phase spectrum of the HTC was not random. The left and right ears received different relative levels of the HTC and the BPN. Therefore, unlike the ITD conditions, the envelope was more peaked in the ear where the HTC was louder. Listening to only the left or right ear, in principle, provided sufficient cues to discriminate between the signals' spatial configurations. For the narrowband conditions, the performance based on monaural listening was even better than the one based on binaural listening. This finding suggests that the processing of binaural information from each left-right pair of auditory filters interfers with the processing of monaural information from the individual auditory filters. For the wideband conditions, performance based only on either monaural or binaural listening was found to be very similar. Therefore, it is not possible to distinguish between monaural or binaural processing on the basis of these results.

In general, differences in the temporal envelope of signals seem to improve the use of their binaural cues for segregation. These differences even may have allowed monaural segregation of the composite signal's components by perceiving continuity in temporal envelope features that are "glimpsed" at momentary differences in level (cf. Noble and Perrett, 2002; Shinn-Cunningham, 2005; Cooke, 2006). To establish such segregation, the auditory system may adopt simultaneously several hypotheses about how the monaural input signals can be segregated based on monaural temporal envelope features. The binaural cues corresponding to each of the envelope features would then assist in selecting the most likely hypothesis, based on the assumption that within one auditory stream the binaural cues have to be coherent across time. Also some top-down processing may be assumed in this context, as prior knowledge about a certain signal would enhance its identification. Thus, for the selection of binaural cues within composite signals, interaural time and level differences are best considered only at time instants at which the sound of a single source is dominant in one or more auditory filters. In the directional cue selection model of Faller and Merimaa (2004), these time instants are identified by analyzing the interaural coherence, and correspond well to the moments in time at which glimpsing would be operative.

In a previous study (van de Par *et al.*, 2005), the discrimination between the same HTC and BPN as used in the current study was investigated, when both were presented interaurally out of phase within an in-phase noise masker. The results showed an ability to discriminate between the two signals at signal-to-masker levels well below monaural detection thresholds. This finding suggests that within the binaural display, i.e., the internal representation of binaural cues, information is available about the temporal structure of the out-of-phase signal. An equalization-cancellation stage (Durlach, 1963; Breebaart *et al.*, 2001) could, in principle, provide such information, because it would cancel the noise masker and not the out-of-phase signal. It would require, however, that within the binaural output, the capacity to process temporal information is sufficiently good to distinguish between the signals.

In the current study, an equalization-cancellation stage could remove one of the two signals in a similar way, allowing for the temporal processing of the other signal. However, the sharp increase in threshold (from 29 via 37 to 97 μs) from decreasing the period of the HTC (from 50 via 25 to 12.5 ms) indicates the vicinity of an upper limit for the temporal resolution at the output of a binaural display for the current experiments. From the current results, it is not evident how equalization-cancellation processing could explain that subjects had such difficulty to identify the target and reference intervals in the condition with the zero-phase HTC and the amplitude-modulated BPN presented in phase.

The glimpsing of signal properties as described above could provide an alternative explanation for the reduced ability to discriminate between the spatial configurations of the zero-phase HTC and the amplitude-modulated BPN presented in phase. For this condition, the peaks in the envelopes of both signals coincide, and the level difference between the maxima of the signals is smaller than for the nonmodulated condition due to maintaining the same overall level for the amplitude-modulated BPN. This made the temporal envelopes similar, and may have facilitated perceptual merging of the two signals into one auditory object. Because of the "discontinuity" of the BPN in between the regular peaks of the composite signal, there was simply less evidence for the presence of a second auditory object with a different lateralization. Similarly, for the random-phase condition there is no evidence in the monaural envelope of the composite signal that allows glimpsing of the constituent signals. This explanation supports the idea that monaural analysis of temporal envelopes on perceived continuity of the glimpsed signal components is required before binaural processing can take place.

IX. CONCLUSION

The ability to segregate two spectrally and temporally overlapping signals based on differences in temporal envelope structure and binaural cues was investigated. As an indication for segregation ability, threshold interaural time and level differences were measured for discrimination between the spatial configurations of the two signals. Discrimination based on interaural level differences was good for all conditions, although absolute thresholds depended on signal bandwidth and center frequency. Discrimination based on interaural time differences depended on the signals' temporal envelope structures.

The HTC and BPN were presented with interaural differences of the same absolute value, but with opposite signs, to yield lateralization to different sides of the medial plane. This way, the composite signal's long-term interaural cross-correlation patterns or the long-term patterns after equalization-cancellation were indistinguishable between the three intervals of the forced-choice procedure, and could not facilitate identification of target and reference intervals. For

successful identification of the intervals in the ITD conditions, the binaural system must have been capable of processing changes in interaural time differences within the period of the HTC. Such processing would require short-term analysis of the binaural cues present in the composite signal, and association of the specific pattern of time-varying binaural cues with the target and reference intervals.

The reported lateralization of the individual signals and the obtained experimental results support the idea that the binaural system uses the short-term monaural information, that is glimpsed from the temporal envelopes at each ear, to process the binaural information of the underlying signal component. This processing facilitates segregation and lateralization of a composite signal's constituent elements. The current findings suggest that monaural information from the temporal envelopes influences the use of binaural information in the perceptual organization of signal components.

ACKNOWLEDGMENTS

We would like to thank Barbara Shinn-Cunningham, Aki Härmä, Nicolas LeGoff, the associate editor Ruth Litovsky, Steve Colburn, and two anonymous reviewers for their valuable comments on the earlier versions of this paper.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model setup," J. Acoust. Soc. Am. **110**, 1074–1088.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S., Ahad, P., Kim, J., and Melnerich, L. (**1994**). "Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or harmonics in a complex tone," Percept. Psychophys. **56**, 155–162.

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J., and Hukin, R. W. (**1997**). "Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity," J. Acoust. Soc. Am. **102**, 2316–2324.

Darwin, C. J., and Hukin, R. W. (**1998**). "Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony," J. Acoust. Soc. Am. **103**, 1080–1084.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089.

Gallun, F. J., Mason, C. R., and Kidd, G., Jr., (**2007**). "The ability to listen with independent ears," J. Acoust. Soc. Am. **122**, 2814–2825.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Heller, L. M., and Trahiotis, C. (**1995**). "The discrimination of samples of noise in monotic, diotic, and dichotic conditions," J. Acoust. Soc. Am. **97**, 3775–3781.

Hukin, R. W., and Darwin, C. J. (**1995**). "Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel," J. Acoust. Soc. Am. **98**, 1380–1387.

Irino, T., and Patterson, R. D. (**2006**). "A dynamic compressive gammachirp auditory filterbank," IEEE Trans. Audio, Speech, Lang. Process. **14**, 2222–2232.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1336.

Schroeder, M. R. (**1970**). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.

Shackleton, T. M., and Meddis, R. (**1992**). "The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs," J. Acoust. Soc. Am. **91**, 3579–3581.

Shinn-Cunningham, B. G. (**2005**). "Influences of spatial cues on grouping and understanding sound," Proceedings of Forum Acusticum, pp. 1539–1544.

Stellmack, M. A., and Lutfi, R. A. (**1996**). "Observer weighting of concurrent binaural information," J. Acoust. Soc. Am. **99**, 579–587.

Stern, R. M., Trahiotis, C., and Ripepi, A. M. (**2006**). "Fluctuations in amplitude and frequency enable interaural delays to foster the identification of speech-like stimuli," in *Dynamics of Speech Production and Perception*, edited by P. Divenyi, S. Greenberg, and G. Meyer (IOS, Amsterdam), pp. 143–151.

Trahiotis, C., and Stern, R. M. (**1994**). "Across-frequency interaction in lateralization of complex binaural stimuli," J. Acoust. Soc. Am. **96**, 3804–3806.

van de Par, S., Kohlrausch, A., Breebaart, J., and McKinney, M. (**2005**). "Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 398–404.

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.

Tuning in the spatial dimension: Evidence from a masked speech identification task

Nicole Marrone,[a] Christine R. Mason, and Gerald Kidd, Jr.
Department of Speech, Language, and Hearing Sciences and the Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 28 February 2007; revised 23 May 2008; accepted 28 May 2008)

Spatial release from masking was studied in a three-talker soundfield listening experiment. The target talker was presented at 0° azimuth and the maskers were either colocated or symmetrically positioned around the target, with a different masker talker on each side. The symmetric placement greatly reduced any "better ear" listening advantage. When the maskers were separated from the target by ±15°, the average spatial release from masking was 8 dB. Wider separations increased the release to more than 12 dB. This large effect was eliminated when binaural cues and perceived spatial separation were degraded by covering one ear with an earplug and earmuff. Increasing reverberation in the room increased the target-to-masker ratio (T/M) for the separated, but not colocated, conditions reducing the release from masking, although a significant advantage of spatial separation remained. Time reversing the masker speech improved performance in both the colocated and spatially separated cases but lowered T/M the most for the colocated condition, also resulting in a reduction in the spatial release from masking. Overall, the spatial tuning observed appears to depend on the presence of interaural differences that improve the perceptual segregation of sources and facilitate the focus of attention at a point in space. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945710]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Lj [RLF] Pages: 1146–1158

I. INTRODUCTION

There are a number of examples in the auditory system of selective responses along a simple stimulus dimension. Perhaps the most obvious and best understood example is stimulus frequency where the tuned responses of the peripheral transduction mechanism have been thoroughly mapped out and examined. Beyond peripheral filtering, however, are instances in which the actions of higher level processes lead to performance that reveals an enhanced degree of selectivity. Greenberg and Larkin (1968), for example, used the probe-signal method to demonstrate tuning in the frequency domain that was not attributable to peripheral filtering but rather to the focus of attention at an expected signal frequency. Similar findings revealing the role of expectation and selectivity have been reported for other dimensions, such as duration (e.g., Wright and Dai, 1994), spectral shape (Hill *et al.*, 1998), and modulation frequency (e.g., Wright and Dai, 1998).

The behavioral evidence for tuning along the spatial (azimuthal) dimension is less compelling and surprisingly limited (cf. Scharf, 1998). Clearly, certain neurons in the brainstem or cortex exhibit selective responses to the primary binaural cues of interaural time and level differences (e.g., Goldberg and Brown, 1968; Middlebrooks and Pettigrew, 1981; Yin and Kuwada, 1983; Tsuchitani, 1988; Yin and Chan, 1990; Sterbing *et al.*, 2003; Stecker *et al.*, 2003; King *et al.*, 2007) so that, akin to the example of frequency noted previously, there is a basis in auditory physiology for expecting tuned responses to spatial location. Further, there is electrophysiological evidence suggesting that higher-level processes affect selectivity in azimuth. Teder-Salejarvi and Hillyard (1998) and Teder-Salejarvi *et al.* (1999) have demonstrated changes in event-related potentials (ERPs) in humans based on attended versus unattended locations of a sound stimulus. When a test stimulus was presented at an attended location, larger ERPs were obtained than when presented at unattended locations. Filter bandwidths inferred from their data were often sharply tuned, in some cases less than 5°. Psychophysical evidence for the important role of attentional focus in location, based on presentation of speech stimuli at likely versus unlikely locations, has also been presented recently by Kidd *et al.* (2005b) and Brungart *et al.* (2006).

Studies of masking in sound fields also provide examples of a selective response to spatial location. More masking occurs when target and masker(s) are colocated than when they are separated in azimuth. This selective response, a progressive reduction in masking as source separation increases, could be interpreted as evidence for tuned spatial channels analogous to frequency channels. However, the release from masking that occurs due to spatial separation of sound sources is a complex phenomenon comprising both lower-level and higher-level components (e.g., Kidd *et al.*, 1998; Freyman *et al.*, 1999; Arbogast *et al.*, 2002; Drennan *et al.*, 2003, 2007; Hawley *et al.*, 2004; Best *et al.*, 2005, 2006) and it is not always apparent which factor drives release from masking. Thus, the mechanism(s) responsible for the apparent tuned response is unclear and may be different in different conditions. The main lower-level components that are thought to provide a basis for spatial release from

[a]Electronic mail: n-marrone@northwestern.edu

masking are the "better ear advantage" [attending to the ear with the more favorable target-to-masker ratio (T/M) caused by the acoustic filtering of the head][1] and "binaural analysis" [defined here as within-channel improvement in T/M due to a masking-level-difference (MLD) type of mechanism, where the terms within and across channel refer to frequency channels]. The relevant higher-level components are less clearly understood but are thought to reflect (at least) the combined actions of perceptual segregation of sources and the focus of attention at a point in space (cf. Kidd *et al.*, 2005b). The type of masking that is present may have a profound effect on the pattern of results as source separation is varied. For example, Kidd *et al.* (1998) found small (less than 5 dB, on average, for mid- and low-frequency targets) amounts of spatial release from masking in a nonspeech pattern identification task for a noise masker producing primarily energetic masking. For the same listeners, targets, and task the spatial release from a primarily informational masker (sequences of random-frequency tone complexes producing little if any energetic masking) was, on average, as much as 20 dB at low and high target frequencies. The Kidd *et al.* (1998) results revealed a progressive and pronounced decline in T/M for the informational masking condition as spatial separation of target and masker varied from colocated to 180°. However, although the results for both energetic and informational maskers were consistent with a tuned response (although of very different magnitudes), and both clearly depended on the binaural cues of interaural time (ITD) and level (ILD) differences, the underlying mechanisms were thought to be fundamentally different. For the energetic masker, by far the largest spatial release occurred for the high frequency target and followed the published values for head shadow (e.g., Shaw, 1974) fairly closely after taking into account the mildly reverberant room. Predictions for spatial release from energetic masking for speech in noise are based on the same two factors and are generally successful in accounting for the empirical results (cf. Zurek, 1993; Bronkhorst, 2000). For the informational masker used by Kidd *et al.* (1998), head shadow may have been one factor, although the across-channel nature of the masking complicates the interpretation of the effect. Binaural analysis almost certainly was not a factor because the masking was not within-channel masking leading to the conclusion that higher-level factors were primarily responsible.

Another example of a spatially tuned response obtained perceptually that appeared to be due largely to higher level processes was reported by Arbogast and Kidd (2000). They used a variant on the probe-signal method to examine tuning in azimuth. The task was to identify the upward or downward trajectory of a sequence of tone pulses presented through a loudspeaker. Sequences of tonal masking sounds were presented from several other loudspeakers roughly concurrently with the target. The target was most likely to be presented at one location but on a small proportion of the trials it was randomly presented from other locations. The masker frequencies were randomized and remote from the target, limiting energetic masking while emphasizing informational masking. Both accuracy and response times were better at the more likely target location than at less likely locations, which Arbogast and Kidd (2000) interpreted as evidence for spatial filtering. The filter-like responses they observed appeared to be fairly sharply tuned although the effects were relatively small and the attenuation characteristics of the "filter" could not be accurately determined.

Thus, there is some psychophysical evidence that suggests tuning in azimuth. However, the evidence is incomplete and in some cases inconsistent. For example, in contrast to the (apparently) sharply tuned pattern of masked responses reported by Arbogast and Kidd (2000), Boehnke and Phillips (1999) found evidence indicating much broader perceptual tuning based on the results of a gap detection task. Based on their results, they proposed two broadly tuned, overlapping spatial channels located in the left and right auditory hemifields. However, that interpretation also has been questioned (Oxenham, 2000). Other studies using different paradigms have also suggested the presence of spatial channels or filters. For example, Carlile *et al.* (2001) used a procedure in which localization judgments following exposure to noise at one location were altered in a manner consistent with the presence of a set of spatially arranged (in azimuth) filters.

For the masking results, at least, the role of head shadow, in which the magnitude of the acoustic effect varies with spatial position and frequency, complicates determining how the other factors contribute to spatial tuning. One approach to minimizing the role of head shadow is the symmetric placement of two maskers around a target. This approach has been used by several investigators (e.g., Helfer, 1992; Peissig and Kollmeier, 1997; Bronkhorst and Plomp, 1992; Noble and Perrett, 2002; Li *et al.*, 2004) to examine spatial release from masking that occurs beyond any acoustical "better ear listening" effects. However, only the study by Noble and Perrett (2002) provides an indication of spatial tuning when the target is one talker and the maskers are other independent talkers, symmetrically placed around the target. They tested two spatial separations, ±30° and ±90° (in addition to colocated) and found that the relatively small amount of spatial release (approximately 4 dB) for the ±30° condition increased by about 1 dB for the ±90° condition. Somewhat like the Arbogast and Kidd (2000) study, it is possible to conclude that the selectivity of a perceptual filter was sharper than the narrowest spacing tested but the precise properties of the filter could not be determined.

In recent years, interest has turned to speech-on-speech masking, in part because it represents a common everyday listening situation, but also because higher level processes seem to be more of a factor in determining spatial release. In two studies (Arbogast *et al.*, 2002; Kidd *et al.*, 2005a), much larger amounts of spatial release were found for speech-on-speech masking than in noise-masked control conditions producing primarily energetic masking. Further, in the Kidd *et al.* (2005a) study, there was also a differential effect of reverberation depending on masker type. Increased reverberation increased the T/M at threshold for both colocated and spatially separated conditions, with a large amount of spatial release still apparent. Other studies, however, (cf. Culling *et al.*, 2003) have reported that increased reverbera-

tion completely eliminates spatial release for speech. Thus, currently, the role of reverberation in spatial release from speech-on-speech masking is uncertain.

In the present study our goal was to examine spatial release from masking under conditions producing different amounts of energetic and informational masking of speech and in rooms with varying degrees of reverberation. The energetic/informational masking distinction was varied by using different types of masking stimuli and reverberation was varied by changing the sound absorption characteristics of the listening environment. This study attempts to answer the question of whether or not a filter-like function can be measured in these situations when the listener's task is to selectively attend to one talker at a particular location and the interfering sources are progressively separated (symmetrically) from the target.

II. METHODS

A. Listeners

Six normal hearing adult volunteers (5 female, 1 male) between 25 and 43 years of age participated in this study. The listeners had audiometric thresholds of 20 dB HL(hearing level re: ANSI 3.6–2004) or better in each ear for octave frequencies from 250 to 8000 Hz. Listeners participated in the experiment in four to six sessions that were approximately two hours each (including short breaks) and were paid for their participation. One additional listener with an anacusic ear after the auditory nerve was severed surgically was tested on a subset of the conditions (see Sec. II D 2).

B. Stimuli

The four female talkers from the Coordinate Response Measure (CRM) speech identification test (Bolia *et al.*, 2000) were used for the target and masker sentences. Each sentence in this corpus has the following structure: "Ready [callsign] go to [color] [number] now." The corpus contains all possible combinations of eight callsigns, four colors, and eight numbers. On each trial, the listener heard three sentences spoken by three different randomly chosen talkers. Each sentence had a different callsign, color, and number. One of the sentences was the target, denoted by the callsign "Baron." The target and masker talkers varied from trial to trial. Once the target sentence was selected, the two masker sentences were then chosen without replacement from the remaining set of possible talkers, callsigns, colors, and numbers.

In addition, a subset of listeners was tested in a condition where the speech of the two masker talkers was temporally reversed (see Sec. III D). The purpose of this manipulation was to test a condition under which the masker had spectrotemporal properties similar to natural speech (and therefore produces about the same amount of energetic masking) but was expected to produce less informational masking because it was not intelligible.

C. Room characteristics

The experiment was conducted in a single-walled Industrial Acoustics Company (IAC) sound booth (12 ft 4 in. long, 13 ft wide, and 7 ft 6 in. high) with stimuli presented through loudspeakers (Acoustic Research 215 PS). This room was designed to allow for changing the reverberation characteristics with panels of different acoustic reflectivity, such as acoustic foam or Plexiglas®. Two room conditions were used for this experiment. In one low-reverberation condition, the surfaces were untreated and had the typical perforated metal surface that is standard in IAC booths and a carpeted floor (the "BARE" room). In the second high-reverberation condition, all surfaces were covered with Plexiglas® panels to increase the acoustic reflections in the room (the "PLEX" room). Various measurements of the room acoustics for these conditions were described by Kidd *et al.* (2005a), documenting the increase in reverberation as a result of adding Plexiglas® panels to the room. For example, there was a greater than 7 dB decrease in the direct to reverberant energy ratio (from 6.3 to −0.9 dB for BARE versus PLEX, respectively) as measured at the approximate position of the listener's head 5 ft from the loudspeaker. Also, the reverberation time (as measured using pulse trains with the loudspeakers in the configuration used in the present study) increased by approximately a factor of 4 (from 0.06 s to just over 0.25 s when the surfaces were covered with Plexiglas®.)

D. Procedures

1. General

Listeners were seated in the sound booth in the center of an array of seven loudspeakers arranged in a semicircle with a 5 ft radius in the horizontal plane at a height approximately level with the listener's ears. The loudspeakers were located at 0° (directly in front of the listener), ±15°, ±45°, and ±90° (directly to the right and left of the listener).

The computer used to control the experiment was located outside the booth along with the Tucker-Davis Technologies (TDT) hardware used to present the stimuli. Stimuli were played at a 40-kHz rate via a 16-bit, 8-channel digital-to-analog converter (DA3-8), low-pass filtered at 20 kHz (FT-6), and attenuated (PA-4). The target sentence was routed through a programmable switch (SS-1). On trials when the target and maskers were colocated (0° spatial separation), the two masker sentences were digitally added, then routed through separate digital-to-analog converter channels, filters, and attenuators before being combined in a mixer (SM3), passed through a power amplifier (Tascam), and routed to the loudspeaker. On trials when the maskers were spatially separated from the target, each sentence was routed through separate digital-to-analog converter channels, filters, attenuators, and power amplifiers (Tascam), and played through separate loudspeakers.

The system was calibrated before each session so that the loudspeakers were correctly positioned (5 ft from listener at head height) and the output level measured with a Brüel &Kjær microphone suspended in that position for a given input was verified and the same from each loudspeaker. For a flat-spectrum Gaussian noise of the same level at the input to the loudspeakers measured at the position of the listener, the

overall SPL was approximately 3 dB higher in the more reverberant room (PLEX). This correction was made when the results are reported in dB SPL.

The task was 1-interval 4×8-alternative forced-choice (four colors: red, white, blue, and green and the numbers 1–8) in which the listeners were asked to identify the color and number from the sentence with the callsign Baron. Listeners were instructed to keep their head facing forward (toward the target loudspeaker at 0° azimuth) but were not restrained. Responses were entered on a handheld keypad with a liquid crystal display (Q-term-II). The word "Listen" appeared on the display at the beginning of each trial. After stimulus presentation, listeners responded to the prompts, "Color [B R W G]?" and "Number [1–8]?" on the keypad. For a response to be scored as correct the listener had to identify both the color and number accurately. Feedback was given on each trial (e.g., "Correct, it was red six"). To familiarize listeners with these procedures, they initially completed two 30-trial blocks with sentences presented in quiet at 60 dB SPL. Every listener had 100% correct speech identification in that condition.

Threshold for target identification was measured in quiet (no maskers) in each room condition using a one-up, one-down adaptive procedure to estimate the 50% correct point on the psychometric function (Levitt, 1971). Each track had a minimum of 30 trials and 9 reversals (typically many more than 9 were obtained) to estimate threshold. The threshold estimate was computed after discarding the first three or four reversals (whichever produced an even number) and thus was based on at least the last six reversals. The initial step size was 4 dB and was reduced to 2 dB after three reversals. Two estimates of threshold in quiet were measured and averaged. If the threshold estimates were more than 5 dB different from one another, an additional two estimates were collected and used in the average.

In all other blocks of trials, the target was presented simultaneously with two maskers. The target level was fixed at 60 dB SPL and the masker level was varied adaptively using the same procedure as for quiet threshold measurements. The two maskers always had the same rms level. At the beginning of each adaptive track, the target was clearly audible above the maskers (+20 dB T/M). The masker level was varied adaptively in 4 dB steps initially and then in 2 dB steps following the third reversal. Each track had a minimum of 30 trials and at least 9 reversals (typically many more than 9 were obtained). The threshold estimate was computed after discarding the first three or four reversals (whichever produced an even number) and thus was based on at least the last six reversals. Threshold estimates were averaged over eight adaptive tracks per condition. Trials were blocked by spatial configuration of the maskers. The maskers were either colocated (0°), near ($\pm 15°$), intermediate ($\pm 45°$), or far ($\pm 90°$). The blocks were presented in random order, such that the masker location changed after every two adaptive tracks. To facilitate comparisons across conditions, masked thresholds will all be expressed as T/M's in decibels. The T/M is calculated as the fixed target level minus the level of the individual maskers at adaptive threshold.

2. "Monaural" control condition

Although there is no acoustically better-ear when the masker talkers are symmetric about the target talker based on the long-term rms value, there is the possibility that the head shadowed representations in the two ears are different enough in the spatially separated cases from the colocated case to provide some benefit of spatial separation. For example, in the $\pm 90°$ case because each masker talker is essentially low-pass filtered by the head before it is received by the far ear, the sum of the two maskers will be frequency-dependent and different from what the spectra are for the colocated case. Thus, it is conceivable that the benefit of spatial separation could be due to monaural cues. In addition, there is presumably a complicated pattern of interaural differences that would lead to moments of improved T/M at each ear. To test the possibility that such information could be useful in performing the task, listeners wore commercially available hearing protectors (an earplug and earmuff) on one ear (left) and repeated a subset of the spatial conditions. This will be referred to as the monaural condition even though it is not strictly monaural listening, but a means of reducing binaural cues (and controlling for the potential acoustic benefit of spatial separation). All six listeners were tested with an earplug and earmuff in a subset of spatial configurations (0° and $\pm 90°$). To compare to true monaural listening, one additional listener with an anacusic ear was tested.

The hearing protectors used were disposable E-A-R® plugs and the AOSafety® Economy Earmuff, both manufactured by the Aearo Company. In order to estimate the amount of attenuation obtained, listeners wore earplugs in both ears and both earmuffs while speech threshold estimates were obtained in quiet. If they did not achieve at least 35 dB of attenuation (± 5 dB) relative to their unoccluded speech thresholds, the earplugs were reinserted, the earmuffs were repositioned, and new threshold estimates were obtained. The earplug and earmuff from the right ear were then removed from the headband, which had been modified so that it could be positioned comfortably on the listener's head. The monaural earmuff was held in place by the tightness of the headband.

3. Time-reversed speech maskers

The speech stimuli and procedures used in the current experiment were chosen to emphasize informational masking. Time-reversed speech maskers were also tested to determine the effect of decreasing the amount of informational masking. Time-reversed speech maskers may be considered less effective informational maskers than time-forward speech because of their lack of meaning, but preserve the spectrotemporal complexity of natural speech. This point is discussed in more detail in later sections and in the Appendix. Five of the six listeners were tested with time-reversed speech maskers for a subset of the spatial configurations (masker locations at 0° and $\pm 90°$). The masker sentences were selected from trial to trial in the same manner as for the forward speech but were played backwards. The procedures for these conditions were otherwise identical to those described earlier.

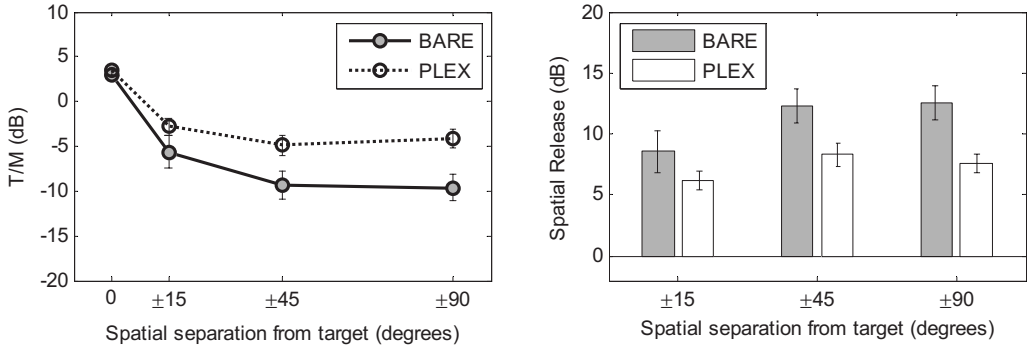


FIG. 1. (Left) Group mean target-to-masker ratios (T/M; expressed in decibels and computed relative to the level of individual maskers) at masked threshold as a function of the horizontal separation of two masker talkers for the BARE (filled circles) and PLEX (open circles) room conditions. The BARE room is a large IAC booth with 0.06 s reverberation time. The addition of Plexiglas® panels increases the reverberation time by a factor of 4 to 0.25 s in the PLEX room. The error bars are ±1 standard error of the mean. (Right) Average spatial release from masking (dB), for each listener, the difference in T/M at threshold between colocated and separated conditions, at three masker separations.

III. RESULTS

The unmasked adaptive thresholds (50% correct points) were within two decibels of one another in the two room conditions. The mean threshold in quiet for target speech at 0° azimuth was 13.5 dB SPL (with a standard error, SE, of 1.3 dB) in the room condition with low-reverberation and 15.5 dB SPL (SE=1.2 dB) in the more reverberant room (after correcting for the 3 dB difference in overall SPL).

The left-hand panel of Fig. 1 displays the group mean thresholds (and ±1 standard error about the mean) for target color and number identification in the presence of two competing talkers as a function of the angular separation between target and maskers. It illustrates the main effects of both spatial separation and reverberation. The abscissa is the degree of separation between the target and maskers in azimuth, from 0° (no separation, or colocated) to ±90°. The ordinate is T/M at threshold in decibels. There are two functions displayed: one for the results obtained in the low-reverberation room (BARE: filled circles connected by solid lines) and the other for the results obtained in the room with more reverberation (PLEX: open circles connected by dotted lines). A repeated-measures analysis of variance (ANOVA) confirmed that there were significant main effects of spatial separation [$F(3,15)=70.1$, $p<0.001$] and room reverberation [$F(1,5)=33.6$, $p=0.002$] on T/M at threshold. Post-hoc analyses (pairwise comparisons) within each room condition indicated that the spatial separations were significantly different from one another ($p<0.05$) with the exception of ±45° versus ±90°. There was also a significant interaction between spatial separation and reverberation [$F(3,15)=11.5$, $p<0.001$] confirming the result that is apparent from the nonparallel functions plotted in the left-hand panel of Fig. 1.

The highest average T/M's were found for the colocated case, regardless of room condition. When the target and masker talkers were colocated, there was essentially no difference between the T/M at threshold in the two room conditions (3.0 dB, SE=0.3 dB for BARE and 3.4 dB, SE =0.4 dB for PLEX). Also, thresholds in the colocated condition were consistent across listeners in both rooms as evident from the small error bars.

Overall, thresholds decreased as the amount of spatial separation between the target and maskers increased for both rooms. In the low-reverberation room (BARE), when the maskers were spatially separated from the target by ±15°, thresholds decreased to −5.7 dB (SE=1.8 dB). Thresholds decreased further with greater spatial separation, but the average results were essentially the same at ±45° and ±90° [thresholds of −9.3 dB (SE=1.6) and −9.6 dB (SE=1.5), respectively]. The same pattern was true in the more reverberant room (PLEX), although the decrease in thresholds was less pronounced. At ±15°, thresholds decreased to −2.7 dB (SE=1 dB). For the greater spatial separations, thresholds decreased further and the values were again comparable to one another [−4.9 dB (SE=1.2) at ±45° and −4.2 dB (SE =1.0) at ±90°]. Individual differences were noted in the maximum decrease in threshold with increasing spatial separation, although the overall pattern of results was consistent across all listeners (see Fig. 2) with most of the benefit occurring in the first ±15° separation.

A. Spatial release from masking

These results can also be evaluated in terms of the amount of spatial release from masking. For each listener this value was calculated as the T/M at threshold for the colocated condition minus the T/M at threshold in the spatially separated conditions. The results of this computation are displayed in the right-hand panel of Fig. 1. The groups of bars are for the 3 spatial separations while the ordinate is the amount of spatial release from masking in decibels. The values shown are group mean differences and standard errors. In the low-reverberation room (BARE), there was 8.6 dB of spatial release from masking (SE=1.7 dB) when the two masker talkers were presented from ±15°. There was nearly 4 dB additional benefit of moving the maskers to either of the wider spatial separations [total spatial release of 12.3 dB (SE=1.4) at ±45° and 12.6 dB (SE=1.4) at ±90°]. In the more reverberant room, there was approximately 2 dB less release from masking when the maskers were at ±15° [PLEX: 6.2 dB (SE=0.8)]. The average amount of spatial release across listeners in the more reverberant room at ±45°

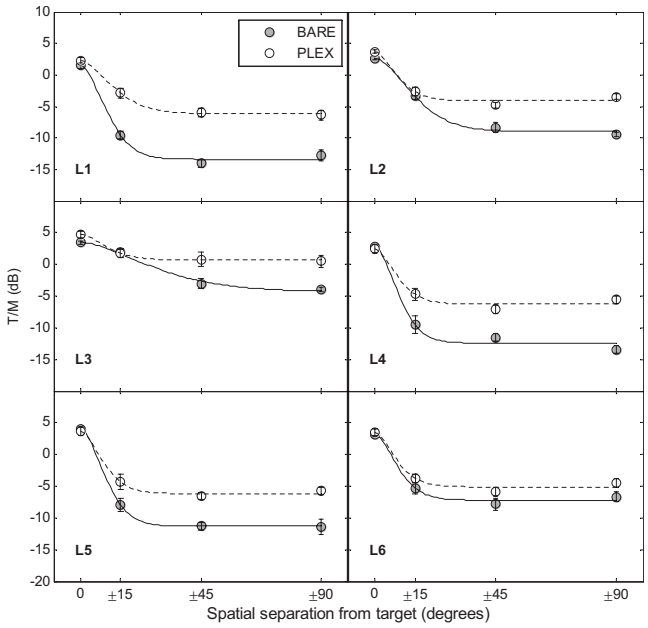


FIG. 2. Individual results plotted for the same conditions as shown in Fig. 1. Error bars are ±1 standard error of the mean. The lines are best-fitting filter functions (see the text and Table I).

and ±90° was still substantial [8.3 dB (SE=0.9) and 7.6 dB (SE=0.7), respectively], although it was less than in the low-reverberation room.

B. Spatial tuning

A useful way of summarizing the relationship between spatial separation of the symmetrically placed maskers and the reduction in masking is to compute best-fitting filter functions based on the masked results. The interpretation of these filter functions as evidence of spatial tuning is considered in the Discussion. The form of the filter applied to the data was the familiar roex(p,r) filter commonly used to characterize auditory filters in the frequency domain (e.g., Patterson *et al.*, 1982; Glasberg and Moore, 1986). This filter was chosen both for computational convenience and because it provides estimates of not only the bandwidth but also the range of the filter (maximum amount of "attenuation"). However, this choice was not meant to suggest that it is the "correct" filter shape for the processes involved. The filters were computed on the results from individual listeners—partly to highlight the differences among listeners—and are shown in Fig. 2. The abscissa is the angular separation between target and maskers in degrees, and the ordinate is the T/M at threshold in decibels. Thus, the curves all begin at the point denoting 0° separation and then display an attenuation characteristic as spatial separation increased and thresholds decreased. In order to show the individual listener T/M's, the curves were *not* normalized to 0 dB attenuation; therefore the attenuation values are equivalent to the amount of spatial release, which is obtained by subtracting the T/M for a spatially separated condition from the T/M for the colocated condition. The two functions and associated data points represent the two room conditions (BARE: filled circles, PLEX: open circles). The bandwidth of the filters (assuming a symmetric filter[2] computed from the 3 dB down point), the maximum attenuation, or range, of the filter (representing the value r in decibels in the roex filter expression) and the angular separation at which the maximum release was achieved are presented in Table I. The angular separation was defined as the smallest separation that was equal to the maximum release in dB on the fitted function. For 5 of the 6 listeners, the −3 dB bandwidth occurred at less than ±10° separation for both rooms' functions with the average value near ±6°. The remaining listener, L3, appears to be something of an outlier with respect to these measurements. The range of the filter/maximum masking release on the fitted functions varied across listeners from roughly 8 to 15 dB for the low-reverberation room (BARE) and 4–10 dB in the more reverberant room (PLEX). The angular separations at which the maximum release was first reached ranged from 31° to 55° with the exception of the function for Listener 3 in the low-reverberation room, which did not show asymptotic behavior in the range tested.

C. Monaural listening

As mentioned in Sec. II, the amount of attenuation obtained with the monaural hearing protectors was estimated by measuring the speech identification threshold in quiet as

TABLE I. Filter characteristics from best-fitting roex (p,r) filter functions computed on masked thresholds from individual listener data displayed in Fig. 2.

	BARE			PLEX		
	3 dB down point	Maximum release		3 dB down point	Maximum release	
Listener	(±deg)	(dB)	(±deg)	(±deg)	(dB)	(±deg)
L1	5	15	37	9	8.5	47
L2	9	11.6	55	7	7.8	33
L3	22	7.7	90	15	4.2	37
L4	5	15.3	33	6	8.8	33
L5	5	15.5	35	6	9.9	33
L6	6	10.4	33	6	8.6	31
Intersubject mean (standard deviation)	±8.7 (6.7)	12.6 (3.2)	±47.2 (22.6)	±8.2 (3.5)	8.0 (3.5)	±35.7 (5.9)

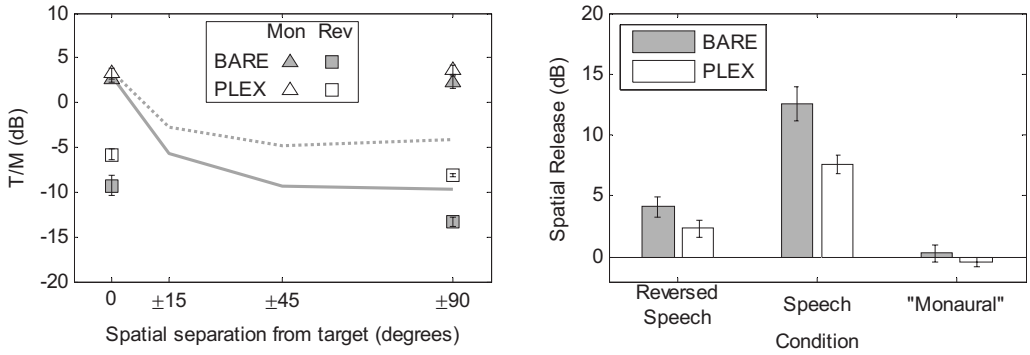


FIG. 3. (Left) Mean T/M at threshold at 0° and ±90° in both rooms for the monaural condition ("Mon," where one ear was occluded with an earplug and muff) and for the condition where participants listened binaurally but the speech maskers were time-reversed ("Rev"). Error bars represent ±1 standard error of the mean. The group mean T/M's from Fig. 1 (binaural listening, forward speech maskers) are replotted for comparison (BARE: solid line, PLEX: dashed line). (Right) Mean spatial release from masking (dB), calculated on an individual listener basis and then averaged across listeners, for the conditions plotted in the left-hand panel.

the listener wore hearing protectors on both ears and comparing that value to the one measured in the unoccluded case. Averaged across all listeners and sessions, the amount of attenuation achieved was 38.1 dB (SE=1.7 dB), relative to each listener's unoccluded speech threshold.

The left-hand panel of Fig. 3 displays the T/M's at threshold in the monaural condition for the colocated and ±90° separation conditions in the two room conditions (BARE: filled triangles, PLEX: open triangles). Also shown for reference are the mean data curves from Fig. 1 for binaural (unoccluded) listening for the same spatial conditions (a solid gray line for the BARE room result and a dashed gray line for the PLEX room result). The average threshold for speech identification with colocated target and maskers was essentially the same regardless of whether the listeners were performing the task monaurally (listening with the earplug and earmuff) or binaurally (unoccluded) for both room conditions. The average T/M at 0° for the monaural listening condition in the low-reverberation room was 2.5 dB (SE =0.2 dB) and was 3.3 dB (SE=0.6) in the more reverberant room; both T/M's were within one decibel of the results for binaural listening.

The most striking (although not unexpected) finding in the monaural data was that wearing the hearing protectors on one ear nearly eliminated the benefit of spatial separation. When the maskers were presented at ±90°, the average T/M's for the masked speech thresholds were approximately equivalent to the colocated thresholds in both room conditions. In the low-reverberation room, the average T/M at ±90° was 2.2 dB (SE=0.7 dB) and in the more reverberant room, it was 3.7 dB (SE=0.5 dB). For both room conditions, the average monaural T/M's at threshold in the colocated and spatially separated conditions were not statistically different from the average T/M for the binaural colocated thresholds in a repeated-measures ANOVA [BARE: $F(2,4)=1.3$, $p=0.36$; PLEX: $F(2,4)=0.410$, $p=0.69$]. In addition, the one truly monaural listener (complete unilateral deafness following vestibular schwannoma removal) who was tested showed no difference in the T/M at threshold for the colocated versus separated conditions in both rooms. The severe reduction in spatial release from masking in the monaural condition is shown clearly in the right-hand panel of Fig. 3, which displays spatial release for the binaural listening condition in the center pair of bars and the monaural listening condition in the right-most pair of bars for both room conditions. By inference, any acoustic differences between the colocated and separated conditions that were present monaurally were insufficient to explain the spatial release from masking found in the binaural condition.

D. Effect of masker type

Figure 3 (left-hand panel) also shows the group means and standard errors for the colocated and ±90° spatially separated conditions using time-reversed speech maskers (squares). The corresponding time-forward mean results are again provided for comparison (BARE: solid gray line; PLEX: dashed gray line). In the colocated condition, the average T/M at threshold was lower with the time-reversed speech maskers than when the maskers were intelligible in both room conditions. In the low-reverberation room, mean threshold T/M at 0° was −9.3 dB (SE=1.2 dB), which was 12.3 dB lower than with the forward-speech maskers. A similar result was found in the more reverberant room, where group-mean threshold at 0° was −5.8 dB (SE=0.6 dB). This was 9.2 dB lower than the corresponding threshold with forward-speech maskers. When the maskers were presented at ±90° there was a modest benefit of spatial separation in both room conditions. In the low-reverberation room, mean threshold at ±90° was −13.4 dB (SE=0.5 dB), corresponding to 4.1 dB spatial release. In the more reverberant room, there was 2.3 dB spatial release, as mean threshold at ±90° was −8.1 dB (SE=0.1 dB). As compared to the forward-speech maskers, all of the threshold T/M's were lower for the reversed speech maskers, but the greater reductions occurred for the colocated condition. Further, the amount of reduction in T/M was similar for both room conditions. This effect reduced the mean spatial release from masking observed with the reversed speech maskers, although the benefit remained statistically significant. Spatial release for the reversed speech condition is plotted in the left-most pair of bars in the right-hand panel of Fig. 3. A two-factor repeated-measures ANOVA on T/M at threshold for the reversed speech maskers with room and spatial condition as between-

subject factors confirmed there were significant main effects for room [$F(1,4)=63.3$, $p=0.001$] and spatial conditions [$F(1,4)=19.2$, $p=0.01$]. The interaction between room and spatial condition was also significant [$F(1,4)=11.7$, $p=0.03$].

IV. DISCUSSION

The current findings support the notion of a tuned response in azimuth resulting in a filter-like pattern of masking results. This conclusion is based on the progressive decrease in T/M at masked threshold with increasing spatial separation of symmetrically placed speech maskers over a narrow range of azimuths. The magnitude of the effect varied across listeners but, on average, was more than 12 dB in the low-reverberation room. This effect appears to be mediated by a variety of aspects of the listening environment, stimuli, and task. However, this benefit did not appear to increase in proportion to increasing differences in target-masker azimuth across the range of values tested (see Fig. 1). Instead, most of this effect was obtained within the first 15° of spatial separation of sources with the full benefit for most listeners realized by about 45°. The −3 dB bandwidths from the fitted filter functions are thus quite narrow with the average value less than ±10°. The bandwidths computed here are similar to the narrow bandwidths found by Teder-Salejarvi *et al.*(1999).

The tuning primarily reflects, we believe, a reduction in informational masking as the target and masker are spatially separated. Although it is difficult to determine the relative amounts of energetic and informational masking that are present in a speech-on-speech masking task, there are several factors that suggest that this is the case. The large release from masking in the colocated condition when the masker speech is reversed is consistent with a reduction in informational masking. However, this technique only provides a rough approximation of the amount of informational masking that occurs and is based on the assumption that energetic masking is equivalent when speech maskers are played backwards or forwards—an assumption that may not hold in some cases (e.g., Rhebergen *et al.*, 2005). Unlike broadband noise, reversed speech has similar temporal fluctuations to normal speech, making it intermediate to noise and speech along a hypothetical continuum of target–masker similarity. Further, there is the possibility that the similar and highly structured nature of target and masker sentences used in the current study differentially affects the amount of energetic masking present in time-forward and reversed maskers. In order to examine this possibility, a control experiment using speech-shaped speech-envelope-modulated noise that took its modulation pattern from forward and time-reversed speech was conducted with four additional listeners. This experiment is described in the Appendix. The conclusion drawn from those data is that the reduction in masking for the time-reversed maskers cannot be attributed to a decrease in energetic masking but must be due to a decrease in informational masking. Further, the spatial release found in both conditions was much less than was found here for speech maskers and was in good agreement with past results in which noise maskers were presented from symmetric locations around a speech target (e.g., Bronkhorst and Plomp, 1992; Helfer, 1992; Peissing and Kollmeier, 1997; Noble and Perrett, 2002). There is also other work, notably that of Brungart *et al.* (2006), that has suggested that the majority of masking that occurs in the coordinate response measure task employing speech maskers is informational in nature.

The type of masking present in the colocated condition of this study has implications for the mechanisms responsible for spatial release. Clearly, the basis for the reduction in thresholds in the spatially separated conditions is the presence of interaural differences. This conclusion is supported by the results of the monaural condition in which no significant spatial release was found. In the symmetric masker configuration, there was no overall acoustical better ear advantage. Measurements made on a Knowles Electronics Manikin for Acoustic Research showed that the overall level (root mean square computed over entire sentence) of either of the maskers alone or target plus maskers differed by less than 1 dB across ears from all of the different spatial configurations tested in the experiment. However, this does not rule out the possibility that there could be short-term fluctuations in level or overall spectral differences due to head shadow effects, resulting in epochs of better T/M when the maskers are spatially separated from the target. Unlike the asymmetry leading to a "better ear" typical with spatial separation of a target and a single masker, the two ears receive roughly equal fluctuations of T/M but either ear may be "better" in the spatially separate case as compared to the colocated case[3] at certain moments in time. If this could explain some of the large spatial release from masking observed, then listeners should exhibit some spatial release when listening monaurally. However, the benefit of spatial separation was eliminated in the condition simulating monaural listening. As a further control, we tested a subset of the listeners while both ears were covered with hearing protectors. After increasing the target level to compensate for the attenuation (i.e., presented at the same sensation level), the spatial release was restored. This assured that the loss of normal pinna cues was not the reason for the difference. Thus, the advantages found here clearly depend on binaural listening.

Previous findings using speech maskers placed symmetrically around a speech target are most relevant to the current results. Noble and Perrett (2002) studied the spatial release from masking for a target speech source masked by symmetrically placed speech, speechlike, or noise maskers. In a soundfield experiment with maskers placed at ±30°, they found about a 4 dB spatial release for two speech maskers. This was smaller than the spatial release found in the current study for ±15° of separation. They found even smaller releases when the maskers were noise. In one experiment, they also measured performance for speech masker placements of ±90° and found only about a 1 dB further increase (relative to ±30°) in spatial release. Given the filter widths computed here (attenuation maxima at about 38° excluding Listener 3 who had atypically large bandwidths), their results appear to be consistent with ours except that the magnitude of their spatial release was substantially less. It seems likely that their smaller spatial release from speech

masking is a consequence of their stimuli and tasks producing smaller amounts of informational masking than those in the present study.

Spatial release from masking in single noise-masker conditions increases with increasing target-masker separation over a range of values (e.g., Plomp, 1976; Plomp and Mimpen, 1981), a result that is consistent with the predictions of Zurek's (1993) model. Also, a speech target masked by a single speech masker follows a similar pattern of spatial release. Asymmetric, single masker placement could result in a function like those shown in Fig. 2 for either type of masker as it is progressively separated from a speech target. However, although both noise-masking-speech and speech-masking-speech produce declining threshold functions in the asymmetric placement condition, the fact that the two differ in the symmetric placement condition is, we believe, quite significant and is consistent with the conclusion that the underlying mechanisms are different as well. This interpretation is in agreement with Noble and Perrett's conclusions despite the discrepancy in the size of the effect. Additionally, the recent work of Brungart *et al.* (2006) suggesting that energetic masking is not a major factor for speech materials and procedures similar to those used here calls into question any strong role of binaural analysis (i.e., within-channel improvements in T/M, or "masking level differences"). Bronkhorst (2000) has proposed a straight-forward approach to the prediction of spatial release for a speech target presented from the front in the multiple masker case. Based on his equation (Bronkhorst, 2000, p. 123) and the parameters he estimated by fitting data from several studies, the prediction for the amount of spatial release for two maskers placed symmetrically at $\pm 90°$ is about 2 dB. That value is consistent with empirical reports of spatial release from noise maskers (e.g., Noble and Perrett, 2002), although it is somewhat smaller than the 4.6 dB effect found by Bronkhorst and Plomp (1992) for a modulated noise masker. For the procedures and target stimuli used in the current study, we found a value of about 1.5 dB in the low-reverberation room when measured empirically using broadband noise maskers[4] and about 3.5 dB for either time-forward or -reversed speech-shaped speech-envelope-modulated noise (refer to the Appendix). Thus, when the primary limitation on performance is energetic masking, the conditions tested in this study yield comparatively small spatial advantages. It appears to be possible that the presence of a high degree of informational masking is necessary (but not necessarily sufficient) to observe the large and sharply tuned effects found here. It is also clear that models of spatial release that only take into account better-ear listening and binaural analysis cannot predict these large effects.

The interpretation of the current results is that the greatest amount of informational masking was present in the colocated condition for the forward-speech maskers. Either spatial separation of the stimuli or time reversing the maskers caused a large reduction in the informational masking. Once the informational masking was reduced by means of decreasing target-masker similarity by time-reversing the masker, further reductions due to spatial separation were minimal and possibly indicate that performance had reached a limit imposed by the remaining energetic masking. The conclusion then is that the amount of informational masking in a task influences the degree of spatial benefit observed after eliminating the better-ear advantage. This interpretation is consistent with the relatively small amounts of spatial release from masking observed in previous studies that used various energetic maskers or presented the stimuli in low-uncertainty conditions or provided other strong perceptual segregation cues (cf. Kidd *et al.*, 1998; Arbogast *et al.*, 2002; Noble and Perrett, 2002; Hawley *et al.*, 2004; Culling *et al.*, 2004; Best *et al.*, 2005).

One of the factors influencing the magnitude of spatial release observed was the amount of reverberation in the listening environment. There was less spatial release from masking in the more reverberant room (the PLEX condition). Although the T/M's at threshold were stable across room conditions for the colocated target and maskers, the T/M's at threshold were higher in the more reverberant room when the maskers were spatially separated.

Increasing reverberation did not affect the colocated thresholds in the current three-talker experiment, but did in the two-talker experiments of Plomp (1976), Culling *et al.*(2003), and Kidd *et al.* (2005a). This might be explained by several factors, including differences between studies in the stimuli and procedures used and in the amount of reverberation present. In the two studies using rooms or simulations with longer reverberation times than those of the current experiment (Plomp, 1976; Culling *et al.*, 2003), increasing reverberation increased T/M by 2–4 dB for the colocated condition. However, in both studies the T/M's in the colocated conditions were generally lower (better) than those found here, likely because there was only a single masker talker. Importantly, the effect of increasing reverberation on the benefit of spatial separation found here is somewhat different than that reported by Kidd *et al.* (2005a) in the same room conditions for a single masker either colocated with the target at 0° or spatially separated at 90°. In that study, when the masker was another speech signal (targets and maskers were comprised of multiple narrow and mutually exclusive frequency bands), a large release from masking was preserved in the more reverberant room (16.0 dB in PLEX versus 16.7 dB in BARE). The main difference between the results of the two studies is that the group mean T/M's at threshold in the colocated condition were more variable and lower in the earlier study. As noted previously, the colocated thresholds found here for two speech maskers were remarkably constant across listeners and room conditions including the monaural control. In the earlier study, the range of performance across listeners in the colocated case for a single masker was much larger. The lower mean thresholds in the earlier study, combined with the larger intersubject differences, suggest that the source segregation cues were different. Although we do not know for certain, it seems possible that thresholds in the colocated, two-talker masker used here are determined simply by relative level. When three talkers of the same sex uttering similar sentences are colocated, there may be insufficient cues (e.g., F0 differences, timbre differences, etc.) to segregate the target until it is the loudest of the three voices. The values of T/M at

threshold were around 2–3 dB meaning that the target was that much higher in level than either masker alone or approximately equal to the sum of the maskers. These values are comparable to performance in a study by Brungart *et al.* (2001) for monaural segregation of three same-sex talkers using the same stimuli. In the Kidd *et al.* (2005a) study, other segregation cues—notably timbre differences resulting from the narrowband processing of the speech targets and maskers (see also Arbogast *et al.*, 2002)—provided another means for segregating the sounds. It is possible that the timbre cue was disrupted by reverberation, causing thresholds to increase in the colocated condition about as much as in the spatially separated condition. Future work is necessary to determine whether longer reverberation times than those used here would disrupt the relative level segregation cue and further increase the colocated thresholds for two or more masker talkers.

A related issue is that the large individual differences in spatial release found here are almost entirely attributable to differences in performance in the spatially separated conditions. The level cue that we speculate forms the basis for segregation in the colocated case seems to be one that most listeners were able to use quite effectively and is robust with respect to this amount of reverberation. However, in the spatially separated conditions, the ability to use interaural differences to segregate and selectively attend to the target—and ignore the maskers—varied widely across listeners. The large individual differences found in the spatially separated condition are not unlike the large individual differences found in the ability to use various segregation cues to overcome informational masking (e.g., Neff and Dethlefs, 1995; Durlach *et al.*, 2003; Richards *et al.*, 2002).

Threshold T/M increased in the spatially separated condition with increasing reverberation. This finding supports earlier work by Helfer (1992) using nonsense syllables masked by symmetrically positioned "cafeteria noise" and by Kidd *et al.* (2005a) and Culling *et al.* (2003) for asymmetric speech maskers. However, the conclusion that can be drawn from the current data differs from that of Culling *et al.* (2003) for several reasons. They observed that the benefit of spatial separation was eliminated with increasing reverberation. Several differences in the design of their experiment and the current experiment could help to explain this apparent discrepancy. First, because they were interested in the interaction of F0 contours (intonated, monotonous, or inverted) with spatial separation and reverberation, the task they used consisted of identifying key words from a male talker in the presence of a single female talker. This potentially involved less informational masking than in the present study as a strong segregation cue was present in all conditions. Their thresholds for the colocated case were better and the release from masking with spatial separation in the simulated anechoic space was smaller than in the current results. In addition, the elimination of the effect in reverberation, as well as the overall increase in thresholds in reverberation, may have been due to the greater amount of reverberation present in their experiment.

Considered across these various studies, reverberation clearly adversely affects performance in multitalker situations. Culling *et al.* (2003), Kidd *et al.* (2005a) and the current results have all demonstrated that the T/M at threshold increases with increasing reverberation when the competing talkers are spatially separated from the target. The most likely reason for this effect, we believe, is that the temporal "smearing" caused by reverberation reduces the interaural time and level differences that normally help the listener segregate the different sound sources and permit the listener to focus attention on the target. Performance in the absence of spatial cues appears to depend on the availability of other cues (e.g., overall level differences) for segregating the target from the colocated maskers that are less sensitive to the temporal smearing caused by this amount of increased reverberation. Therefore, whether spatial release per se is reduced, and, if so, to what extent, appears to depend on the specific conditions tested in the experiment.

V. SUMMARY AND CONCLUSIONS

The motivation for the current study was to better understand the processes that allow listeners to selectively attend to one source while ignoring competing sources in realistic room conditions. Listeners gave a selective response to a talker located straight ahead in the presence of two colocated or symmetrically positioned interfering talkers. Overall, filter-like properties were observed in the pattern of responses and it seems likely that this effect depends on the specifics of the stimuli and task. The spatial advantage observed appears to be a consequence of the listener using interaural differences to improve perceptual segregation of the target from the maskers and to focus attention at a point in space in order to overcome informational masking.

Spatial release from masking increased in these experiments as a function of increasing target–masker separation in azimuth from 0° to ±45° with negligible improvement for increasing spatial separation further to ±90°. Most of this release occurred for the initial separation of ±15°, suggesting that even small spatial separations provide large perceptual benefits. The spatial filters that were computed from the results were quite narrow (generally less than ±10° at the 3 dB down points) and maximum values of attenuation were mostly achieved in the range of ±30° to ±45°. A relatively large amount of spatial release was observed in this task, in excess of 12 dB for the larger separations, when listening binaurally in the low-reverberation room. When participants listened with one ear occluded by an earplug and earmuff to simulate monaural listening, spatial release was nearly eliminated; the average T/M's in the colocated and spatially separated conditions were not statistically different from the average T/M for the binaural colocated condition. The control conditions of time-reversed speech maskers, modulated noise maskers, and Gaussian noise maskers all produced much less spatial release than was obtained for the time-forward speech maskers. The large effects produced when informational masking was emphasized allow us to highlight the importance of the perceptual aspects of a process that has traditionally been thought of as explainable by fairly low level processes as in the model proposed by Zurek (1993), to account for spatial release from masking for speech in noise.

When room reverberation was increased, a fairly large spatial release from speech-on-speech masking (over 8 dB) was still present. As in the low-reverberation room, most of the spatial release occurred for the initial ±15° separation. In the more reverberant room, listeners needed a more favorable T/M to identify the target when the talkers were spatially separated, but the same T/M when the talkers were colocated. Thus, increasing reverberation reduced the spatial advantage. We speculate that the segregation cues in the colocated condition resulted from stimulus level differences that were unchanged by increasing reverberation. In the spatially separated conditions, the segregation cues were likely differences in perceived location based on interaural timing and level differences that were disrupted by the increase in reverberation.

Time reversing the maskers resulted in a greater change in performance for the colocated condition than in the spatially separated condition and this reduced the spatial release from masking. The reduction in similarity between target and masker provided a large segregation benefit. When the talkers were spatially separated, there was less improvement due to time-reversing the maskers possibly because of the large improvement already achieved through spatial separation. This suggests that informational masking is greater when there are fewer segregation cues.

ACKNOWLEDGMENTS

This work was supported by Grant No. FA9950-05-1-2005 from the Air Force Office of Scientific Research and by Grant Nos. DC008440, DC00100, DC04545 and DC04663 from the National Institute on Deafness and other Communication Disorders (NIDCD). This work partly fulfills the requirements for the doctor of philosophy degree at Boston University for the first author. The authors wish to thank the listeners that participated in the experiment, as well as Steven Colburn, Melanie Matthies, Barbara Shinn-Cunningham, Frederick Gallun, Virginia Best, and Nathaniel Durlach for helpful discussions of the results. Deborah Corliss and Jacqueline Therieau provided assistance with the Plexiglas® in the Sound Field Lab. They also thank Richard Freyman and two anonymous reviewers for useful comments on earlier versions of this manuscript.

APPENDIX

A potential explanation advanced for the reversed speech result (reduced T/M at threshold especially for the colocated condition) is that the forward speech produced more energetic masking than the reversed speech because of the coherence of the target and masker envelopes. Because all of the sentences—both targets and maskers—have the same structure and are to some degree time aligned, the envelopes of the targets and maskers are positively correlated (near 0.5 by our estimates). Reversed speech, however, diminishes the correlation between target and masker (near zero or slightly negative). If the peaks in the envelopes are more highly correlated in the time-forward case, then it is possible that the spectral overlap is also greater, whereas a lower envelope correlation with time-reversed maskers could provide a better opportunity for extracting target information in envelope minima. Because this alternative explanation for the reversed-speech result affects the extent to which we can attribute the colocated findings to informational masking, an additional experiment was conducted to examine this possibility in more detail. A new group of four listeners was recruited and tested using the envelopes of the forward and time-reversed speech maskers to modulate speech-shaped noise (derived from the corpus of speech used in the experiment). Following some preliminary analysis and listening experience, the envelope low-pass cutoff was set at 10 Hz. Speech-shaped speech-modulated noise was used because it is generally considered to produce primarily energetic masking of speech with little concomitant informational masking. Because the envelopes are derived from forward versus reversed speech, the extent to which the target and masker envelopes overlap is about the same as for actual speech. The test procedures used were identical to those employed in the main experiment with samples of the time-forward or -reversed noises replacing the speech maskers. If the difference between forward and reversed speech maskers was due to greater energetic masking in the forward speech maskers, or to a better opportunity to extract target information in the masker envelope minima for the reversed speech, then there should be less masking obtained for the reversed speech-modulated noise masker than for the forward speech-modulated noise masker.

The results indicated no significant difference between forward and reversed noise at either spatial condition: colocated and ±90°. For the colocated condition, the group mean T/M's were about −6.5 dB, whereas the corresponding values for the spatially separated condition were about −10 dB. These values, including the approximately 3.5 dB spatial release from masking, are in good agreement with similar values reported by Bronkhorst and Plomp (1992).

Our interpretation of the results of this control experiment is that, despite the greater temporal overlap of the target and masker envelopes in the time-forward condition than in the time-reversed condition, the amount of energetic masking was about the same. Therefore, the previous conclusions regarding the large improvements in T/M at threshold for reversed speech maskers compared to forward speech maskers being attributable to a release from informational masking appears to be supported by these results. It seems likely that the generally sparse spectral overlap of the CRM materials, as reported by Brungart *et al.* (2006), causes only small amounts of energetic masking even when the target and masker envelopes are somewhat coherent. The generalization of this finding to other speech intelligibility results should be made with caution, if at all, because of the closed-set highly structured nature of the CRM test and stimuli.

[1]What happens to the information from the "poorer ear" when considering the advantage of better ear listening is not always stated explicitly. If the two ears are considered separate channels, then the best strategy should be to optimally combine information from each. So, the poorer ear would contribute to the overall percept. If combining the inputs to the two channels results in a single representation or image that is, in some sense, noisier than the better input, then summation would be disadvantageous.

However, if the listener is able to select only a single channel, then obviously attending to the better ear is advantageous.

[2]Although we do not have any direct evidence regarding the (a) symmetry of the filter, there is some evidence from closely related conditions (Kidd *et al.*, 2005b) suggesting a preference for attending to stimuli presented from the right-hand side in highly uncertain listening conditions. Whether this would affect filter symmetry is not known and because of the symmetric placement of the maskers, there is not a way to evaluate filter asymmetry in the current design.

[3]Apart from the complex spectrotemporal patterns that may be different for the colocated and spatially separated conditions, imagine this extremely over simplified example to illustrate the point. The target presented from 0° arrives roughly equally at the two ears however, each of the two maskers when presented from ±90° are essentially low-pass filtered by the head when received in the opposite ears. Therefore, the high frequency masker energy from one of the two sources is reduced in each ear relative to the case of both maskers arriving equally at the two ears when colocated with the target. Whether or not these effects should actually aid in identifying the target is not obvious but the results of the monaural control condition would appear to suggest that they do not.

[4]It has been demonstrated previously by Bronkhorst and Plomp (1992) and Noble and Perrett (2002) that the spatial release for two symmetrically placed independent Gaussian noise maskers is quite small—on the order of one or two decibels. We double checked this finding for the stimuli and procedures used in the current experiment by replacing the speech maskers with two independent broadband noises. The average spatial release from masking for a group of four listeners (two who participated in the speech masking experiments and two new listeners) for the symmetrically placed noises was essentially the same in both room reverberation conditions (BARE: 1.5 dB and PLEX: 0.9 dB). This is consistent with the previous results referenced earlier.

American National Standards Institute. (2004). American national standard specification for audiometers. ANSI 3.6–2004

Arbogast, T. L., and Kidd, G., Jr. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Best, V., Gallun, F. J., Ihlefeld, A., and Shinn-Cunningham, B. G. (**2006**). "The influence of spatial separation on divided listening," J. Acoust. Soc. Am. **120**, 1506–1516.

Best, V., Ozmeral, E., Gallun, F. J., Sen, K., and Shinn-Cunningham, B. G. (**2005**). "Spatial unmasking of birdsong in human listeners: Energetic and informational factors," J. Acoust. Soc. Am. **118**, 3766–3773.

Boehnke, S. E., and Phillips, D. P. (**1999**). "Azimuthal tuning of human perceptual channels for sound location," J. Acoust. Soc. Am. **106**, 1948–1955.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. **86**, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Carlile, S., Hyams, S., and Delaney, S. (**2001**). "Systematic distortions of auditory space perception following prolonged exposure to broadband noise," J. Acoust. Soc. Am. **110**, 416–424.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Culling, J. F., Hodder, K. I., and Toh, C. Y. (**2003**). "Effects of reverberation on perceptual segregation of competing voices," J. Acoust. Soc. Am. **114**, 2871–2876.

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.

Drennan, W. R., Won, J. H., Dasika, V. K., and Rubenstein, J. T. (**2007**). "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," J. Assoc. Res. Otolaryngol. **8**, 373–383.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003**). "Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

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.

Goldberg, J. M., and Brown, P. B. (**1968**). "Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613–636.

Greenberg, G. Z., and Larkin, W. D. (**1968**). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," J. Acoust. Soc. Am. **44**, 1513–1523.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Helfer, K. S. (**1992**). "Aging and the binaural advantage in reverberation and noise," J. Speech Hear. Res. **35**, 1394–1401.

Hill, N. I., Bailey, P. J., and Hodgson, P. (**1998**). "A probe-signal study of auditory discrimination of complex sounds," J. Acoust. Soc. Am. **102**, 2291–2296.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005b**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd, G., Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005a**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acust. Acta Acust. **91**, 526–536.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

King, A. J., Bajo, V. M., Bizley, J. K., Campbell, R. A. A., Nodal, F. R., Schulz, J., and Schnupp, J. W. H. (**2007**). "Physiological and behavioral studies of spatial coding in the auditory cortex," Hear. Res. **229**, 106–115.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.

Middlebrooks, J. C., and Pettigrew, J. D. (**1981**). "Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location," J. Neurosci. **1**, 107–120.

Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.

Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1336.

Oxenham, A. (**2000**). "Influence of spatial and temporal coding on gap detection," J. Acoust. Soc. Am. **107**, 2215–2223.

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.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **110**, 1660–1670.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Plomp, R., and Mimpen, A. M. (**1981**). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," Acustica **48**, 325–328.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native

interfering speech," J. Acoust. Soc. Am. **118**, 1274–1277.
Richards, V. M., Tang, Z., and Kidd, G., Jr. (**2002**). "Informational masking with small set sizes," J. Acoust. Soc. Am. **111**, 1359–1366.
Scharf, B. (**1998**). "Auditory attention: The psychoacoustical approach," in *Attention*, edited by H. Pashler (Psychology Press Ltd., Hove, East Sussex) pp. 75–117.
Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **58**, 1848–1861.
Stecker, G. C., Mickey, B. J., Macpherson, E. A., and Middlebrooks, J. C. (**2003**). "Spatial sensitivity in field PAF of cat auditory cortex," J. Neurophysiol. **89**, 2889–2903.
Sterbing, S. J., Hartung, K., and Hoffmann, K. P. (**2003**). "Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig," J. Neurophysiol. **90**, 2648–2659.
Teder-Salejarvi, W. A., and Hillyard, S. A. (**1998**). "The gradient of spatial auditory attention in free-field. An event-related potential study," Percept. Psychophys. **60**, 1228–1242.
Teder-Salejarvi, W. A., Hillyard, S. A., Roder, B., and Neville, H. J. (**1999**). "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," Brain Res. Cognit. Brain Res. **8**, 213–227.
Tsuchitani, C. (**1988**). "The inhibition of cat lateral superior olivary unit excitatory responses to binaural tone bursts. I. The transient chopper discharges," J. Neurophysiol. **59**, 164–183.
Wright, B. A., and Dai, H. (**1994**). "Detection of unexpected tones with short and long durations," J. Acoust. Soc. Am. **95**, 931–938.
Wright, B. A., and Dai, H. (**1998**). "Detection of sinusoidal amplitude modulation at unexpected rates," J. Acoust. Soc. Am. **104**, 2991–2997.
Yin, T. C., and Chan, J. C. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.
Yin, T. C., and Kuwada, S. (**1983**). "Binaural interaction in low frequency neurons in the inferior colliculus of the cat. III. Effects of changing frequency," J. Neurophysiol. **50**, 1020–1042.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Needham Heights, MA), 255–276.

Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles

Asitha Mallawaarachchi and S. H. Ong[a)]
Department of Electrical and Computer Engineering, National University of Singapore, 9 Engineering Drive 1, Singapore 117576, Singapore

Mandar Chitre
Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore, 12a Kent Ridge Road, Singapore 119223, Singapore

Elizabeth Taylor
Marine Mammal Research Laboratory, Tropical Marine Science Institute, National University of Singapore, 14, Kent Ridge Road, Singapore 119223, Singapore

(Received 2 January 2007; revised 23 April 2008; accepted 27 May 2008)

Marine mammal vocalizations are often analyzed using time-frequency representations (TFRs) which highlight their nonstationarities. One commonly used TFR is the spectrogram. The characteristic spectrogram time-frequency (TF) contours of marine mammal vocalizations play a significant role in whistle classification and individual or group identification. A major hurdle in the robust automated extraction of TF contours from spectrograms is underwater noise. An image-based algorithm has been developed for denoising and extraction of TF contours from noisy underwater recordings. An objective procedure for measuring the accuracy of extracted spectrogram contours is also proposed. This method is shown to perform well when dealing with the challenging problem of denoising broadband transients commonly encountered in warm shallow waters inhabited by snapping shrimp. Furthermore, it would also be useful with other types of broadband transient noise. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945711]

PACS number(s): 43.66.Gf [MCH] Pages: 1159–1170

I. INTRODUCTION

Marine mammal vocalizations are recorded and studied for a variety of purposes, including research on behavioral and contextual associations, animal detection and localization, and census surveys. Some of these natural signals of interest are nonstationary waves well suited for analysis using time-frequency representations (TFRs). In this paper, we focus on the analysis of dolphin vocalizations using spectrograms, which are TFRs based on the short-time Fourier transform (STFT).

The ability of dolphins to communicate acoustically has been instrumental in many field studies. For example, communication signals commonly referred to as "whistles" have been used to identify individuals or groups of animals,[1,2] and to determine the acoustic features salient to the animals.[3] These studies provide valuable insights into the dolphin's vocal repertoire and behavioral associations.

Dolphin whistles are commonly characterized by their time-frequency contours in spectrograms that highlight nonstationarities and provide effective visual means of differentiating whistles from other acoustic signals. Features extracted from spectrogram contours of dolphin whistles have been used in many classification studies.

However, dolphin whistle contours are often corrupted by other underwater acoustic sources (noise), making extraction difficult, and it is therefore done manually. However, when a large number of whistles are to be extracted, time spent on extraction could be significant. Therefore, an automated method is highly desirable.

The semiautomated method of Buck and Tyack[4] is able to extract whistle contours if the start and end points are known *a priori*. For each time bin between the beginning and end of a whistle, the algorithm selects the frequency bin with the highest energy as a pixel on the contour. A further check is performed to avoid choosing pixels on high-energy harmonics. However, this method only works well for recordings with a high signal-to-noise ratio (SNR), which can be a difficult requirement to satisfy in natural waters.

A noise removal method to facilitate contour extraction in natural waters has been proposed by Sturtivant and Datta.[5,6] In this method, dolphin echolocation clicks (broadband transient signals) are removed by the sequential application of a vertical edge suppression filter and an exponential smoothing filter. After denoising, potential whistles are detected by local segmentation, followed by connectivity thresholding to obtain tonal components that are longer than a predefined time duration. The contour is traced on the original spectrogram using an "inertial" whistle-following technique, starting from candidate points detected by the previous segmentation step. The start and end points of whistles are identified by drops in local SNR.

However, this algorithm is not widely known, and there is no indication in the original publication to how it would perform for warm shallow water recordings. Since we had

[a)]Also with Division of Bioengineering. Electronic mail: eleongsh@nus.edu.sg

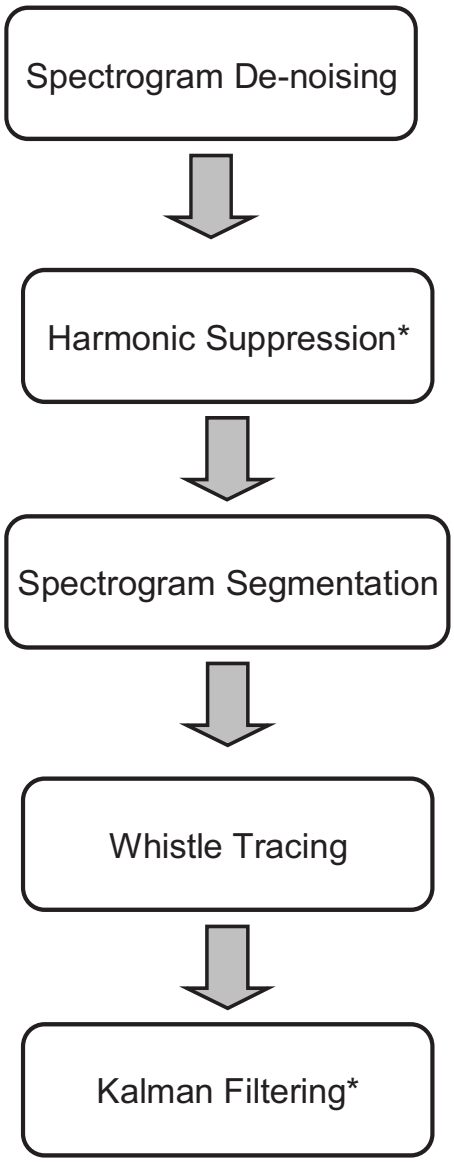


FIG. 1. Block diagram of the proposed algorithm. An asterisk denotes optional processing steps.

independently developed an algorithm consisting of (1) an image-based transient suppression filter to remove acoustic noise and (2) an adaptive image segmentation and tracing method to extract whistle contours from spectrograms, we conducted a comparative study to analyze the performance of the two methods. A flowchart which illustrates the method described in this work is given in Fig. 1.

Comparing the constituent elements of the two approaches, it may be noticed that while the denoising step of the Sturtivant and Datta method (SD) includes only a vertical edge suppression element, our proposed method also includes a horizontal smoothing element. Our experimental results (Sec. VII) indicate that this leads to better preservation of the signal while suppressing transient noise. The local normalizations used for denoising in SD are computed locally, whereas our method uses static directional filter kernels which are more efficiently implemented. Furthermore, the core algorithm components of denoising, segmentation, and tracing (excluding the optional steps—Kalman filtering and harmonic suppression) require fewer manual settings of parameters.

A denoised spectrogram may be converted back into a clean acoustic signal for play-back provided phase information is retained. Although the algorithm was developed for extracting dolphin vocalizations it has the potential to be used to extract other narrow-band nonstationary signals such as the humpback whale song.

Also introduced is an objective method of measuring the accuracy of tracing a vocalization spectrogram contour. This not only enables the comparison of tracing methods, but also helps tune the algorithm parameters such that optimal results can be obtained in a particular noise environment. Results obtained using recordings of vocalizations made by bottlenose dolphins (*Tursiops truncatus aduncus*) and Indo-Pacific humpback dolphins (*Sousa chinensis*) in warm shallow waters around Singapore confirm that our proposed method effectively denoises a wide variety of whistle contours and performs better than other methods in most instances tested.

II. SPECTRAL PATTERNS OF COMMON ACOUSTIC SIGNALS

A spectrogram is produced by converting a time-domain signal to the joint time-frequency domain by the STFT. Formally, the STFT of a discrete-time function $x[n]$ with respect to the window function $w[n]$ evaluated at the location $[\omega,m]$ in the frequency-time plane is defined as

$$X[\omega,m] = \sum_{n=-\infty}^{\infty} x[n]w[n-m]\exp(-j\omega n). \quad (1)$$

The columns of the matrix $X[\omega,m]$ contain the time-localized frequency content of the discrete signal $x[n]$. The values of X are generally complex, and it is customary to log-compress the absolute value of this transform for visual inspection due to the large dynamic range. This log-compressed gray-level two-dimensional (2D) image is called the spectrogram of $x[n]$, and is denoted by $\hat{X}[\omega,m]$.

In underwater environments where dolphin vocalizations are recorded using hydrophones, there are various types of acoustic sources: mechanical, such as produced by ships; natural physical sources such as waves; and biological sources. It is important that the spectral characteristics of this noise are taken into consideration when attempting to isolate dolphin whistles.

The two most commonly encountered dolphin vocalizations are narrow-band, frequency-modulated whistles and short-duration, broadband echolocation clicks. Whistles give rise to smooth frequency-localized contours [Fig. 2(a)] in a spectrogram, while clicks create vertical line patterns [Fig. 2(b)]. Signals generated by mechanical processes usually have low, constant frequencies, resulting in spectral patterns consisting of horizontal lines in the lowest regions of a spectrogram [Fig. 2(c)].

Ambient noise in warm shallow waters worldwide is dominated by the short-duration broadband crackling or popping sounds made by snapping shrimp, which has been shown to have a non-Gaussian energy distribution.[7] This

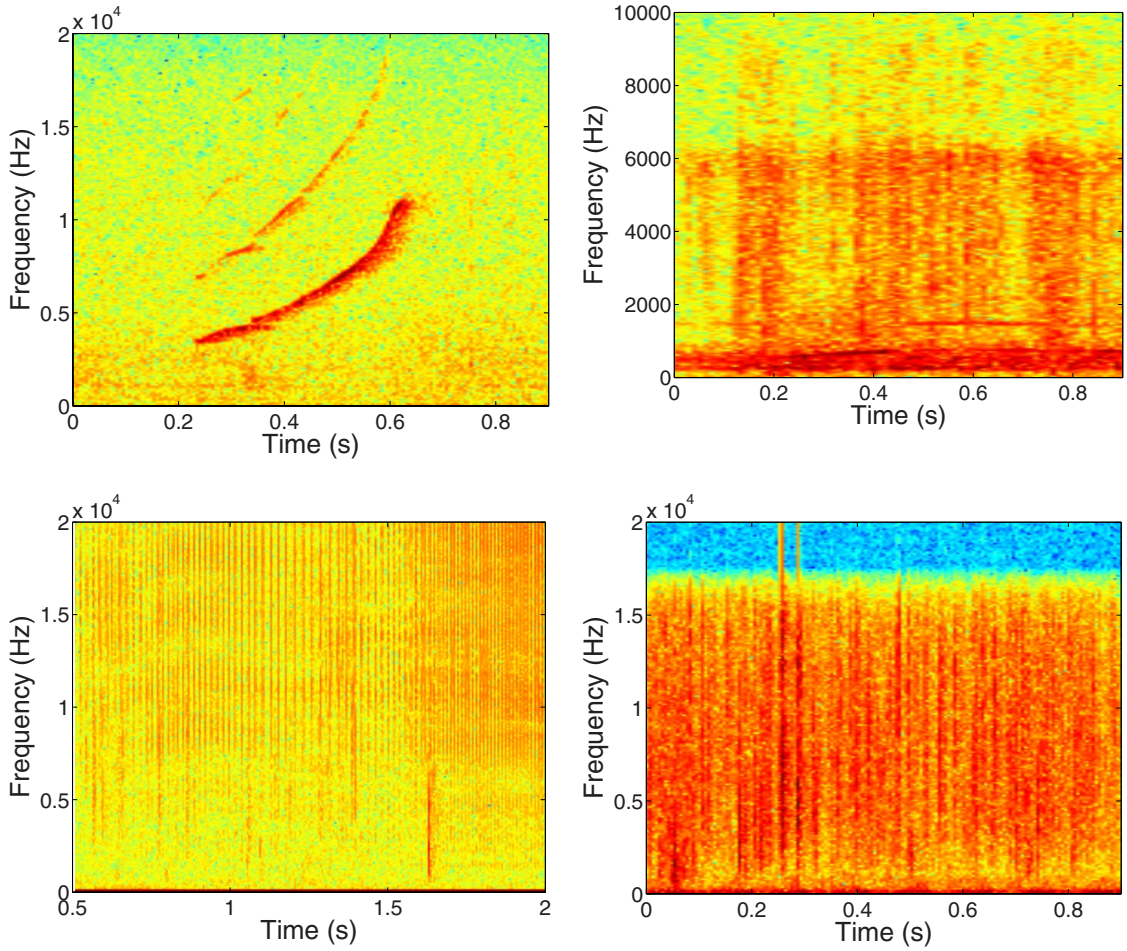


FIG. 2. (Color online) Common spectral patterns.

sound creates spectral patterns resembling narrow vertical lines. However, since many snapping shrimp produce these sounds simultaneously, individual "snaps" overlap and the resulting vertical structures are not as clearly defined as dolphin clicks [Fig. 2(d)].

III. SPECTROGRAM DENOISING

Spectrogram images are a special class of images that are not the product of conventional optical imaging. Noise in these images must be defined according to the higher-level detection task one is attempting. Unlike optical images, where the image noise usually arises from imperfections in acquisition or transmission, spectrogram noise is due mainly to the presence of "undesirable" acoustic sources. Some of these noise sources are introduced by human activities, while others are inherent in the recording environment and are described in Sec. II.

When detecting dolphin whistles all other acoustic sources are treated as noise sources; hence the spectral patterns to which they give rise in the resulting spectrograms are defined as noise. The aim of denoising a spectrogram is to facilitate the extraction of the desired type of spectral patterns by attenuating all other patterns. This paper introduces image processing methods to achieve this objective.

As a preprocessing step, the low-frequency tonal sounds created by mechanical devices such as motors and engines can easily be removed by high-pass filtering with a cut-off frequency set slightly below the lowest frequency at which dolphin whistles are expected ($\sim$1.5 kHz). However, if this method were to be used for denoising other types of signals, such as the low frequency calls of baleen whales, the band of interest would have to be modified accordingly. In the latter case, the use of a low-pass filter might be appropriate.

A. Denoising in nonimpulsive noise environments

The quality of spectrogram images of recordings made in pool environments and not excessively corrupted by transient noise can be improved significantly by an edge-preserving local-smoothing filter such as the bilateral filter.[8] This is essentially a neighborhood averaging filter with the kernel coefficients computed from the geometric closeness and the gray level similarity between the neighborhood center and the other neighborhood pixels.

If the center pixel of the local neighborhood being processed in spectrogram $\hat{X}$ is $\hat{X}[\omega,m]$ and any other pixel belonging to the same local neighborhood is denoted by $\hat{X}[\omega',m']$, the geometric closeness function, c, depends only on the relative positions of the two pixels $x=[\omega,m]$ and $\xi=[\omega',m']$. The gray level similarity function s, on the other hand, is a function of the relative pixel intensities $\hat{X}[x]$ and $\hat{X}[\xi]$.

In the shift-invariant Gaussian implementation of the bilateral filter, both the closeness and similarity functions are Gaussian functions of their respective arguments. Thus, we have

$$c(\xi,x) = \exp\left[-\frac{1}{2}\left(\frac{d(\xi,x)}{\sigma_d}\right)^2\right], \tag{2}$$

where the Euclidian distance $d(\xi,x)$ between ξ and x is

$$d(\xi,x) = d(\xi - x) = \|\xi - x\|$$

and

$$s(\xi,x) = \exp\left[-\frac{1}{2}\left(\frac{\delta(\hat{X}[\xi],\hat{X}[x])}{\sigma_r}\right)^2\right] \tag{3}$$

where

$$\delta(\hat{X}[x],\hat{X}[\xi]) = \|\hat{X}[\xi] - \hat{X}[x]\|.$$

Using the the aforementioned definitions, bilateral filtering can be described by

$$\hat{X}_{\mathrm{BF}}(x) = \frac{1}{k(x)}\sum_{\xi}[\hat{X}[\xi]c(\xi,x)s(\hat{X}[\xi],\hat{X}[x])], \tag{4}$$

where $\hat{X}_{\mathrm{BF}}[x]$ is the output of the filter operation on pixel $\hat{X}[x]$, and the summation is performed over all the neighborhood points ξ. The normalization factor $k(x)$ is obtained by

$$k(x) = \sum_{\xi}[c(\xi,x)s(\hat{X}[\xi],\hat{X}[x])]. \tag{5}$$

The parameter values of the bilateral filter depend on the size of the features one desires to preserve and the amount of smoothing preferred. The window size should be larger than the whistle-contour thickness, which depends on the time-frequency resolution set by the fast Fourier transform (FFT) window size. The Gaussian kernel width σ_d should be a fraction of the window size, while σ_r should be a fraction of the range of gray levels.

B. Short-duration transient suppression

Recordings made in open waters are typically more challenging to denoise. The vertical line patterns created by dolphin clicks and snapping shrimp can overlap whistles and complicate the tracing process. Pixels that are part of these spectral patterns should be detected and attenuated before whistle tracing.

The dominant direction of energy distribution in the local neighborhood of each pixel is detected using a set of four asymmetric kernels generated from the Gaussian functions (7),

$$G_1(p,q) = \exp\left[-\frac{1}{2}\left(\left(\frac{p}{\sigma_p}\right)^2 + \left(\frac{q}{\sigma_q}\right)^2\right)\right] \tag{6}$$

and

TABLE I. Asymmetric Gaussian kernels.

Kernel	Generating function	Value of σ_p	Value of σ_q	Orientation
$\nu_1(p,q)$	$G_1(p,q)$	a	$6a$	Horizontal
$\nu_2(p,q)$	$G_1(p,q)$	$6a$	a	Vertical
$\nu_3(p,q)$	$G_2(p,q)$	a	$6a$	Diagonal
$\nu_4(p,q)$	$G_2(p,q)$	$6a$	a	Diagonal

$$G_2(p,q) = \exp\left[-\left(\left(\frac{q-p}{\sigma_p}\right)^2 + \left(\frac{p+q}{\sigma_q}\right)^2\right)\right]. \tag{7}$$

The four Gaussian kernels ν_i $i \in \{1,\ldots,4\}$ given in Table I are oriented in the horizontal, vertical, and two diagonal directions. The values of σ_p and σ_q are chosen according to the kernel size, and in Table I, a is used to indicate their relative magnitudes. Generally, for a kernel of size $M \times M$, $a \approx M/10$ is recommended. In this work, a 9×9 window is used.

The spectrogram image $\hat{X}$ is filtered by ν_i to produce four intermediate images $\hat{X}_i$. Denoting the neighborhood center by $[p,q]$ and a pixel in the neighborhood by $[p',q']$, the intermediate images are given by

$$\hat{X}_i[p,q] = \frac{1}{t(p,q)}\sum_{p'}\sum_{q'}[\hat{X}[p',q']\nu_i(\|p'-p\|,\|q'-q\|)] \tag{8}$$

with the normalization factor $t(p,q)$ defined by

$$t(p,q) = \sum_{p'}\sum_{q'}\nu_i(\|p'-p\|,\|q'-q\|). \tag{9}$$

Since the output of the functions ν_i can be precomputed for a given size of the local neighborhood, the above operations can be efficiently implemented.

To remove the vertical spectral patterns, pixels belonging to local neighborhoods with vertical energy distributions are attenuated. Let $r(\xi)$ be the highest nonvertical energy average, and $v(\xi)$ the vertical energy average (Algorithm 1, see Table II). Therefore the expression $r(\xi) - v(\xi)$ evaluates to a positive value when the primary direction of energy distribution is nonvertical, and negative when it is vertical. Adding the expression $r(\xi) - v(\xi)$ to the original pixel value therefore has the effect of attenuating the pixels with a vertical energy distribution. The constants α and β are used to control the degree of attenuation.

Only the relative values of α and β are important. Higher relative α preserves more of the original detail while higher relative β increases the amount of attenuation of ver-

TABLE II. Algorithm 1: Transient suppression.

1:	**for all** pixel $\xi=(p,q)$ of $\hat{X}(\xi)$ **do**
2:	$r(\xi)=\arg\max \hat{X}_1(\xi),\hat{X}_3(\xi),\hat{X}_4(\xi)$
3:	$v(\xi)=\hat{X}_2(\xi)$
4:	$\hat{X}_{TS}(\xi)=[\alpha\times\hat{X}(\xi)+\beta\times(r(\xi)-v(\xi))]/[\alpha+\beta]$
5:	**end for**

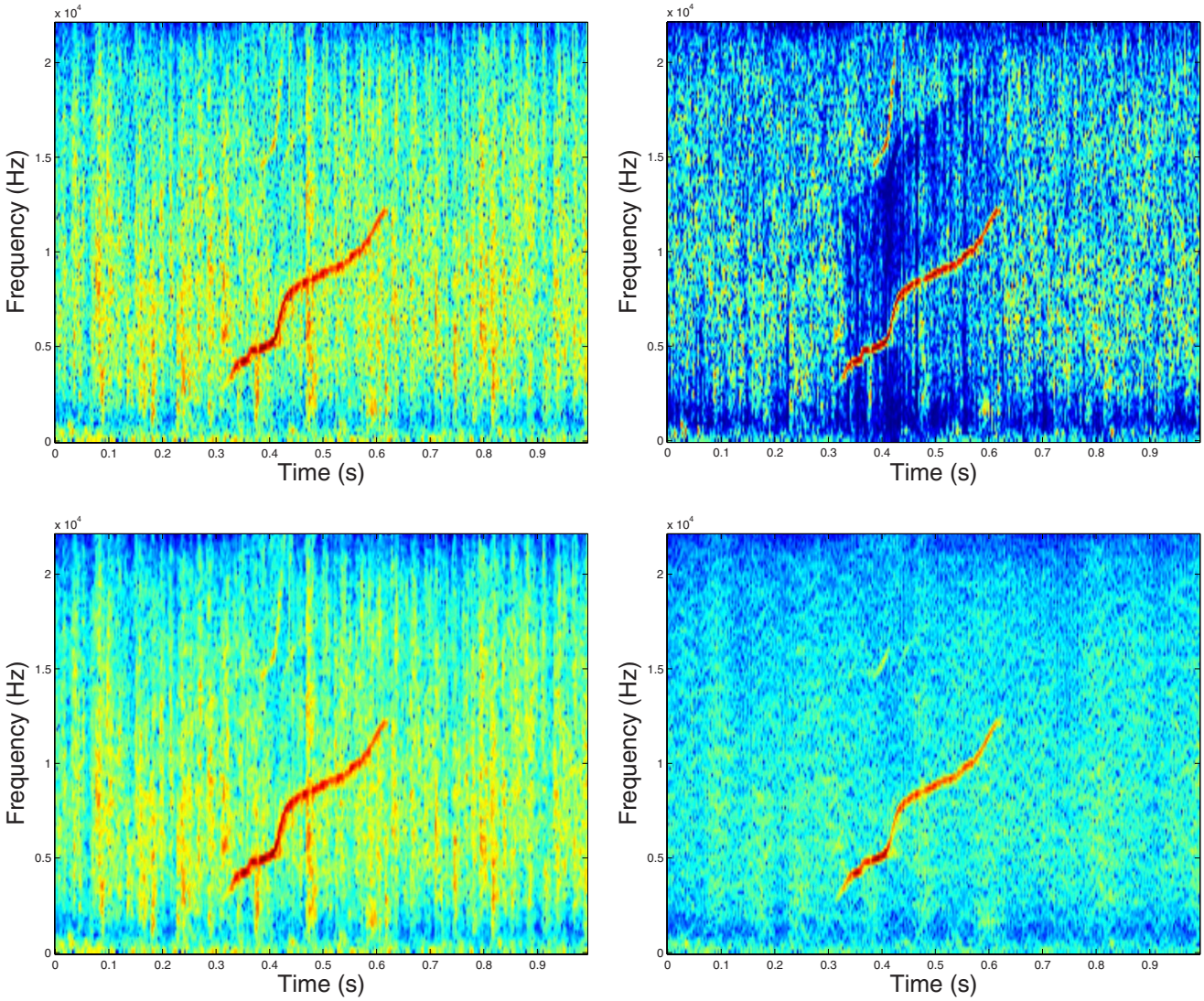


FIG. 3. (Color online) Denoising in the presence of transient noise.

tical spectral patterns. Although the procedure is described here in its noniterative mode of operation, there is no restriction on using the filter iteratively.

An example of applying this filter to a whistle corrupted by snapping shrimp noise is shown in Fig. 3 together with the output of bilateral filtering and an implementation of the method of SD.[6] Bilateral filtering [Fig. 3(b)] enhances connected high-energy regions irrespective of orientation and therefore produces poor denoising performance. The method of SD [Fig. 3(c)] reduces most of the transients and creates a "noise trough" around the whistle, but leaves behind a significant amount of noise pixels. In comparison, the proposed transient suppression filter [Fig. 3(d)] removes most of the undesired vertically oriented spectral patterns while preserving the whistle contour.

IV. HARMONIC SUPPRESSION

Whistles often contain harmonics similar in shape to the fundamental frequency variation with only a shift in frequency, and these can potentially hinder accurate tracing of the fundamental. The instantaneous frequency of a harmonic is an integer multiple of the fundamental, a property that can be exploited to automatically remove it from the spectrogram image.

A row of pixels in a spectrogram represents the time variation of a discrete frequency bin f_i, and, from bottom to top, the rows represent a linear increase in frequency. Let us define a pixel intensity vector $\mathbf{I}_i$ that contains the pixels in the ith row of the spectrogram. The harmonic suppression update equation for the ith row is expressed as

$$\mathbf{I}_i - \mathbf{I}_i - k_h \mathbf{I}_j, \tag{10}$$

where k_h is a user-defined scalar constant and the vector $\mathbf{I}_j$ contains the pixel values of the jth row for which $f_j = f_i/N$, where N is an integer. For example, if $N=2$ is used, harmonics which are $\{2, 4, 6,\ldots\}$ multiples of the base frequency will be attenuated. Therefore by repeating the procedure for different values of N, most harmonic patterns may be attenuated. A good choice for N is the set of the largest common divisors of the integer multiples of the fundamental that produced the harmonic pattern. In practice, $N \in \{2,3,5\}$ is used, and Eq. (10) is applied to every row from top to bottom, and iterated for each value of N.

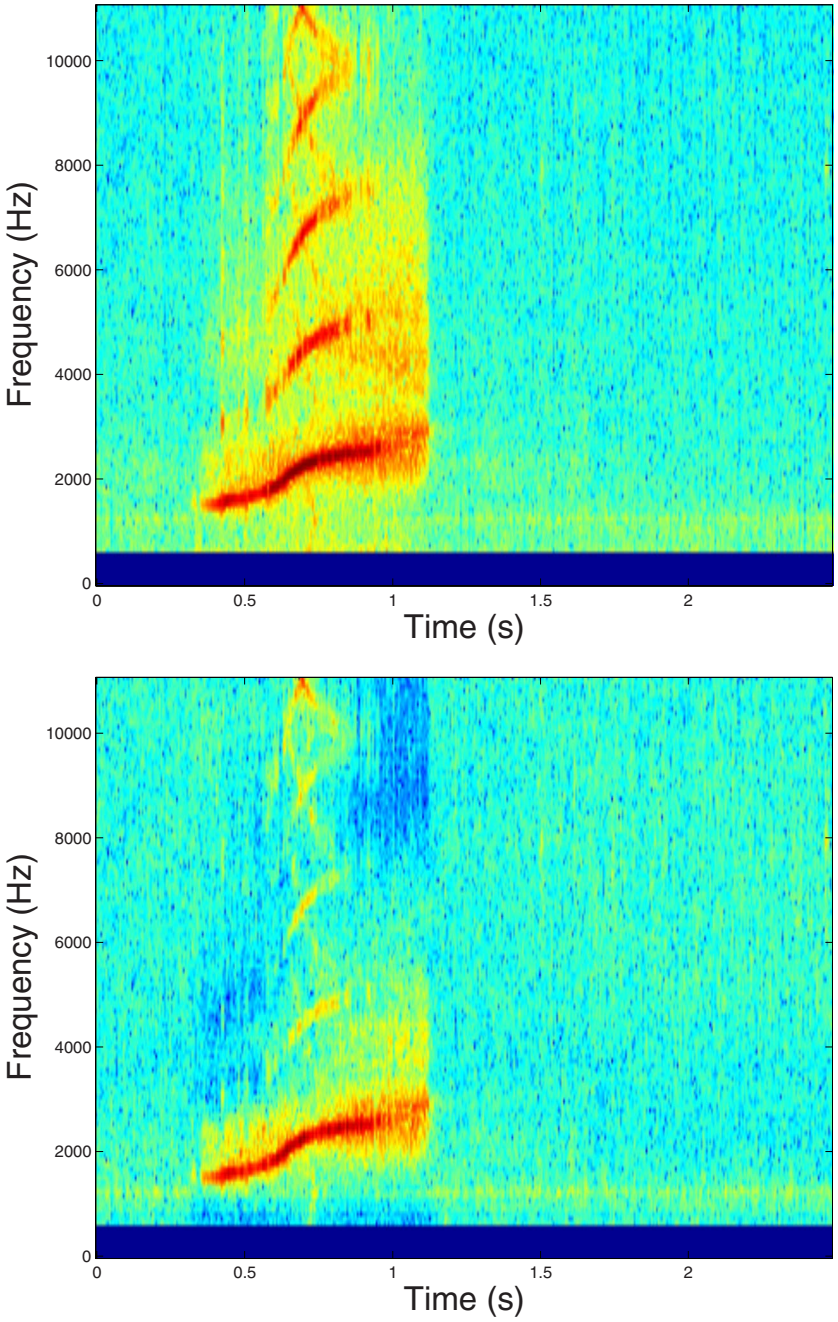


FIG. 4. (Color online) Iterative harmonic suppression.

Figure 4 shows how harmonics are suppressed from the spectrogram by this iterative procedure; the harmonics now have lower energy, while the fundamental signal retains intensity values similar to the original. This step should be used selectively since not all whistle recordings capture strong harmonic patterns. If multiple whistles overlap in time, and the fundamental (frequency variation) of one whistle intersects the harmonic pattern of another whistle, the harmonic suppression algorithm will attenuate the pixels belonging to the fundamental of the intersecting whistle. This algorithm should therefore not be used in such situations.

V. SPECTROGRAM SEGMENTATION

After the spectrogram has been denoised to remove unwanted spectral patterns, whistles can be extracted using image segmentation techniques. Although many segmentation techniques have been proposed, the objective of the current work is to choose a method applicable to spectrogram images and simple enough to be efficiently implemented. The latter objective is important because an on-line tool for whistle extraction would be very helpful to field scientists studying dolphins and other marine mammals.

Thus, a three-stage image segmentation technique is proposed. Thresholding is chosen as the first step as it can be efficiently implemented and requires only one parameter (the threshold) to be determined. Furthermore, there are several known methods of adaptively computing a threshold. Having the ability to adaptively select a threshold is crucial in spectrogram images because the ambient intensity values can greatly vary from one image to the other, even in spectrograms taken from different sections of the same recording.

Factors contributing to these variations include the dynamic nature of underwater acoustic sources, and the settings of the recording equipment used. Further details on thresholding are given in Sec. V A.

As an intermediate step, a morphological clean-up operation is performed on the output of the thresholding operation (Sec. V B).

Dolphin whistles are generally continuous contours with fading starts and ends. Therefore, the start and end segments are likely to be misclassified as background by thresholding. However, these segments are connected to other high intensity parts of the whistle that are readily identified by thresholding. We, therefore, propose the use of an intensity-based region-growing algorithm for the final stage as described in Sec. V C.

A. Adaptive thresholding

A thresholding function f operates on an *intensity* image I and produces a *binary* image J, using a global threshold T,

$$J(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geqslant T \\ 0 & \text{else.} \end{cases} \tag{11}$$

In the first segmentation stage, we use a higher global threshold T value to ensure noisy pixels are unlikely to be segmented as foreground. This creates the possibility of coclassifying some segments of the whistle as background noise, but the subsequent region growing step compensates for this.

Since spectrograms vary significantly in average energy level, depending on the recording environment and the type of dolphin vocalizations, the threshold T has to be calculated adaptively. The *Niblack* method[9] is a well-known and simple method for computing an adaptive threshold,

$$T = \mu + k\sigma, \tag{12}$$

where μ is the mean gray value, σ the standard deviation, and k a user-defined constant. We determine k from

$$\int_{-\infty}^{\mu+k\sigma} N_{\mu,\sigma}(x)dx = \rho, \tag{13}$$

where ρ is the percentage of background pixels. A normal distribution of gray values may be assumed even though the actual distribution can differ.[10] Using the *a priori* knowledge that over 90% of a spectrogram is background, our proposed method computes a high threshold by setting $\rho=0.96$. The calculation of k using Eq. (13) is implemented by using precalculated values for a Gaussian cumulative distribution function, stored in a table format. This is also referred to as a Z table.

Note that a simple and fast algorithm is preferred because, in the first stage of segmentation, we are mostly interested in obtaining a set of seed points that we can feed into the region-growing algorithm.

B. Morphological clean-up operations

As an intermediate step, mathematical morphology is used to improve the segmentation and remove any noisy outlying pixels from J. In morphology, the structure of a group of pixels is considered rather the the pixel intensity values. All morphological operators are therefore defined with respect to a "structuring element." It is possible to use morphology to remove noisy outlying pixels based on structural characteristics such as size, by using an appropriate structuring element.

Two basic morphological operators are opening and closing. Opening with a structuring element S will erase image structures (groups of connected pixels) smaller than S, while closing will fill gaps (holes) smaller than S. Morphological closing was first performed on J with a 2×2 square structuring element (SE) s_1, followed by opening with a 2×3 SE s_2. This dual operation preserves original detail and removes any remnant noisy outlying pixels.

C. Region growing

After removing outlying pixels with the morphological operators, the pixels of the segmented image are input as seed points in a 2D region-growing algorithm. Region-growing algorithms take one or more seed points together with a threshold value T' as inputs, and initially the output image contains only the seed(s). As the algorithm progresses, the neighborhood pixels of the seed(s) above the threshold T' are located and added to the output image, and then the neighbors of the newly added points are searched. This recursive algorithm continues until all the pixels meeting the criteria are added to the output image.

VI. WHISTLE TRACING

The final goal of the algorithm is to extract the time-frequency contour of the fundamental frequency variation of a whistle. A trace of this extracted contour should also be drawn on the spectrogram for visual inspection.

The candidate points for the trace consist of the strongest peaks in each time bin of the segmented spectrogram.[4] This simple process is used for whistle tracing as the denoising and segmentation steps contain most of the algorithmic intelligence. However, strong harmonics or remnant background noise can cause incorrect segmentation and, hence, outlying points in the automatically obtained trace. In such situations, an additional step which can correct outlying points is beneficial.

Based on the assumption that dolphin whistles are smooth curves without sudden jumps in frequency, we propose the use of a Kalman filter,[11] which is a model-based estimator. A Kalman filter takes imprecise measurements of a process that can be mathematically modeled, and produces, as output, a weighted average of the model prediction and the measurement. The weights depend on the relative confidence in the measurements and the model prediction.

The confidence values for each measurement (a point on the automated trace) are calculated and saved during tracing. Low confidence values are assigned to whistle points that exhibit sudden jumps in frequency. The confidence assignment function has a memory of 1 in the sense that if the

TABLE III. Second-order Kalman filter model used for whistle smoothing.

State vector	$x=[f\ v\ a]$
	f—frequency (position)
	v—rate of change of frequency (velocity)
	a—second derivative of frequency (acceleration)
State equations	$f(k)=ut(k)+\frac{1}{2}at^2(k)$
	$v(k)=u+at(k)$
	where u is the initial velocity
Discrete update equations	$f(k+1)=f(k)+v(k)\delta_T+\frac{1}{2}a\delta_T^2$
	$v(k+1)=v(k)+a\delta_T$
	$a(k+1)=a(k)$ where $\delta_T=t(k+1)-t(k)$
	$\begin{bmatrix} f_{k+1} \\ v_{k+1} \\ \alpha_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \delta_T & \frac{1}{2}\delta_T^2 \\ 0 & 1 & \delta_T \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_k \\ v_k \\ a_k \end{bmatrix}$

previous measurement had low confidence, the confidence attributed to the next measurement will be partially based on the previous value.

The same process model was used to "track" a smooth dolphin whistle as when tracking the trajectory of a particle moving in a straight line under constant acceleration. Frequency (f), which is the filtered variable with respect to time, is analogous to the position of the particle. In this context, velocity and acceleration correspond, respectively, to the first and second derivatives of frequency with respect to time. The Kalman model is summarized in Table III. The filter output is continuous and will be quantized to the nearest discrete Fourier transform (DFT) bin value. Figure 5 shows the plot of a traced whistle overlaid with the Kalman filter output. The whistle points adjusted by the filter are marked by ellipses.

One shortcoming of the Kalman filter is the slight shift of the trace to the right (on the time axis) caused by the filter. This indicates a "filter inertia," which resists changes in direction contrary to the model prediction. Effective use of Kalman filtering requires an appropriate trade-off between smoothness and flexibility (to change direction), and this can be done by setting the process and measurement error covariances. Another problem is that in situations where a large number of continuous outlying points are present, the filter will gradually adapt to those incorrect values and will fail to make the expected filter corrections. Despite these imperfections, Kalman filter corrections are beneficial in some situations.

VII. EXPERIMENTAL RESULTS

A. Experiment design

To evaluate the effectiveness of the automated tracing methods, 26 dolphin whistles were chosen from underwater recordings of vocalizations made by bottlenose dolphins (*Tursiops truncatus aduncus*) and Indo-Pacific humpback dolphins (*Sousa chinensis*) in waters surrounding Singapore. The sampling rate was 44.1 kHz, and spectrograms were created with an FFT window size of 256, with successive windows overlapping by 50%. Each whistle was approximately 0.5–1 s in duration and contained natural background noise dominated by snapping shrimp

The Marine Mammal Research Laboratory (MMRL) at the National University of Singapore provided the underwater recordings and the reference traces for the selected whistles. To obtain reference traces of the whistle contours, the spectrograms of the selected whistles were manually traced by the MMRL, and quantized to fit the spectrogram bins using a nearest neighbor algorithm. The whistle traces obtained by automated methods were then compared with the reference traces for quantitative evaluation of their performance using the metrics defined in the following.

B. Performance metrics

A quantitative method to evaluate the accuracy of an extracted whistle contour trace has not previously been proposed. This paper defines three metrics that can be used to gauge the completeness and accuracy of tracing a known whistle contour. They are defined with respect to a reference contour $\omega(m)$, with associated times $t(m)$ obtained by

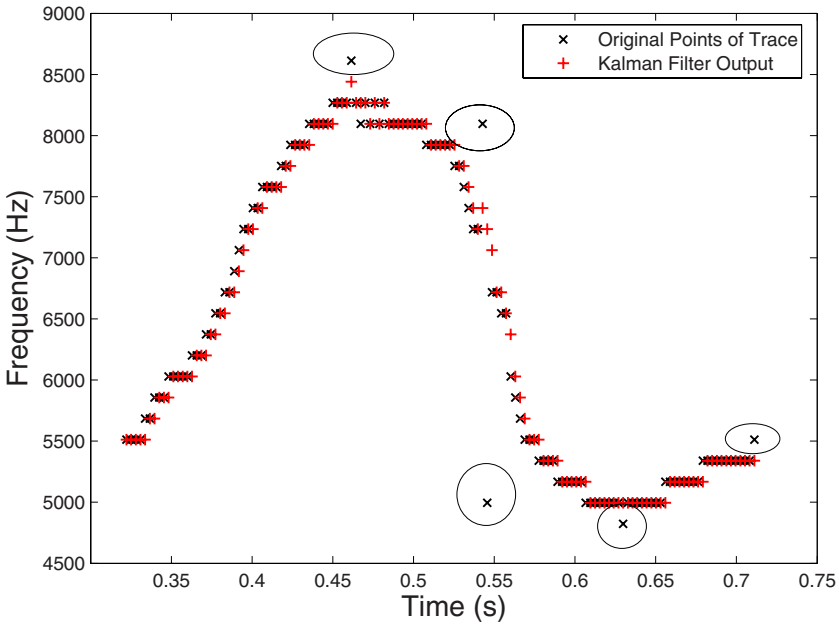


FIG. 5. (Color online) Plot of a traced whistle overlaid with the Kalman filter output. The whistle points adjusted by the filter are marked with ellipses.

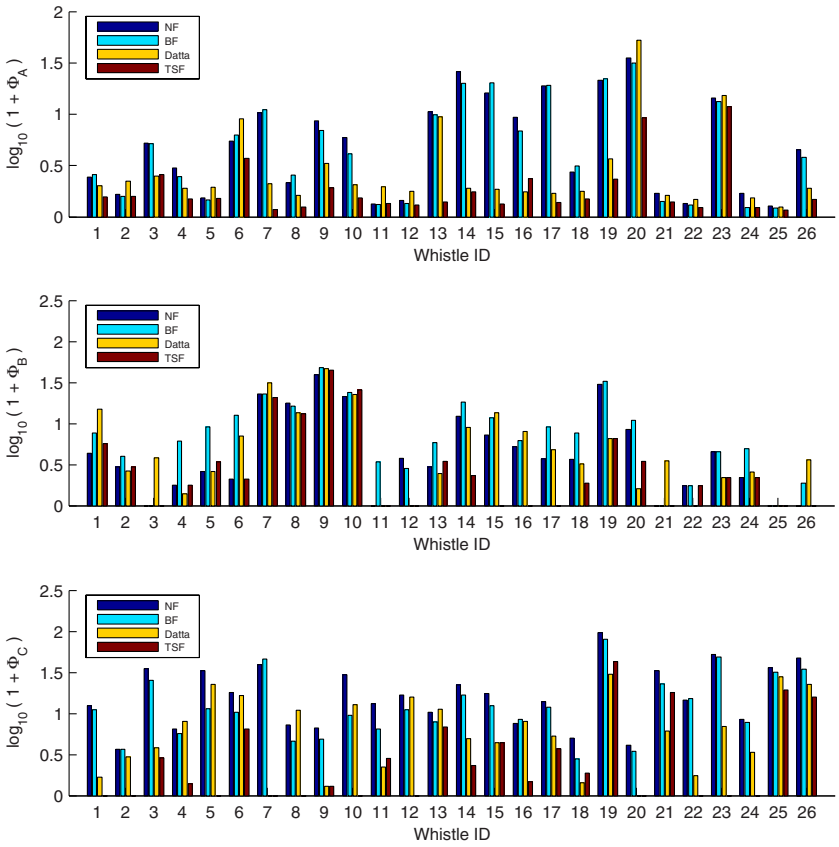


FIG. 6. (Color online) Performance metrics of tracing 26 recorded whistles.

manual tracing (Sec. VII A). Similarly, a whistle contour obtained using an automated method can be denoted by $\hat{\omega}(m)$, with associated times $\hat{t}(m)$.

Metric 1: Mean percentage error of tracing,

$$\Phi_A(\omega,\hat{\omega}) = \frac{1}{N}\sum_m \frac{\|\omega(m)-\hat{\omega}(m)\|}{\omega(m)} \times 100\,\%\,, \tag{14}$$

where N is the number of matching points in ω and $\hat{\omega}$. This is the relative percentage displacement of the contour $\omega(m)$ with respect to $\hat{\omega}(m)$, averaged over all the whistle points. The displacements are calculated only over the matching points of the trace, i.e., where $t(m)=\hat{t}(m)$. Here the absolute relative displacement is used instead of squared error since it provides a more direct measurement of the actual deviation from the reference trace.

Metric 2: Percentage of missing points,

$$\Phi_B(\omega,\hat{\omega}) = \frac{\beta}{|t(m)|} \times 100\,\%\,. \tag{15}$$

This is computed by using the number of missing time bins in $\hat{t}(m)$ compared to $t(m)$, denoted by β.

Metric 3: Percentage of extra points,

$$\Phi_C(\omega,\hat{\omega}) = \frac{\eta}{|t(m)|} \times 100\,\%\,. \tag{16}$$

This is computed by using the number of extra time bins in $\hat{t}(m)$ compared to $t(m)$, denoted by η.

The average of all three metrics provides a measure of the total tracing error in terms of the extent and accuracy of the traced contour. This metric is defined as follows.

Metric 4: Average percentage tracing error,

$$\Phi(\omega,\hat{\omega}) = \frac{\Phi_A+\Phi_B+\Phi_C}{3}. \tag{17}$$

C. Results and discussion

The set of 26 whistles was traced using four different approaches, three of which use the general procedure proposed in this work, with differing denoising methods. The fourth approach is an implementation of the method proposed by Sturtivant and Datta.[2,5] Specifically, the methods are:

(1) NF: No spectrogram denoising is performed in this method. The rest of the algorithm takes the original spectrogram images and performs segmentation and tracing.
(2) BF:Spectrogram denoising is performed using bilateral filtering.[8]
(3) TSF:Spectrogram denoising is performed using transient suppression filtering proposed in this work.
(4) SD: Whistle tracing is performed by implementing the method proposed by Sturtivant and Datta.[2,5]

Each algorithm must first be calibrated to determine the parameter values for optimal performance. This is done by a combination of visual verification of tracing results and computing the performance metrics. Note that the parameter settings used here are chosen to be generic to the recording environment and not optimized for each individual whistle. The first three metrics (Φ_A,Φ_B,Φ_C) are calculated on the

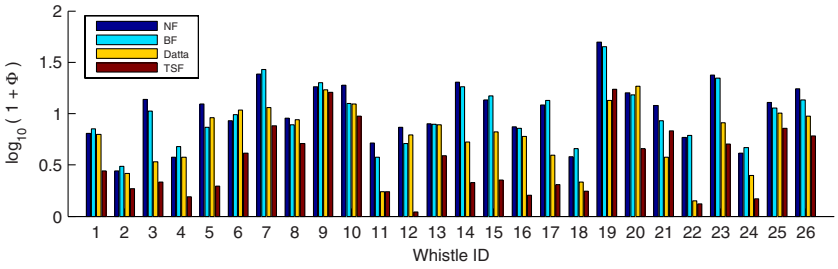


FIG. 7. (Color online) The average of the three metrics Φ_A, Φ_B, Φ_C for 26 whistle recordings.

tracing results of the whistle set and are graphically presented in Fig. 6. The average metric Φ is presented in Fig. 7.

NF is included as a "control sample" to quantify the effect of image denoising on tracing performance, particularly for BF and TSF. Compared to NF, BF tends to decrease the mean percentage error (Φ_A) and the number of extra trace points (Φ_C), indicating a reduction of outlying points in the traces [Figs. 6(a) and 6(c)]. However, the average percentage of missing points (Φ_B) rises in general, indicating that a significant number of signal pixels are also attenuated by the filter.

SD performs better compared to the BF method (on all metrics), and better than NF on Φ_A and Φ_C (but not Φ_B). From the tracing metrics and the visual inspection of the spectrogram plots, it can be inferred that the vertical edge suppression filter of SD is more effective in removing transients than the smoothing operation performed by BF.

TSF has an even greater impact on reducing the number of outlying points; at the same time, it manages to retain more of the whistle points of $\omega(m)$ (overall reduction of all metrics compared to NF, BF, and SD). This behavior is attributable to the vertical suppression and horizontal smoothing property of the directional filter kernels, which manage to effectively suppress high energy broadband transient spectral patterns while retaining frequency modulated narrowband whistles.

Statistical analysis of the results were performed on the results using STATISCA 6.0 (STATSOFT, Inc., 2001, Tulsa, OK). One-way multivariate analysis of variance (MANOVA) was performed simultaneously on the three scoring categories for the four methods. Log $(x+1)$ transformation successfully reduced homogeneity of variances for Φ_A (percentage error: Cochran's $C_{df3}=0.328$, $p>0.05$), Φ_B (percentage missed points: Cochran's $C_{df3}=0.264$, $p>0.05$), and Φ_C (percentage extra points: Cochran's $C_{df3}=0.348$, $p>0.05$). The results showed significant effects of the various methods on the scores (Wilk's $s_{9,239}=0.619$, $p<0.001$): univariate decomposition demonstrated that effects were significant for all three scoring categories [$\log(\Phi_A+1)$: $F_{3,100}=5.592$, $p<0.01$; $\log(\Phi_B+1)$: $F_{3,100}=5.418$, $p<0.05$; $\log(\Phi_C+1)$: $F_{3,100}=8.873$, $p<0.01$].

The SD method was slightly modified to improve its performance. First, a low-pass filter was introduced to remove high frequency tonals in addition to the high-pass filter stipulated in the original work. Second, when the inertial whistle-following algorithm stops upon encountering a sudden drop in SNR, it is automatically restarted to increase detection probability. The metrics for the SD method applied to filtered and unfiltered data were: square-root transformed "error" ($F_{1,50}=11.38$, $p<0.01$); and log (average+1) ($F_{1,50}=4.110$, $p<0.05$); and were significantly lower after imposing a low-pass filter than in unfiltered data.

Post-hoc Tukey's honest significant difference test demonstrated that TSF scores were consistently significantly lower (at $p=0.05$) than BF scores; the Φ_A and Φ_C (but not Φ_B) were significantly lower for TSF than NF; TSF also scored consistently lower than SD for all metrics but the difference was statistically significant only for Φ_B (see Fig. 8).

In summary, the statistical analysis of the results indicate that TSF consistently performed better than the other denoising methods. All metrics derived from the TSF method were consistently lower than the SD method, although this was only statistically significant for the Φ_B. SD metrics significantly outperformed the BF and NF methods only for Φ_C.

However, for some whistles, e.g., whistles 19 and 21 [Figs. 9(b) and 9(c), respectively], TSF does not perform well. In both cases a closely packed cluster of transients has caused an incomplete segmentation, resulting in a higher number of extra points on the trace. In these two situations, SD has not included the extra portion due to the SNR drop between the two segments and hence performs better.

Both SD and TSF methods perform poorly on whistle 9 [Fig. 9(a)], which contains rapid rising and falling segments. In such scenarios, methods employing vertical edge suppression filters, while useful for removing broadband patterns, also attenuate sections of the signal with similar characteristics. Additionally, due to its low SNR, all methods miss a large percentage of the whistle. Another reason for the poor performance for this whistle is due to the nature of the reference trace. Human vision is known to be remarkably adept at linking disconnected line segments to form a "complete"

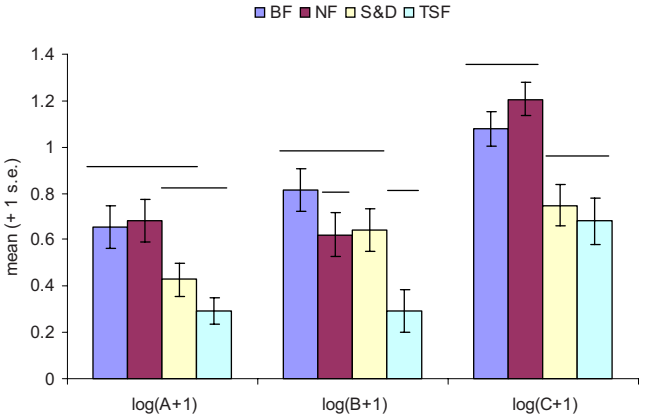


FIG. 8. (Color online) Statistical analysis of tracing results.

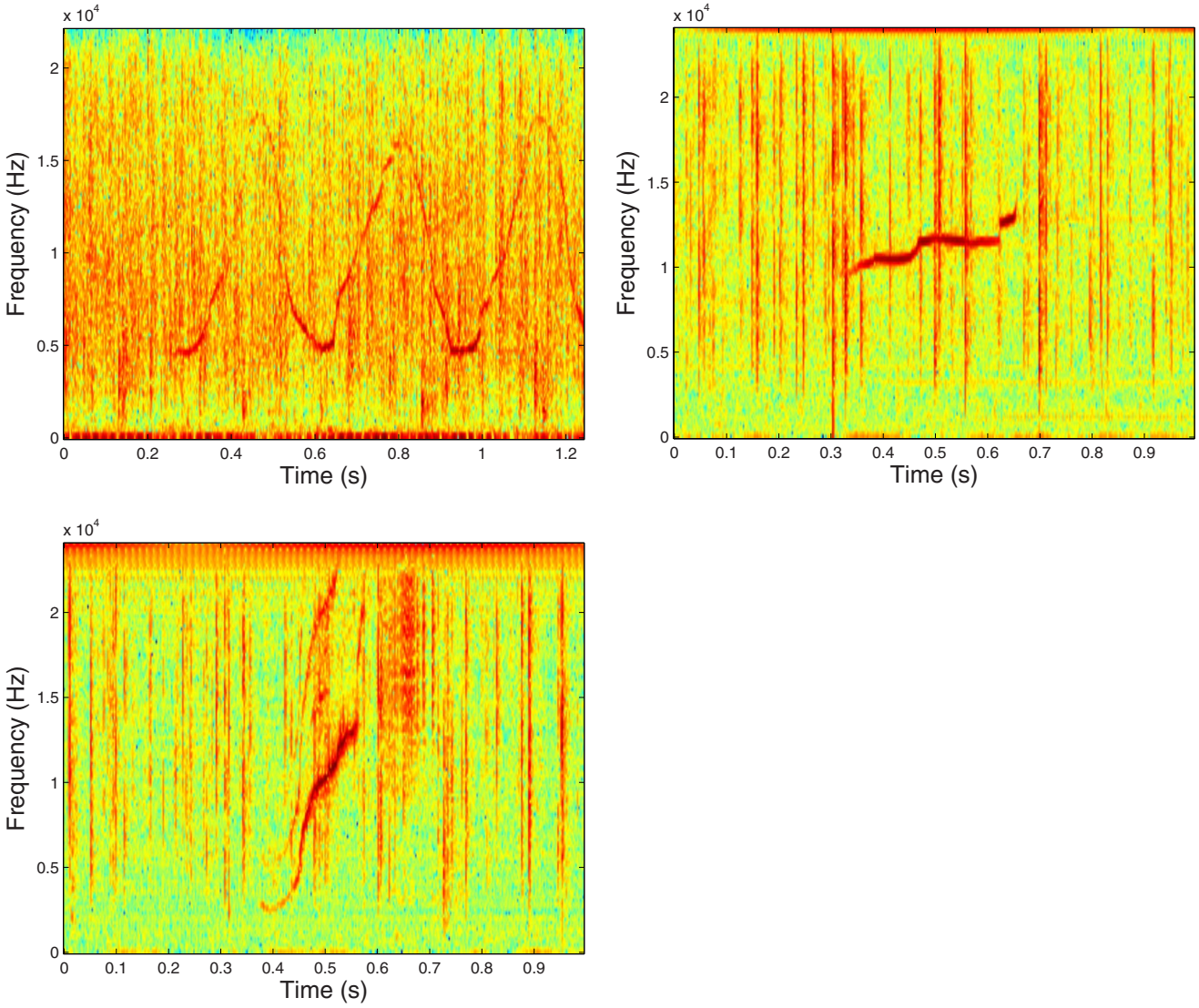


FIG. 9. (Color online) Spectrograms of problematic whistles for tracing.

picture. Therefore the manual trace contains pixels that are actually "breaks" in the whistle but are "artificially" linked together. However, the automated methods fail to identify these pixels as being part of the whistle and this leads to a higher number of missing points for such whistles.

Apart from measuring tracing performance, metrics Φ_B and Φ_C shed light on the values of operational parameters used in the tracing algorithm. For example, a high percentage of missed points Φ_B coupled with a low percentage of extra points Φ_C indicate that the segmentation (or detection) thresholds are probably set too high, and vice versa. Appropriate action can then be taken to determine the optimal point at which both metrics are sufficiently low. This method was followed in order to obtain the best performance for all the algorithms.

This study demonstrates that several different image processing-based approaches can be considered for extracting whistle contours from recordings. It also confirms that TSF performs better for recordings made in warm shallow waters.

VIII. CONCLUSION

This paper introduces an algorithm based on image processing techniques to denoise and extract contours of dolphin whistles from spectrograms. The algorithm presented in Sec. III is well suited for denoising recordings made in warm shallow waters, where the ambient noise is dominated by snapping shrimp. It exceeds the performance of existing algorithms[5,6] by incorporating the pixel values of adjacent time bins for greater preservation of the signal while effectively attenuating transient spectral patterns. The use of static directional smoothing kernels yields more efficient implementations and reduces the processing time of denoising. The objective method introduced for testing the tracing performance using known time-frequency contours not only enables the selection of a particular algorithm, but can also be used to tune its parameters.

The modularity of the algorithm enables easy integration of other techniques at any stage of the process, and further studies may be carried out to examine combinations that produce the best results. Although an effort has been made to

make the algorithm self-adaptive as far as possible, some parameters still have to be manually set depending on any particular acoustic recording. However, suggestions are made to guide the choice of the values of those parameters. Features extracted from whistle contours are available for classification tasks, paving the way for faster analysis of dolphin vocalizations. Although this implementation has not yet been optimized for speed, and currently works on off-line data, an on-line system coupled with a whistle recognition module would be an invaluable tool for field biologists. We believe the techniques presented here can be applied to extract other animal vocalizations such as whale calls and bird songs from acoustic recordings. Further studies are recommended to determine the feasibility of such extensions.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Gao Rui and Suranga Chandima Nanayakkara of the Department of Electrical and Computer Engineering, NUS, Dr. Sin Tsai Min of TMSI, and Yeo Kian Peen of MMRL, for their assistance and encouragement at various stages, including recording of dolphin whistles, manual whistle tracing, and statistical analysis.

[1]V. M. Janik, "Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods," Anim. Behav. **57**, 133–143 (1999).
[2]S. Datta and C. Sturtivant, "Dolphin whistle classification for determining group identities," Signal Process. **82**, 251–258 (2002).
[3]J. R. Buck, H. B. Morgenbesser, and P. L. Tyack, "Synthesis and modification of the whistles of the bottlenose dolphin, tursiops truncatus," J. Acoust. Soc. Am. **108**, 407–416 (2000).
[4]J. R. Buck and P. L. Tyack, "A quantitative measure of similarity for tursiops truncatus signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506 (1993).
[5]C. Sturtivant and S. Datta, "Techniques to isolate dolphin whistles and other tonal sounds from background noise," Acoust. Lett. **18**, 189–193 (1995).
[6]C. Sturtivant and S. Datta, "The isolation from background noise and characterisation of bottlenose dolphin (tursiops truncatus) whistles," J. Acoust. Soc. India **23**, 199–205 (1995).
[7]M. Chitre, J. Potter, and S. H. Ong, "Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise," IEEE J. Ocean. Eng. **31**, 497–503 (2006).
[8]C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE International Conference on Computer Vision, Bombay, India*, Jan. 4–7, 1998.
[9]W. Niblack, *An Introduction to Digital Image Processing* (Prentice Hall, Englewood Cliffs, NJ, 1986).
[10]F. Yan, H. Zhang, and C. R. Kube, "A multistage adaptive thresholding method," Pattern Recogn. Lett. **26**, 1183–1191 (2005).
[11]R. E. Kalman, "A new approach to linear filtering and prediction problems," ASME J. Basic Eng. **82**, 35–45 (1960).

Experimental investigation of the influence of a posterior gap on glottal flow and sound

Jong Beom Park[a] and Luc Mongeau[b]
Department of Mechanical Engineering, McGill University, 817 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada

(Received 3 July 2007; revised 14 April 2008; accepted 15 May 2008)

The influence of a posterior gap on the airflow through the human glottis was investigated using a driven synthetic model. Instantaneous orifice discharge coefficient of a glottal shaped orifice was obtained from the time-varying orifice area and the velocity distribution of the pulsated jet measured on the axial plane using a single hot-wire probe. Instantaneous orifice discharge coefficient values were found to undergo a cyclic hysteresis loop when plotted versus Reynolds number and time, indicating a pressure head increase and a net energy transfer from the air flow to the orifice wall. The net energy transferred was estimated to be around 10% of the value presumably required to achieve self-sustained oscillation. The radiated sound pressure was measured to characterize the influence of the minimum flow through the posterior gap on the broadband component of the radiated sound. The presence of a posterior gap was found to significantly increase the broadband sound level produced over the frequency range in which human hearing is most sensitive. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945116]

PACS number(s): 43.70.Bk, 43.70.Aj, 43.28.Ra [AL] Pages: 1171–1179

I. INTRODUCTION

Previous studies of normal speakers as well as dysphonic patients have shown that normal voice production does not necessarily involve complete vocal fold closure. Incomplete closure of the vocal folds near the posterior glottal wall, referred to as a *posterior (glottal) chink* or as a *posterior gap*, is commonly found in normal phonation, especially in women (Biever and Bless, 1989; Bless *et al.*, 1986; Koike and Hirano, 1973; Peppard *et al.*, 1988). The posterior gap is the result of an incomplete closure of the posterior cartilaginous portion of the glottis. This incomplete adduction of the vocal folds can be caused by low muscular tension or, for pathological voices, by lesions such as vocal fold nodules and polyps, or else by functional dysphonia of the arytenoid muscles (Belisle and Morrison, 1983; Morrison *et al.*, 1983).

Air leakage through the posterior gap occurs during the closed phase of the phonatory cycle. The steady air flow through the gap, referred to as minimum flow, intermittently merges with the pulsated jet emanating from the vibrating membranous portion of the glottis. The resulting flow field is generally turbulent and involves merging of two independent jet streams, which was postulated to generate significant broadband radiated sound. Clinical studies of the influence of a posterior gap have reported that the voice quality inherently features breathiness (Fritzell *et al.*, 1986; Hammarberg *et al.*, 1984) and that the glottal gap can extend into the posterior part of the folds, preventing complete closure of the membranous folds (Holmberg *et al.*, 1988). Quantitative investigations of the minimum flow through the posterior gap *in vivo* have been hampered by the difficulty of making direct and precise measurements of flow parameters in a clinical setting. There have been few *in vitro* experimental and theoretical stuides of the effects of the posterior gap on voice production dynamics (Cranen and Schroeter, 1995). *In vitro* experiments in controlled conditions using a replica are useful for basic quantitative studies of the relation between glottal flow rate and transglottal pressure. The accuracy of the flow parameters obtained depends, of course, on the realism of the replica geometry.

In the present study, the air flow through a physical replica of human vocal folds with a posterior gap was investigated using hot-wire anemometry. One goal was the determination of the instantaneous orifice coefficient and comparisons with comparable quantities for stationary orifices. The results provided accurate flow resistance data, allowed the importance of unsteady effects to be assessed, and characterized the influence of the posterior gap on laryngeal flow resistance. The broadband sound emissions were measured and compared with those of comparable orifices with no gap.

II. EXPERIMENTAL METHODS

A forced-oscillation rubber model of the vocal folds was employed to allow external independent control of the glottal flow rate. The vocal fold replica and the experimental setup were the same as those used in previous studies (Park and Mongeau, 2007; Zhang *et al.*, 2002).

A. Apparatus

The synthetic model was fabricated using a commercially available silicone compound, GE RTV-11. A converging orifice profile in the coronal plane was used as a base line. The converging model was chosen over a diverging model because its flow is stable and symmetric (Erath and Plesniak, 2006; Park and Mongeau, 2007). The orifice in-

[a]Electronic mail: jong.b.park@mail.mcgill.ca
[b]Electronic mail: luc.mongeau@mcgill.ca

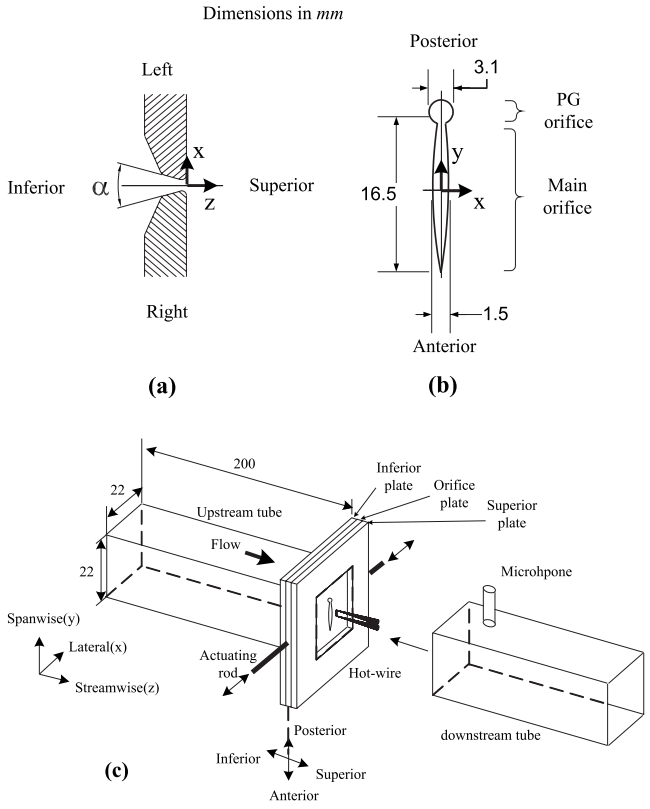


FIG. 1. Schematic of the experimental apparatus: (a) orifice included angle, $\alpha=30°$, (b) geometry of the orifice of the PG model, (c) apparatus and orientation of the coordinate system. Dimensions in millimeters.

cluded angle and the length of vocal folds were 30° and 16.5 mm, respectively. To create a posterior gap, a 3.18 mm diameter hole was made at the posterior commissure of the base line model. The base line model and the model with the posterior gap are referred to as the "NPG model" and the "PG model," respectively. The posterior gap and the remaining vibrating portion of the orifice are denoted as "PG orifice" and "main orifice," respectively, as shown in Fig. 1. The jets emanating from each orifice are referred to as the "PG jet" and the "main jet," respectively. The PG jet during closure of the main orifice constitutes the "minimum flow."

The synthetic model orifice plate was sandwiched between two rectangular aluminum plates, with 2.2×2.2 cm^2 rectangular openings at their center, as shown in Fig. 1. A matching $2.2\times2.2\times20$ cm^3 Plexiglas tube located upstream of the orifice was attached to the inferior plate to simulate the trachea. In most experiments, the orifice discharge was open to atmosphere to facilitate velocity measurements; in some cases, a tube of the same dimensions as those of the upstream tube was added to play the role of the vocal tract for radiated sound pressure measurements.

Two electrodynamic linear actuators, Labworks ET-126, were used to drive the synthetic model. The transglottal time-averaged pressure was measured using a differential pressure transducer, MKS Baratron model 100T. The radiated sound was measured using a B&K 4939 microphone, located 3 cm downstream from the superior plate. A single hot-wire probe, TSI 1210, mounted on a three-dimensional linear traverse was used to measure the axial velocity of the glottal jet in the x-y plane (referred to as the *measurement plane*) located 1 mm downstream from the rubber orifice. A digital high speed camera, NAC FX-K3, and a photoelectric sensor were used for measuring the glottal opening area.

B. Experimental procedures

The orifice area, the flow velocity, and the flow rate were directly measured. The orifice area was calculated from pixelwise area integration of digital images captured by the high speed camera at a rate of 6000 frames/s. The orifice area was correlated with the light intensity measured by a photoelectric sensor located inside the upstream tube. The time-varying area, $A_o(t)$, was then subsequently obtained online from measuring the photoelectric detector signals simultaneously with the flow velocity.

The axial velocity distribution over the cross sectional area of the jet plume, i.e., the *velocity map* $u_z(x,y)$, was obtained by scanning the measurement plane with the hot-wire probe (Park and Mongeau, 2007). The spatial resolution was 0.2 mm along x and 1.5 mm along y. The acquisition of the anemometer signal at each measuring point was phase locked with respect to the actuator excitation signal. The velocity map recordings were then assembled to construct the time history of the glottal flow velocity distribution, $u_z(x,y,t)$.

The mean time-averaged flow rate, Q_m, was obtained using an electronic mass flow meter located upstream of the test section. The instantaneous flow rate, $Q(t)$, was obtained through the numerical integration of the velocity map. The mean flow rate was used for calculating the discharge coefficient for stationary orifices and for verifying the accuracy of the time-averaged value of flow rate obtained from the numerical integration of the velocity maps, $Q(t)=\int_{\mathbf{x}}\{u_z(\mathbf{x},t)>u_{\text{cut}}\}\cdot d\mathbf{x}$. A cutoff velocity threshold value, u_{cut}, was established based on a steady flow calibration to reduce the influence of artifacts that tend to bias the flow rate, such as nonzero velocity fluctuation in the absence of flow or flow entrainment from the jet shear layers. The value of u_{cut} was determined by matching $\overline{Q(t)}$ with Q_m, where both flow rates were measured using a nonpulsated stationary orifice. It was found that a fixed u_{cut} value yielded acceptable accuracy when it was applied to the time-varying orifice. A typical estimation error, $|(Q_m-1/T\int Q(t)dt)/Q_m|$, was less than 5% using $u_{\text{cut}}=4$ m/s for a time-averaged transglottal pressure of 10 cm H_2O. Note that this procedure may not be necessarily applicable to the case of self-oscillating synthetic models due to possible asymmetry, larger jet flow angles, out-of-plane model deformations, and reduced controllability of the model vibration leading to abrupt transitions of the glottal flow.

The PG model was driven using a sinusoidal input signal at a frequency of 100 Hz. The time-averaged transglottal pressure was set to 10 cm H_2O for the unsteady glottal jet configuration. This operating condition was the same as that used for the NPG model in a previous study (Park and Mongeau, 2007). Time-varying instantaneous orifice discharge coefficients, C_d^i, for the vibrating orifice were calculated from the velocity map and orifice area using

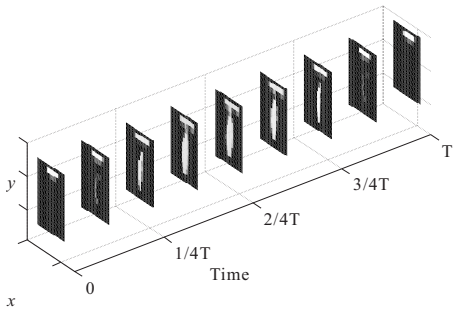


FIG. 2. Temporal sequence of velocity maps for the unsteady PG model. $f=100$ Hz, $\nabla p=10$ cm H_2O, and $Q_m=3.21\times10^{-4}$ m^3/s.

$$C_d^i(t_j,\mathrm{Re}) = \left.\frac{Q(t)}{A_o(t)u_c(t)}\right|_{t=t_i}, \tag{1}$$

where $u_c(t)$ is the flow velocity at the center of the orifice and Re is Reynolds number, defined as $u_c\sqrt{A_o}/\upsilon$. Stationary discharge coefficients, C_d^{st}, were also measured for comparisons with C_d^i values. Independent experiments were performed, with the PG model held fixed, over a range of steady opening areas and flow rates that spanned those recorded in the experiments for the dynamic case. The coefficients for stationary orifices were then computed as

$$C_d^{\mathrm{st}}(\mathrm{Re}) = \frac{Q_m}{A_o u_c}. \tag{2}$$

The radiated sound pressure was measured for cases with time-varying orifices for both NPG and PG models. The influence of downstream confinement was investigated in each case by fitting a tube downstream of the superior plate. The sound measurements were performed separately from the other parameters being measured to avoid possible probe interference effects.

III. RESULTS

A. Aerodynamic features of the glottal jet

A sequence of velocity maps (Fig. 2) shows the temporal evolution of the glottal jet plume within the measurement plane. The main jet merged with the PG jet approximately at $t\sim T/4$ during the opening phase and separated from the PG jet at $t\sim 3T/4$ during the closing phase. The discharged flow rate wave form was periodic and symmetric over one period, as shown in Fig. 3, as for the case of the NPG model from previous studies. The time-averaged flow rate estimated from the velocity map integration, $\overline{Q(t)}=3.24\times10^{-4}$ m^3/s, was equal to the value from the mass flow meter, $Q_m=3.21\times10^{-4}$ m^3/s, for the cutoff value $u_{\mathrm{cut}}=4$ m/s. The minimum flow rate through the PG orifice ($A_{\mathrm{PG}}\sim3.7$ mm^2) was about 37% ($Q_{\mathrm{PG}}=1.23\times10^{-4}$ m^3/s) of the total flow rate. The orifice discharge coefficient of the PG orifice is approximately $C_{d,\mathrm{PG}}=Q_{\mathrm{PG}}/(A_{\mathrm{PG}}\cdot\sqrt{2\nabla p/\rho})\sim0.8$, which is very similar to (slightly smaller than) the orifice discharge coefficient of a converging orifice (NPG) model (Park and Mongeau, 2007; Zhang *et al.*, 2002).

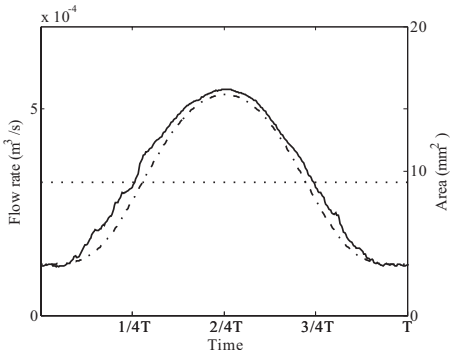


FIG. 3. Instantaneous flow rate and area versus time: PG model, $f=100$ Hz, $\nabla p=10$ cm H_2O, and $Q_m=3.21\times10^{-4}$ m^3/s;—: $Q(t)$;---: $A_o(t)$;---: Q_m.

The presence of the minimum flow requires a modification in the calculation of the discharge coefficients for the PG model. Because only the PG orifice discharges air flow during the *closed* phase, when the main orifice is completely closed, it no longer makes sense to use the flow velocity measured at the center of the main orifice, $u_c(=u_{c,\mathrm{main}})$, in the calculation of the instantaneous coefficient. One may use the PG center velocity, $u_{c,\mathrm{PG}}$, because it should be equal to $u_{c,\mathrm{main}}$ from Bernouilli's equation. However, $u_{c,\mathrm{PG}}$ was slightly smaller than $u_{c,\mathrm{main}}$ over $T/4<t<3T/4$, as shown in Fig. 4(a). The small discrepancy is believed to have been due to the fact that wall deformations during the forced motion tilted the PG jet slightly upward. As discussed in Park and Mongeau (2007), single wire probes measure the velocity magnitude and are not very sensitive to flow direction in any plane containing the wire. The maximum difference between

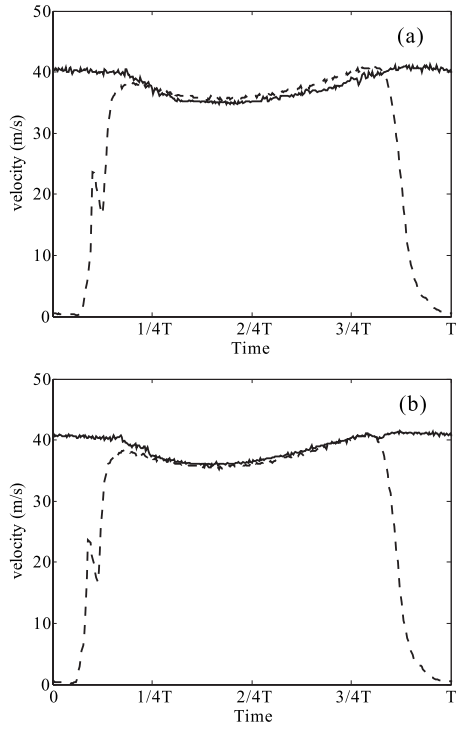


FIG. 4. Comparison of center flow velocities for the PG and main jets: PG model, $f=100$ Hz, $\nabla p=10$ cm H_2O, $Q_m=3.21\times10^{-4}$ m^3/s. (a)—: u_c,PG;---: $u_{c,\mathrm{main}}$, (b)—: $u_{\max}$;---: $u_{c,\mathrm{main}}$.

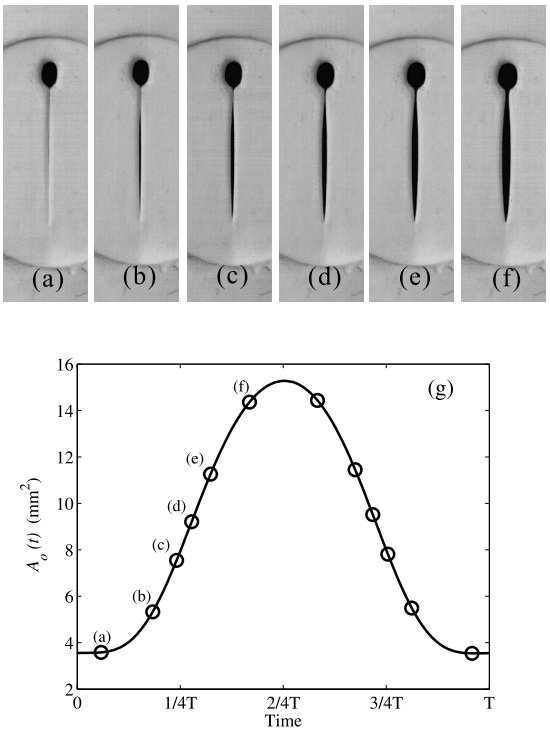


FIG. 5. Stationary PG model with area values of (a) 3.7 mm^2, (b) 5.3 mm^2, (c) 7.6 mm^2, (d) 9.3 mm^2, (e) 11.3 mm^2, and (f) 14.4 mm^2. These were used to obtain the stationary orifice discharge coefficients, C_d^{st}. Each opening area is indicated on the area function, $A_o(t)$, of the unsteady PG model in (g) to show the approximate time each area corresponds to during one cycle.

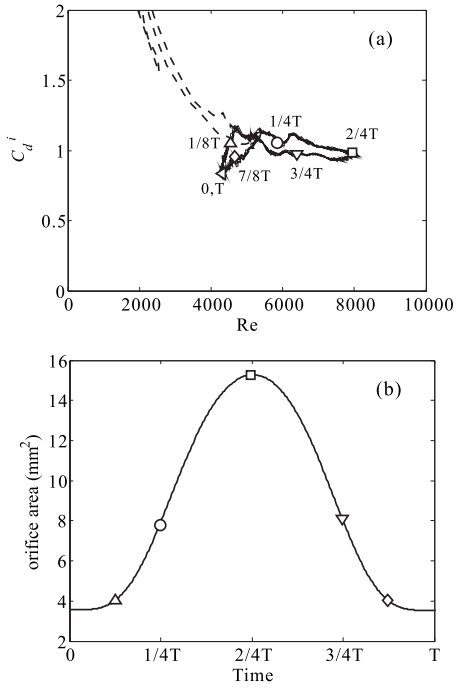


FIG. 6. Instantaneous orifice discharge coefficient of the unsteady PG model vs Reynolds number: $f=100$ Hz, $\nabla p=10$ cm H_2O, and $Q_m=3.21\times10^{-4}$ m^3/s. (a) The curve was calculated for several cycles with key time values denoted by open symbols. The two line types show the effects of employing different jet center velocities (—: $u_c=u_{max}$ and---: $u_c=u_{c,main}$) for the C_d^i calculation. (b) Orifice area, $A_o(t)$, of the PG model. Key time values are indicated with open symbols, △: $t=1/8T$; ○: $t=T/4$; □: $t=T/2$; ▽: $t=3/4T$; ◇: $t=7/8T$.

$u_{c,PG}$ and $u_{c,main}$ was approximately 2 m/s at $t\sim3/4T$. This is only slightly larger than the difference estimated from the cosine of the jet plume angle with respect to the axial direction, which was 10° at most, but small with respect to the overall velocity magnitude (<5%). To correct the problem, the maximum value of the glottal jet at every time instant,

$$u_{max}(t)=\max\{u(\mathbf{x},t):\mathbf{x}\subset(x,y) \text{ for all } x,y \text{ in measurement plane}\} \quad (3)$$

was used in place of u_c in Eq. (1). The resulting velocity, u_{max}, is shown in Fig. 4(b). Note that the maximum flow velocity is now the same for the PG and the main jets, i.e., $u_{c,main}$ and $u_{c,PG}$ over most of the cycle. The use of this modified velocity in the orifice coefficient calculations helped compensate for the inaccuracies associated with the skewed jet angle.

B. Orifice discharge coefficients

The areas of the stationary orifices for which C_d^{st} were obtained are shown in Fig. 5. Steady flow measurements were performed for transglottal pressures varied from 0 to 20 cm H_2O in increment of 1 cm H_2O. The maximum velocity was used for the C_d^{st} calculation in the same way as for the C_d^i calculation. It was found that $u_{max}=u_{c,main}\sim u_{c,PG}$ for $A_o\geqslant5.3$ mm^2 [Figs. 5(b)–5(f)] and $u_{max}=u_{c,PG}$ for $A_o=3.7$ mm^2 [Fig. 5(a)], which is consistent with the jet centerline velocity differences described in Sec. III A.

The calculated C_d^i values are shown for a few complete cycles using solid lines in Fig. 6(a). Open symbols indicate key instants of time within the cycle; the corresponding area values are shown in Fig. 6(b). The dotted line in Fig. 6(a) shows the C_d^i values obtained using $u_c=u_{c,main}$ instead of $u_c=u_{max}$ for the calculation. Note that the coefficient values based on $u_{c,main}$ diverge at low Reynolds numbers due to the vanishing magnitude of $u_{c,main}$ during the closed phase. The time history of C_d^i values based on u_{max} is cyclic, which was also observed in the case of the unsteady NPG model from a previous study (Park and Mongeau, 2007). However, C_d^i values for the PG model exhibit a distinct hysteretic behavior, which was not the case for the unsteady NPG model. The values of C_d^i are larger during the opening phase than during the closing phase. They asymptotically approach 0.8 at $t\sim0$ or T and unity at $t\sim2T/4$. The first asymptotical value of 0.8 indicates that the discharge coefficient of the PG orifice only, and the second value of 1 indicates that of the entire orifice. A sharp transition in C_d^i values occurred at $T/8<t<T/4$ and $3T/4<t<7T/8$, which corresponds to the time intervals where the location of the maximum velocity of the glottal jet changed from within the PG jet to within the main jet, or vice versa (Sec. III A). These trends are most likely the consequence of the merging of the PG and the main jets, as will be discussed further in Sec. IV A.

To evaluate the quasisteady approximation, C_d^i values are compared to the values of C_d^{st} in Fig. 7 for the same orifice area values shown in Fig. 5. A 5% tolerance was allowed on the orifice areas for computing the C_d^i values. Figures 7(b)–7(f) show that the C_d^{st} values are in good agreement with C_d^i values over both the opening and closing

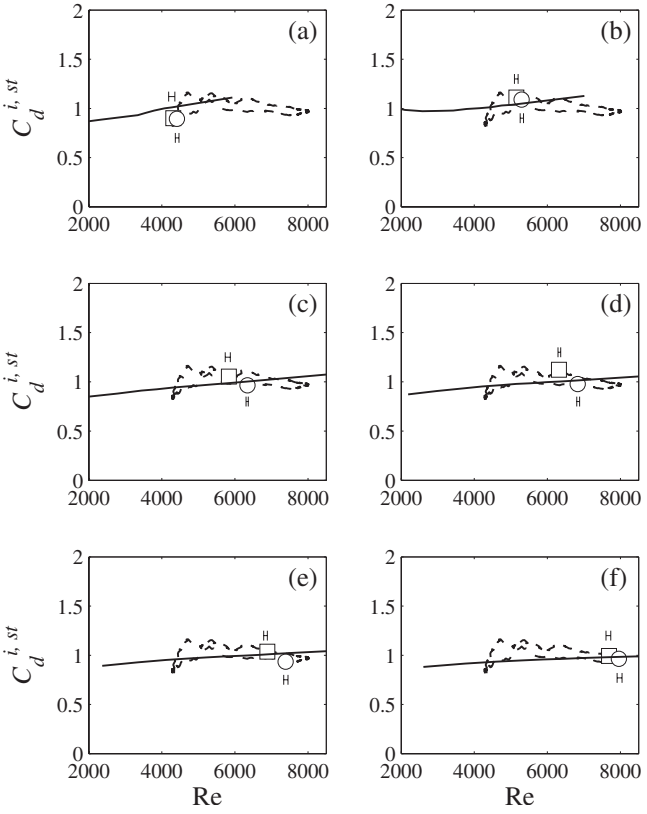


FIG. 7. Orifice discharge coefficients of unsteady (---: C_d^i) and steady (—: C_d^{st}) PG models for the orifice areas of Fig. 5: (a) 3.7 mm^2, (b) 5.3 mm^2, (c) 7.6 mm^2, (d) 9.3 mm^2, (e) 11.3 mm^2, and (f) 14.4 mm^2. The instantaneous coefficient is indicated for opening (□) and closing (○) phases for each area. Error bars above and below the symbol indicates the range of C_d^i computed with a ±5% tolerance on the orifice area.

phases, which implies that unsteady effects associated with wall motion and flow acceleration can be regarded as negligible. For the closed phase, as shown in Fig. 7(a), the C_d^i values were smaller than the C_d^{st} values. Such trend was also observed in the case of the NPG model where the C_d^i values were smaller than the C_d^{st} values for very small orifice areas, when the orifice was almost closed. It appears that unsteady effects did become significant for the NPG orifice near the time of complete closure, or flow stream discontinuity, due presumably to the influence of the displacement flow and viscosity. Similar trends were observed for the PG model, despite the absence of sharp discontinuities in the flow stream due to the presence of the minimum flow. Flow displaced or pumped by the motion of the walls may thus be the dominant cause of intrinsically nonsteady behavior in this case.

C. Influence of posterior gap on radiated sound

It is generally anticipated that the presence of the posterior gap would cause an increased level of broadband sound, which is one feature of breathy voices. One-third octave band (Beranek and Vér, 1992) radiated sound spectra were measured to quantify the influence of the posterior gap on the perceived broadband noise levels underlying the desired tonal voice signature. Frequency spectra in the presence of

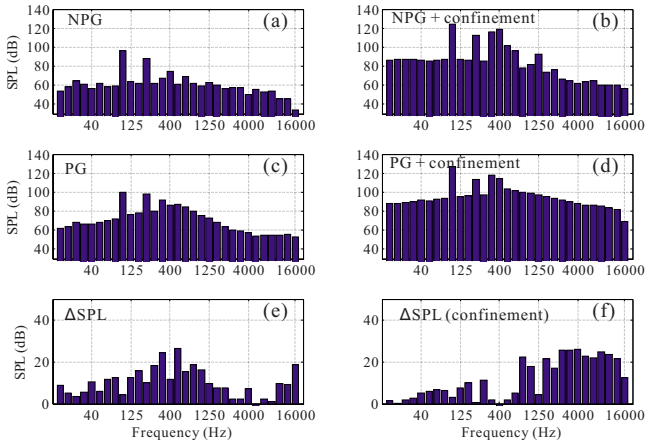


FIG. 8. (Color online) One-third octave band sound pressure levels for unsteady jets: (a) NPG with no confinement, (b) NPG with confinement, (c) PG with no confinement, (d) PG with confinement, (e) spectral level increase due to the presence of PG jet with no confinement [i.e., (c)–(a)], and (f) spectral level increase due to the presence of PG jet with confinement [i.e., (d)–(b)].

downstream confinement were also measured to study the influence of the vocal tract on the broadband sound radiation.

Figures 8(a) and 8(c) show radiated sound spectra for the NPG and PG models with no confinement, respectively. Dominant tonal components at $80 < f < 400$ Hz are associated with the fundamental pulsating frequency and its harmonics. The differences between the underlying broadband levels for both configurations were computed and are shown in Fig. 8(e). The spectral levels were significantly increased over the range $125 < f < 4000$ Hz. When a downstream confinement was added for both NPG and PG models, it caused the overall sound levels to increase, as shown in Figs. 8(b) and 8(d). This well known phenomenon is because of the increase in acoustic radiated power per unit area caused by an increase in radiation efficiency for ducted jets. Considering that the source is a compact dipole, local flow kinetic energy is more effectively converted into acoustic energy for a planar wave duct than for radiation in an open space (Zhang *et al.*, 2004). In the presence of confinement, the spectral level difference between NPG and PG models is distributed over the frequency range $400 < f < 16\,000$ Hz, as shown in Fig. 8(f). This includes the range over which the human ear is most sensitive, i.e., 1–4 kHz (Zwicker and Fastl, 1999). The presence of the PG jet or any substantial changes in the posterior gap may thus be readily perceivable, which may in part explain the breathy voice quality of subjects with posterior gaps. It is presumed that the posterior gap, together with the downstream confinement of the vocal tract, enhances glottal jet mixing, leading to increased broadband sound production over a frequency band whose wavelengths are commensurate with the size of the posterior gap and also the supraglottal vocal tract diameter. A correlation between the morphologic structure of the posterior gap and the resulting broadband spectral characteristics could be developed in the future and would perhaps allow the remote estimation of the size of the posterior gap.

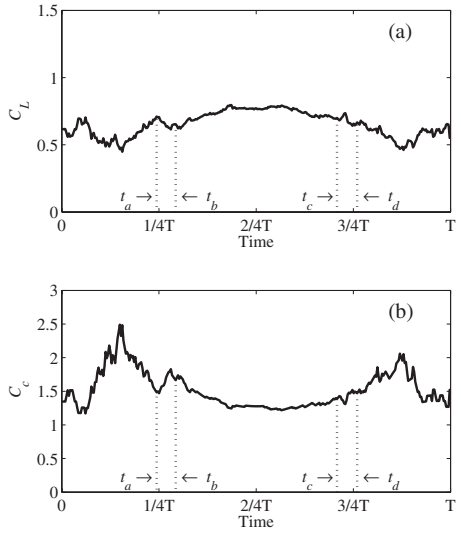


FIG. 9. (a) Loss coefficient, C_L, and (b) contraction coefficient, C_C, vs time (PG model, $f=100$ Hz, $\nabla p=10$ cm H_2O, $Q_m=3.21\times10^{-4}$ m^3/s). Each vertical line denotes the approximate time when the two orifice domains area connected ($t_a<t<t_b$) and disconnected ($t_c<t<t_d$) during the opening and closing phases.

IV. DISCUSSION

A. Hysteresis of the instantaneous orifice discharge coefficient

To explain the hysteresis phenomenon, it is useful to divide the orifice coefficient into the product of the loss coefficient, C_L, and the contraction coefficient, C_C. The loss coefficient captures the difference between the estimate of the flow rate assuming a plug flow, i.e., the product of the center flow velocity and the orifice area, and the actual flow rate. The contraction coefficient is the ratio of the cross sectional flow stream area, A^M, evaluated on the measuring plane to the orifice area. It quantifies the pressure variation and viscous loss along the stream line (Park and Mongeau, 2007).

The loss coefficient, C_L, shown in Fig. 9(a) undergoes a temporally symmetric transition over the period, with the exception of a transient decrease at $t_a<t<t_b$. The contraction coefficient, C_C, on the other hand, is slightly larger during the opening phase with a transient increase over $t_a<t<t_b$. This asymmetric variation of C_C causes the hysteresis, implying that pressure rise or viscous loss is greater during the opening phase.

The period of time, $t_a<t<t_b$, corresponds to the moment when the PG and the main orifices coalesced into one single continuous domain during the opening phase. The orifice shapes before and after the domain connection are shown in Figs. 10(b) and 10(c). This orifice transformation caused the PG jet and the main jet to merge. Jet merging can also be observed from the change in velocity profile over the region $3<y<5$ mm between $t=t_a$ (dashed line) and t_b (solid line) from Fig. 11, which shows the velocity profile along the anterior-posterior line running through the center of the orifice ($x=0$, $-10\leqslant y\leqslant 10$).

The change in C_C and C_L during the jet merging indicates a transient variation in properties of the glottal jet. The abrupt increase in C_C during the jet merging period corresponds to the increase in A^M at a steeper rate than that of the orifice area, as shown in Fig. 10(a). It suggests that the jet merging brings about a sudden increase in the pressure head along the stream. A greater value of C_L at $t=t_a$ implies that the velocity distribution is more uniform at $t=t_a$ than at $t=t_b$ because the flow rate is closer to that of an ideal flow with uniform velocity over the jet cross sectional area. The distribution at $t=t_a$ in Fig. 11 is, however, disrupted more severely at 3 mm$<y<$5 mm and may therefore seem less uniform than the distribution at $t=t_b$. The distribution at $t=t_a$, however, occurs at the time when the two orifice domains are not connected [Fig. 10(b)] and does not reduce an overall uniformity of the velocity distribution. The local drop at 3 mm$<y<$5 mm in the velocity profile at $t=t_b$, on the other hand, occurs over the connected orifice domain and therefore leads to a less uniform velocity profile over the orifice area, resulting in a lower value of C_L. During the closing phase, no significant change in C_C and C_L was observed when the glottal jet was separated into two distinct plumes at $t_c<t<t_d$.

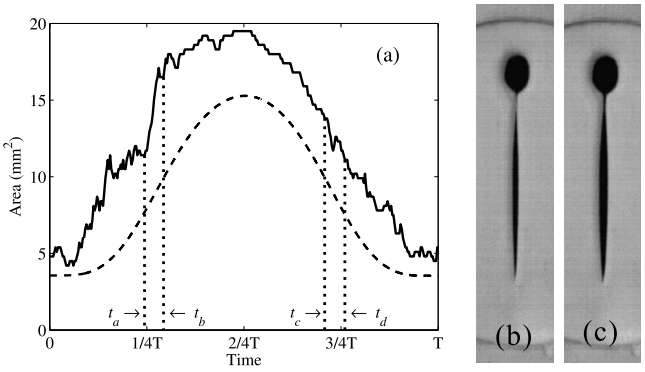


FIG. 10. Jet stream cross sectional area variation vs time: (a)—: stream area, A^M;–––: orifice area, A_o. Each vertical line denotes the approximate time when the two orifice domains area connected ($t_a<t<t_b$) and disconnected ($t_c<t<t_d$) during the opening and closing phases. (b) Orifice shape at $t=t_a$ and t_d. (c) Orifice shape at $t=t_b$ and t_c.

B. Estimation of energy transfer to the physical model

The rate of energy transfer from the glottal flow to the vocal folds, i.e., the power extracted from the flow field, can

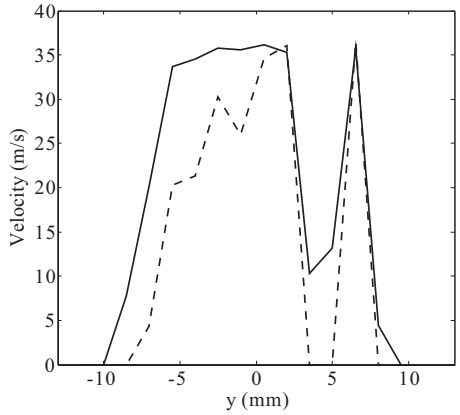


FIG. 11. Jet velocity profile in the anterior-posterior direction at the center of the orifice before and after jet merging (—: $t=t_a$;—: $t=t_b$). The main orifice and the PG orifice are located approximately at -9 mm$<y<$4 mm and 4 mm$<y<$8 mm, respectively.

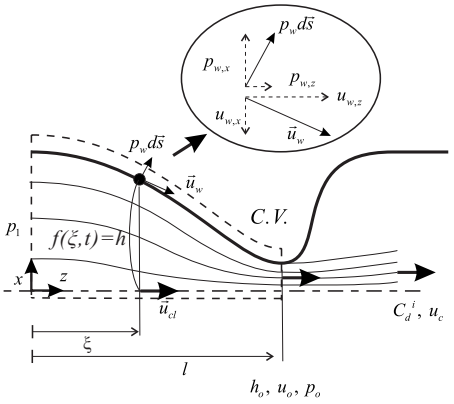


FIG. 12. Diagram of the glottal flow domain (sub- and supraglottal space in the coronal plane). The glottal wall geometry is defined by $h=f(z,t)$, where h is the distance (height) from the line of symmetry to the wall. The orifice throat (of height h_o) is located at the point of minimum area $z=l$.

be estimated from the empirical C_d^i data, provided that some simplifying assumptions on the flow field and the orifice motion are made. The energy transfer occurring at the inferior surface of the vocal folds is assumed to be most significant, and hence only the upstream domain of the physical model is taken into account for the estimation. Let the vocal fold wall shape be described by a function of streamwise distance and time, $f(z,t)$. This orifice shape function $f(z,t)$ represents the height h, of the vocal fold wall from the line of symmetry and spans from the upstream end of the vocal fold ($z=0$) to the orifice throat of the minimum height, h_o at $z=l$, with a unit depth, d, as depicted in Fig. 12. The flow field inside the control volume is assumed to be inviscid, irrotational, and incompressible; i.e., a potential flow is assumed with the additional condition of no viscous loss, which implies that the whole flow domain has a constant total pressure head, or *Bernoulli constant*. Work done by the pressure on the wall surface is considered much larger than work done by the shear stress on the surface or by the adverse pressure gradient at the flow separation point downstream of the orifice. Therefore the effects of turbulence and boundary layer are not considered in the estimation.

At the orifice throat ($z=l$) the glottal flow is assumed to have a uniform velocity profile, $u_o(t)$, which can be obtained by imposing the actual flow rate, $Q(t)$, measured downstream as

$$u_o(t) = \frac{Q(t)}{A_o(t)} = C_d(t) \cdot u_c(t). \tag{4}$$

The pressure at the orifice throat, $p_o(t)$, is then obtained from Bernoulli's equation with the measured inlet subglottal pressure, p_1, assuming that the inlet velocity is small, $u_1(t) \ll u_o(t)$,

$$p_o(t) = p_1 - \tfrac{1}{2}\rho C_d^2(t) u_c^2(t). \tag{5}$$

The orifice height at the throat is also a function of time and may be approximated by the effective radius of the measured orifice area function, $A_o(t)$, as

$$h_o(t) = f(z,t)|_{z=l} = \sqrt{A_o(t)/\pi}. \tag{6}$$

Note that a unit depth, d, in the y-direction is assumed such that the cross sectional area inside the control volume can be calculated as $A(z,t)=df(z,t)$ for $0 \leqslant z \leqslant l$.

It is assumed that the flow has a uniform streamwise velocity distribution, and the flow rate is only time dependent, i.e., constant at any cross section within the control volume $[Q(z,t)=Q(t)=dh_o(t)u_o(t)]$, from the conservation of mass flux. The flow velocity along the centerline, i.e., along the line of symmetry, is expressed as

$$u_{\mathrm{cl}}(\xi,t) = \left.\frac{Q(t)}{dh}\right|_{z=\xi} = \left.\frac{h_o(t)u_o(t)}{f(z,t)}\right|_{z=\xi} \tag{7}$$

at any point $z=\xi$, and is then equated to the streamwise velocity component of the wall flow velocity,

$$u_{w,\xi}(\xi,t) = u_{\mathrm{cl}}(\xi,t), \tag{8}$$

from the assumption of uniform streamwise velocity. Because the wall flow velocity can be expressed using the tangential vector for the known wall geometry,

$$\mathbf{u}_w(\xi,t) = (u_{w,z}, u_{w,x}) = \left.\frac{|\mathbf{u}_w(z,t)|}{\sqrt{1+f'(z,t)^2}}[1, f'(z,t)]\right|_{z=\xi}, \tag{9}$$

where the prime denotes $\partial/\partial z$, the dynamic pressure on the wall is obtained from Eqs. (7)–(9) as

$$\frac{\rho}{2}|\mathbf{u}_w(\xi,t)|^2 = \left.\frac{\rho}{2}\left(\frac{\sqrt{1+f'(z,t)^2}}{f(z,t)}h_o(t)u_o(t)\right)^2\right|_{z=\xi}. \tag{10}$$

By using Eq. (10) and Bernoulli's equation, the pressure at the wall is obtained as

$$p_w(\xi,t) = \left.\frac{\rho}{2}u_o(t)^2\left[1 - \frac{1+f'(z,t)^2}{f^2(z,t)}h_o(t)^2\right] + p_o(t)\right|_{z=\xi}. \tag{11}$$

No work is done in the z-direction because there is no orifice motion in that direction. Only the power in the x-direction is significant and can be calculated as

$$\Pi(t) = \int_0^l \tilde{p}_{w,x}(\xi,t)\frac{\partial}{\partial t}f(\xi,t)d\xi \tag{12}$$

where $\tilde{p}_{w,x}$ is the vertical component of the wall force per unit length (N/m),

$$\tilde{p}_{w,x}(\xi,t) = \left.\frac{p_w(z,t)d}{\sqrt{1+f'(z,t)^2}}\right|_{z=\xi}. \tag{13}$$

Finally, the net power transferred to the wall is obtained as

$$\Pi_{\mathrm{net}} = \frac{1}{T}\int_0^T \Pi(t)dt. \tag{14}$$

An orifice geometry having a triangular shape with a negative slope, $-k$ ($k>0$), was employed to estimate the net power transfer for the current driven converging model. The shape function is expressed as

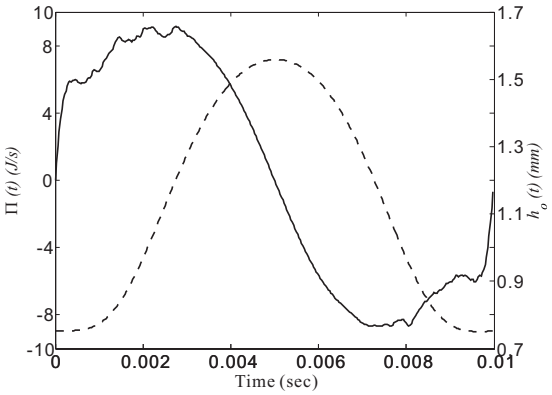


FIG. 13. Rate of energy transfer to the orifice model surface over one period.—: $\Pi(t)$;---: $h_o(t)$. The root mean squared value of the instantaneous power, Π_{rms}, is 6.7 J/s and the net power is 0.14 J/s (~2% of Π_{rms}).

$$f(z,t) = -k(z-l) + \frac{h_m}{2}\left[\sin\left(\omega t - \frac{\pi}{2}\right) + 1\right] + \Delta h, \qquad (15)$$

where h_m is the maximum opening height, $\max[h_o(t)]$, ω is the angular velocity of the oscillation, and Δh is a minimum opening height to avoid numerical singularity. The sinusoidal function has a phase shift to match the orifice throat motion, i.e., $f(z=l,t)=h_o(t)$. By utilizing an orifice half angle, $k=\tan(15°)$, $h_m=1.6$ mm, $f=100$ Hz and $l=10$ mm and by enforcing measured data for $C_d(t)$, $u_c(t)$, and $A_o(t)$, Eq. (12) results in

$$\Pi_{\text{rms}} = \sqrt{\frac{1}{T}\int_0^T \Pi(t)^2 dt} \sim 6.7 \text{ (J/s)} \qquad (16)$$

and

$$\Pi_{\text{net}} \sim 0.14 \text{ (J/s)}. \qquad (17)$$

The resulting power is shown in Fig. 13 together with the orifice throat opening, $h_o(t)$. The energy transferred, i.e., the area of $\Pi(t)$, during the opening phase is 3.25×10^{-2} J, whereas it is -3.11×10^{-2} J during the closing phase. The net rate of energy transfer is approximately 2% of the root mean squared value of the instantaneous power ($\Pi_{\text{net}}/\Pi_{\text{rms}} \sim 0.02$). These results follow a similar trend that was reported in a previous study (Thomson *et al.*, 2005) where the energy transfer for the self-oscillating synthetic model was computed numerically. The estimated power in the present study presents a generally odd-symmetric transition because the model retains its converging orifice shape throughout the period. The wave form of the power during the first and last quarter of the period corresponds to the temporal transition of the orifice coefficients observed in Fig. 6(a).

The hysteresis of the orifice discharge coefficients indicates the occurrence of net power transfer from the glottal flow to the physical model. The amount of power transfer in the PG model is compared in Table I with that of the NPG (converging) model, which has no hysteresis in the orifice discharge coefficient over one complete phonatory cycle (Park and Mongeau, 2007). The same shape function, i.e., Eq. (15), was used in the calculation of power transfer of the NPG model with experimental data, $C_d(t)$, $u_c(t)$, and $A_o(t)$ from the previous study of Park and Mongeau (2007). Note

TABLE I. Flow resistance and power transfer from the glottal flow to the vocal fold wall for the physical models with and without a posterior gap (PG and NPG). Power calculations are based on the shape function defined in Eq. 15, which is applicable only to the converging vocal fold profile.

	PG model (converging)	NPG model (converging/diverging)
Flow resistance (kPa s/l)	3.05	6.62/4.4
Π_{rms} (J/s)	6.7	4.7/NA
Π_{net} (J/s)	0.14	0.004/NA

that the shape function is not applicable to the NPG model for a diverging orifice wall shape. The net power transferred for the NPG model is negligibly small with respect to that of the PG model. The flow resistance, $\Delta P/Q$, is lower for the PG model because the glottal flow is sustained through the posterior gap even during the closed phase.

The net power transfer of 0.14 J/s is approximately 10% of the power required in order to drive Thomson's model oscillations (~1.3 J/s). The transferred amount of energy may not be significant enough to significantly affect the fold vibration or cause a local deformation of the folds. However, it is possible that the posterior gap may cause a change in the mechanical property of the fold tissue and that the minimum flow initiates an energy exchange that may have some benefit on the dynamics of the vocal folds as the parameters of the posterior gap vary further. Further studies with posterior gaps of varying size would be helpful to characterize this further.

C. Broadband sound and jet mixing scale

The energy source of jet noise is turbulence that grows downstream as the jet entrains the surrounding air. Broadband noise is produced by the coherent turbulent mixing within the jet, with large scale turbulence a more efficient radiator (Hubbard, 1994; Zhang *et al.*, 2004). The length scales of the jet mixing regions may be estimated from the frequency at which the broadband sound spectral level is dominant.

The range of length scale, L, corresponding to a given convection velocity, u_c, and Strouhal number, St, is obtained by

$$\frac{\text{St}u_c}{f_{\max}} < L < \frac{\text{St}u_c}{f_{\min}}, \qquad (18)$$

where $f_{\min}<f<f_{\max}$ is the frequency range of interest. Assuming a Strouhal number of 0.2 for vortex shedding for $400<\text{Re}<10\,000$ (Blake, 1986) and $u_c=40$ m/s from Fig. 4, together with frequency ranges of $125<f<4000$ Hz for the open PG jet and $400<f<16\,000$ Hz for the confined PG jet [Figs. 8(e) and 8(f)], the length scale associated with the elevated broadband sound level can be further normalized to

$$2 < \frac{L}{D} < 64 \text{ (open)}, \quad 0.5 < \frac{L}{D} < 20 \text{ (confined)}, \qquad (19)$$

where the glottal height, $D\sim1$ mm, is used for normalizing L. This indicates that the minimum flow enhances turbulent mixing over a scale on the order of the glottal height and

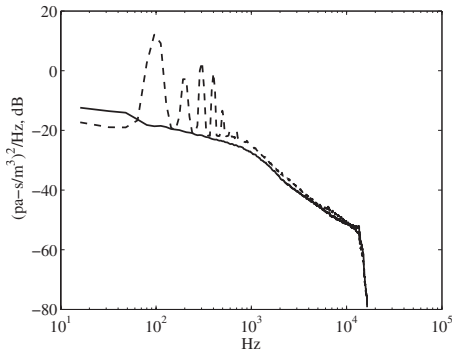


FIG. 14. Power spectral density of the radiated sound for steady (—) and unsteady (---) PG models, normalized the time-averaged flow rate, Q_m. Orifice area is $A_o(t)$, as defined in Fig. 5, for the unsteady PG model, and $A_o(t=1/2T)$ for the steady PG model. The transglottal pressure is $\nabla p = 10$ cm H_2O for both cases.

larger, while the (vocal tract) confinement is likely to confine mixing over length scales up to $\sim 20D$.

It is interesting to note that broadband sound production was observed to be independent of the orifice motion in the presence of the posterior gap. Figure 14 shows a comparison of power spectral densities of the sound produced by unsteady and steady PG models with the pressure normalized by the time-averaged flow rate. The broadband level is almost identical regardless of the presence of the glottal pulsations. This suggests that the broadband sound produced by the PG jet may overwhelm that from the pulsating main jet. This implies that accurate measurements of the broadband sound production may be made using a rigid replica of the PG model.

V. CONCLUSIONS

The effects of a posterior gap on the glottal orifice discharge coefficients and on broadband sound production in voicing were investigated. The quasisteady approximation was found to be reasonably accurate for $\mathrm{Re} > 4000$. The orifice discharge coefficient was observed to undergo a hysteretic transition. The net rate of energy transfer from the flow to the inferior glottal wall was approximated by using measured variables and an orifice shape function with further restrictive assumptions on the flow field. Approximately 2% of the root mean squared value of the instantaneous power is the net power transferred over one period due to the orifice discharge hysteresis. The presence of the minimum flow through the posterior gap enhanced turbulent mixing and increased the broadband radiated sound pressure level at high frequencies, within the frequency range of maximum human hearing sensitivity. This frequency band was found to correspond approximately to the shedding frequency of vortices with length scales that are comparable to the lateral and longitudinal dimensions of the glottal orifice and the vocal tract duct.

Belisle, G., and Morrison, M. (**1983**), "Anatomic correlation for muscle tension dysphonia," J. Otolaryngol. **12**, 319–321.

Beranek, L. L., and Vér, I. L. (**1992**), *Noise and Vibration Control Engineering: Principles and Applications* (Wiley, New York).

Biever, D., and Bless, D. (**1989**), "Vibratory characteristics of the vocal folds in young adults and geriatric women," J. Voice **3**, 120–131.

Blake, W. K. (**1986**), *Mechanics of Flow-Induced Sound and Vibration, Volume 1: General Concepts and Elementary Sources* (Academic, London).

Bless, D., Biever, D., and Shaik, A. (**1986**), "Comparisons of vibratory characteristics of young adult males and females," edited by J. Hibi, M. Hirano, and D. Bless, Proceedings of the International Conference on Voice, Kurume, Japan, pp. 46–54.

Cranen, B., and Schroeter, J. (**1995**), "Modeling a leaky glottis," J. Phonetics **23**, 165–177.

Erath, B. D., and Plesniak, M. W. (**2006**), "An investigation of bimodal jet trajectory in flow through scaled models of the human vocal tract," Exp. Fluids **40**(5), 683–696.

Fritzell, B., Hammarberg, B., Gauffin, J., Karlsson, I., and Sundberg, J. (**1986**), "Breathiness and insufficient vocal fold closure," J. Phonetics **14**, 549–553.

Hammarberg, B., Fritzell, B., and Schiratzki, H. (**1984**), "Teflon injection in 16 patients with paralytic dysphonia: Perceptual and acoustic evaluations," J. Speech Hear Disord. **49**, 72–82.

Holmberg, E., Hillman, R., and Perkell, J. (**1988**), "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.

Hubbard, H. H. (**1994**), *Aeroacoustics of Flight Vehicles: Theory and Practice Vol 1. Noise Sources* (Acoustical Society of America, New York).

Koike, Y., and Hirano, M. (**1973**), "Glottal-area time function and subglottal-pressure variation," J. Acoust. Soc. Am. **54**, 1618–1627.

Morrison, M., Rammage, L., Belisle, G., Pullan, B., and Nichol, H. (**1983**), "Muscular tension dysphonia," J. Otolaryngol. **12**, 302–306.

Park, J. B., and Mongeau, L. (**2007**), "Instantaneous orifice discharge coefficient of a physical, driven model of the human larynx," J. Acoust. Soc. Am. **121**, 442–455.

Peppard, R., Bless, D., and Milenkovic, P. (**1988**), "Comparison of young adult singers and nonsingers with vocal nodules," J. Voice **2**, 250–260.

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.

Zhang, Z., Mongeau, L., and Frankel, S. (**2002**), "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by jets in tubes," J. Acoust. Soc. Am. **112**, 1652–1663.

Zhang, Z., Mongeau, L., Frankel, S., Thomson, S., and Park, J. B. (**2004**), "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**, 1720–1728.

Zwicker, E., and Fastl, H. (**1999**), *Psycho-Acoustics: Facts and Models*, 2nd ed. (Springer, New York).

Patterns of acquisition of native voice onset time in English-learning children

Joanna H. Lowenstein[a] and Susan Nittrouer[b]
Department of Speech and Hearing, Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, Ohio 43210

(Received 2 November 2007; revised 16 May 2008; accepted 16 May 2008)

Learning to speak involves both mastering the requisite articulatory gestures of one's native language and learning to coordinate those gestures according to the rules of the language. Voice onset time (VOT) acquisition illustrates this point: The child must learn to produce the necessary upper vocal tract and laryngeal gestures and to coordinate them with very precise timing. This longitudinal study examined the acquisition of English VOT by audiotaping seven children at 2 month intervals from first words (around 15 months) to the appearance of three-word sentences (around 30 months) in spontaneous speech. Words with initial stops were excerpted, and (1) the numbers of words produced with intended voiced and voiceless initial stops were counted; (2) VOT was measured; and (3) within-child standard deviations of VOT were measured. Results showed that children (1) initially avoided saying words with voiceless initial stops, (2) initially did not delay the onset of the laryngeal adduction relative to the release of closure as long as adults do for voiceless stops, and (3) were more variable in VOT for voiceless than for voiced stops. Overall these results support a model of acquisition that focuses on the mastery of gestural coordination as opposed to the acquisition of segmental contrasts. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945118]

PACS number(s): 43.70.Ep, 43.70.Fq [CHS] Pages: 1180–1191

I. INTRODUCTION

How does children's speech acquire the phonetic structure of mature speakers in their linguistic environment? This question has intrigued child-language scientists for decades. Early accounts of language development were predicated on the notion that phonetic segments are integral to speech production; that is, they are the so-called "building blocks" of language. Accordingly, experimental approaches to language development were largely transcription based, focusing on when children learn to produce and combine discrete phonetic segments (e.g., Bean, 1932; Jakobson, 1941; Osgood, 1953; Prather, Hedrick, and Kern, 1975; Velten, 1943).

Some approaches to the study of speech development instead propose that the first unit of productive organization is something other than the phonetic segment, and these approaches use methods of analysis not focused as strongly on phonetic transcription. Generally, alternatives to the phoneme as the initial unit of speech organization focus on larger units, such as unanalyzed syllables or words. Some of the terminology invoked to convey this concept includes "word recipes" (Vihman and Velleman, 1989), "articulatory routines" (Menn, 1978, 1983), and "templates" (Vihman and Velleman, 2000). Speech development in this view involves gradually differentiating these routines into their component consonantal and vocalic gestures, and then learning to recombine these gestures as appropriate to produce sequences of speech with proper organization so as to allow the listener to retrieve phonetic structures (e.g., Goodell and Studdert-Kennedy, 1991). This process may be termed gestural reorganization.

A. Gestural reorganization

The notion that the organization of speech gestures is modified as children acquire experience producing and hearing their native language complements the concept of the phonetic segment as a perceptuomotor unit (e.g., Studdert-Kennedy, 1987). According to this account, phonetic segments do not exist separately from longer utterances. Rather, phonetic structure arises from the carefully coordinated set of gestures involved in the production of long stretches of speech. Children do not learn phonemes in this view, but rather discover phonetic structure by hearing how the gestures of their native language are produced and organized and simultaneously by learning how to accomplish that coordination themselves. In accordance with these models, some studies of speech development instead focus on analyzing patterns of articulatory movement in young children's speech through both acoustic analysis and direct measures of articulation (e.g., Goodell and Studdert-Kennedy, 1993; Green *et al.*, 2000).

In the speech of adults, individual gestures are precisely specified and generally produced with a great degree of consistency across utterances and exacting coordination among adjacent gestures such that phonetic implementation is in line with what is appropriate for the language being spoken and the situation in which an utterance is being produced [e.g., adaptive variability (Lindblom, 1990)]. It is this precision in spatial composition and interarticulatory timing rela-

[a]Electronic mail: lowenstein.6@osu.edu.
[b]Electronic mail: nittrouer.1@osu.edu.

tions that gives rise to segmental and suprasegmental structures in speech (e.g., Kelso *et al.*, 1984; Munhall, 1985; Nittrouer, 1991; Nittrouer *et al.*, 1988). However, unlike adults' speech, children's early speech gestures appear to be only roughly specified (i.e., spatial targets are general), to vary in form across similarly shaped utterances (which follows from them being roughly specified), and to lack precision in interarticulatory timing (Browman and Goldstein, 1989; Goffman and Smith, 1999; Goodell and Studdert-Kennedy, 1993; Green *et al.*, 2000; Macken and Barton, 1980a; Nittrouer, 1993; Nittrouer *et al.*, 1989; Nittrouer *et al.*, 1996; Smith and Goffman, 1998; Smith and McLean-Muse, 1986). Moreover, unlike the speech of adults, children's first babbling utterances show little language specificity. In fact, the patterns of early productions are strikingly similar across languages, at least when activity that can be correlated with specific segments is examined (e.g., Locke, 1983; Oller and Eilers, 1982).

However, when speech production is examined at a more "global" level, we find that children acquire the productive patterns of their native language at an early age. For example, de Boysson–Bardies *et al.*, (1986) derived the long-term spectra of adults with different native languages. When the long-term spectra of separate groups of 10-month-olds from those language environments were derived, it was found that those spectra differed across language groups in precisely the same way as the adults' spectra did. Thus, by 10 months of age children must know something about the gestural organization of their native language, although only at a global level, that is, one less clearly associated with individual phonetic segments and instead focused more on overall patterns of production.

A number of investigators have described the articulatory patterns of children's initial productions: specifically, babbling and/or first words (generally meaning when the child has less than roughly a dozen words). For example, MacNeilage and Davis (1991, 1993) reported that children's earliest utterances can be adequately described as cycles of mandibular elevation (resulting in poorly specified constrictions) and depression, accompanied by phonation; that is, children's early speech consists largely of complete closures of the lips or tongue against the palate (resulting in /b, d, m, n/) alternating with different degrees of mandibular opening (resulting in central vowels of varying heights). Stretches of long voicing lag are extremely rare in babbled productions (e.g., Locke, 1980; Whalen *et al.*, 2007), probably because young speakers are unable to incorporate vocal-fold abduction gestures into the already complicated (for young children) task of moving upper vocal-tract articulators while phonating.

Other studies suggest that at least some of young children's gestures actually differ from those of adults in form. For example, a motion tracking study of bilabial production found significant differences in articulator coordination between 1-year-olds, 2-year-olds, 6-year-olds, and adults (Green *et al.*, 2000). The 1- and 2-year-olds achieved oral closure primarily through jaw displacement, while the 6-year-olds and adults achieved oral closure through a combination of lower lip movement and jaw displacement. Furthermore, the contribution of the jaw to oral closure was found to differ significantly for 1-year-olds versus 2-year-olds, supporting the suggestion of reorganization at these early ages. In agreement with the findings of Green *et al.* (2000), Smith and McLean-Muse (1986) found that children as old as 11 years made labial gestures with different lower lip and jaw displacements and peak velocities than adults. Finally, coordination among gestures is not as precisely handled in children's speech as in the speech of adults: Goodell and Studdert-Kennedy (1993) recorded children (at 22 and 32 months) and adults imitating minimal pair nonsense disyllables that consisted of the syllable [bə] followed by a stressed [ba], [bi], [da], [di], [ga], or [gi]. They found that the proportions of total utterance duration assigned to each syllable became more adultlike at 32 months, and there was an age-related decrease in gestural overlap across the two syllables in terms of tongue front-back position. They also found evidence of more adultlike differentiation in tongue height for the schwa and stressed-syllable vowels at 32 months, compared with 22 months. Smith and Goffman (1998) found that children (aged 4 and 7 years) produced speech with less stable movement trajectories than adults (also see Goffman and Smith, 1999).

The process of gestural reorganization appears to continue into the early school years. For example, children as old as 7 years have been found to produce adjacent consonant and vowel gestures with more overlap than adults for both fricatives (Nittrouer *et al.*, 1989) and stops (Nittrouer, 1993); in addition, they have difficulty consistently coordinating vocal-tract closure with laryngeal abduction in the production of words with syllable-final voiced or voiceless stops (Nittrouer *et al.*, 2005). Several studies have found that children generally produce gestures at slower rates than adults, at least up to eight years of age (Goffman and Smith, 1999; Goodell and Studdert-Kennedy, 1993; Green *et al.*, 2000; Nittrouer, 1993; Smith and Goffman, 1998; Smith and McLean-Muse, 1986). Children have also been found to produce individual articulatory gestures with greater temporal variability than adults (e.g., Nittrouer, 1993). In sum, we find a much more convoluted and protracted developmental course for gestural speech reorganization than some phonetic transcriptions indicate.

B. Acquisition of syllable-initial stop voicing

The acquisition of voicing categories for syllable-initial stops can help inform us about developmental changes in gestural coordination. When stops occur at the start of words the timing between the release of the oral closure and the onset of phonation must be coordinated in order to provide the listener with appropriate information for voicing judgments. This timing, termed voice onset time (VOT), was first demonstrated to differ across languages by Lisker and Abramson (1964). Cross linguistically they found three patterns for VOT: long voicing lead, where phonation starts well before the oral release; short voicing lag, where phonation begins just after the oral release; and long voicing lag, where phonation begins well after the oral release. Because lan-

guages differ in their phonetic requirements, children need to learn the appropriate gestural timing specific to their native language.

Generally speaking, most words spoken are not produced in isolation; instead they occur in a stream of speech. In order to produce a voiceless aspirated stop, phonation is interrupted in continuous speech by a glottal abduction gesture. This gesture involves the posterior cricoarytenoid muscle, which begins to contract shortly before the glottis opens and reaches maximum activation before the peak of glottal opening. As the activity of the posterior cricoarytenoid muscle decreases, the interarytenoid muscles contract, resulting in the return of the glottis to a position appropriate for phonation (Löfqvist *et al.*, 1984; Löfqvist and Yoshioka, 1980). Because voiceless aspirated sequences are so rare in babbled speech (e.g., Locke, 1980; Whalen *et al.*, 2007), it is likely that children are unable to incorporate these abduction/adduction gestures into their productions. However, when they begin producing words, they must learn to do so.

Studies of stop acquisition using methods other than phonetic transcription have found that children only gradually develop the ability to produce long-lag VOTs. For example, Macken and Barton (1980a) audiotaped four English-learning children at 2 wk intervals starting around 18 months and ending around 26 months. Each taped session was 20–30 min in length, and toys were used in the sessions that would explicitly elicit stop productions. The examiners isolated words that began with stops, with the criterion that there had to be an interruption in voicing between the preceding word and the isolated target word. In other words, children in the study had at least mastered the ability to abduct or adduct vocal folds during continuous speech. Spectrographic and oscillographic analyses were done on the speech samples to measure VOT. Based on the results, Macken and Barton (1980a) proposed a three-stage model for the acquisition of stop consonants. These three stages are age independent and describe a linear path for the development of stop production. The first proposed stage of stop acquisition consisted of children producing primarily short-lag stops (defined as VOTs between 0 and 20 ms). In this stage, children's articulation seems to fit the description of "everything moves at once" offered by Kent (1983) because they appear to simultaneously, or nearly simultaneously, begin phonation and release the oral closure. Longitudinal studies of languages other than English confirm Macken and Barton's (1980a) findings that children initially produce primarily short-lag stops (Eilers *et al.*, 1984; Kehoe *et al.*, 2004; Macken and Barton, 1980b).

In Macken and Barton's (1980a) second stage, children begin to differentiate between word-initial voiced and voiceless stops. Words that have voiced initial stops in the adult version still have stops with VOTs close to 0 ms. However, in words that should have (according to the adult version) a voiceless aspirated initial stop, the stops are consistently produced with longer VOTs. Macken and Barton (1980a) suggested that this stage can be further divided into two substages. In the first, VOTs for stops intended as voiceless are consistently longer than those of stops intended as voiced but are clearly in the perceptual category of "voiced" for English listeners (i.e., less than 20 ms). In the second substage, these stops have VOTs on the order of 40 ms, and so are ambiguous in perceived voicing for adult English speakers. This second stage, where children are producing a contrast that is not clearly perceptible to adults [e.g., "covert contrast" (Scobbie *et al.*, 2000)], can be seen as evidence for gestural reorganization—children have begun to expand their options for coordinating voicing onset and oral release, but have not yet perfected the process.

Macken and Barton's (1980a) proposed third stage of VOT development is also described as having two substages. In the first, children produce voiceless stops with VOT means considerably longer than adult means (over 100 ms). In the second part of the third stage, children produce VOTs more similar to those produced by adults. In summary, Macken and Barton's (1980a) model of VOT acquisition is one suggesting that children initially have great difficulty producing long-lag VOTs; after some early modest attempts at delaying the onset of laryngeal vibration relative to closure release, children exaggerate this pattern, obtaining adultlike productions at just over 2 years of age.

Of course, the question arises as to why children do not produce the long-lag VOTs of their native language in their earliest stop productions: Do they fail to discriminate short- and long-lag VOTs in perception, or are they unable to coordinate laryngeal and supralaryngeal gestures appropriately? Numerous studies of infant speech perception have shown that infants as young as 2 months of age can discriminate syllables beginning with short-lag versus long-lag VOTs (e.g., Eimas *et al.*, 1971; Werker and Tees, 1999), and so we conclude that children's perceptual capacities are adequate to hear the differences in lag duration for English voiced and voiceless initial stops. Instead, the initial inabilities of infants to produce long-lag VOTs very likely represent production constraints.

Although elegant in design, Macken and Barton's (1980a) proposed model has not gone unchallenged. In particular, the suggestion that young children go through a period of producing voiceless VOTs longer than those of adults has mixed support in subsequent studies. Several studies have found that children up to six years of age produce shorter mean VOTs than adults for words intended to have initial voiceless stops (e.g., Kewley-Port and Preston, 1974; Nittrouer, 1993; Zlatin and Koenigsknecht, 1976). Nittrouer (1993) found that VOT for initial /t/ produced by 3-year-olds is more than 10 ms shorter than those produced by older children and adults. There is also evidence that children attain the voiceless VOTs characteristic of their native language at varying ages: the 17 normal children in three related studies first produced adultlike VOT values at ages ranging from 18 to 29 months (Macken and Barton, 1980a; Snow, 1997; Tyler and Saxman, 1991).

C. Variability in VOT development

Regardless of claims about the patterns of VOT acquisition, a common finding of all studies is that children (at least up to 7 years of age) produce VOTs for voiceless stops with

greater intraspeaker variability than adults (Barton and Macken, 1980; Gilbert, 1977; Kewley-Port and Preston, 1974; Koenig, 2000; Macken and Barton, 1980a; Menyuk and Klatt, 1975; Nittrouer, 1993; Zlatin and Koenigsknecht, 1976). Variability is generally viewed by child phonologists as the reciprocal of skill in coordinating vocal-tract gestures; that is, variability diminishes as skill in coordinating gestures increases (e.g., Goodell and Studdert-Kennedy, 1993; Hodge, 1990; Kent, 1976). From this perspective, variability is viewed similarly to how investigators interested in the development of other skilled actions view it (e.g., Bruner, 1973), and diary studies are replete with reports of young children exhibiting great variability in their attempts at word productions. For example, Ferguson and Farwell (1975) listed ten transcribed attempts of a toddler attempting to say "pen," all of them different. It is clear from the list that the child had all the component gestures of the adult form, but was unable to coordinate timing among those gestures. In the case of VOT production, high variability relative to that which is encountered in more mature speakers likely signifies that the child is in the process of mastering the coordination between oral release and the onset of vocal-fold vibration that produces long-lag VOTs. We would not expect variability in VOT to be as great in children's attempts at voiced initial stops precisely because of the everything-moves-at-once principle: if one can begin all gestures at the same time, or nearly at the same time, coordination is simplified. For these reasons, the present investigation examined VOT variability, with the prediction that it would be greater for voiceless stops.

D. Avoiding difficult gestural sequences

One factor that complicates the study of VOT acquisition is that babbled productions and real words coexist in the speech of toddlers. Locke (1980) reviewed three studies of babbling in 11–12 month old American infants and found that complete closures of the vocal tract accompanied by even minimal vocal-fold abduction (i.e., voiceless aspirated stop-like sounds) were present in only 1–11% of babbled utterances; similarly, Oller and Eilers (1982) found that English- and Spanish-learning infants produced aspirated initial plosives only 6–11% of the time. This means that investigators are largely obliged to wait until children begin attempting words intended to have initial voiceless stops to examine the acquisition of these gestural sequences. However, what if, as some have suggested (Schwartz and Leonard, 1982; Schwartz *et al.*, 1987), children avoid words with gestural sequences that are difficult for them? This could complicate analyses, particularly those that rely on counts of voiceless and voiced syllable-initial stops. For example, Kewley-Port and Preston (1974), who analyzed all vocalizations that began with non-nasal, nonlateral, pulmonic egressive apical stop consonants (both babble and words), found for two of the children in their study that there was an increase in the number of apical stops with long-lag VOTs (defined as over 25 ms) from less than 15% of all apical stops to more than 50% between the ages of 1 and 2 years. It is possible that this change resulted from an increase in the number of words (rather than babbled sequences) with intended voiceless stops, although it is not clear from their report because they did not distinguish between the words and babble. In a study of spontaneous speech, Dobrich and Scarborough (1992) found that children between the ages of 2 and 3 years produced fewer target words, which in their adult form contained final consonant clusters, than their mothers did, showing some evidence of avoidance for children in that age range.

However, the avoidance hypothesis begs the question of how children acquire production skills: If they are truly avoiding words with certain gestural sequences, how do they ever get the practice needed to master those sequences? Schwartz *et al.* (1987) answered that question by suggesting that children analyze the segmental contents of words in order to determine whether they are currently able to produce them and gradually relax their selection constraints in what they will attempt as they gain more general speaking experience. We might modify the account to suggest that children analyze the *gestural* demands of words and avoid those for which those demands are too great. Then, as they gain more experience generally producing speech, they become more willing to try new gestural sequences. In any event, counting the numbers of words produced by children that in the adult form begin with long-lag VOTs should reveal whether children initially avoid words that begin with voiceless aspirated stops, and then increase their attempted productions of these words as they become more experienced.

E. The current study

The present investigation explored developmental trends in gestural coordination by focusing on the acquisition of VOT because long-lag voiceless stops require careful timing between the oral release and the onset of phonation, are rarely found in babble and early words, and only slowly approach adult VOT values. This protracted developmental course of voiceless initial stop acquisition makes this gestural pattern an ideal candidate for study. The investigation posed three specific questions: First, do children's productions of VOT for syllable-initial stops intended to be voiceless become more like those present in their native language by means of a gradual acquisition process? The answer to this question would inform us regarding how children's speech productions acquire phonetic structure. Finding that children only gradually master the coordination needed for initial stops to be heard as voiceless aspirated stops would support the position that phonetic structure emerges as children master component speech gestures and language-specific coordination among them. The specific hypothesis to be tested stemmed from Macken and Barton's (1980a) three-stage model, but with modifications: children less than three years of age were expected to (1) initially produce only short-lag VOTs for stops intended as both voiced and voiceless, then (2) produce a difference in VOTs for voiced and voiceless stops that do not correspond to adult values, and (3) finally produce VOTs that approach but are still shorter than adult values. The prediction for the third stage of this model differs from that of Macken and Barton (1980a) but is commensurate with the results of others (Kewley-Port and

Preston, 1974; Nittrouer, 1993; Zlatin and Koenigsknecht, 1976), showing that children continue to have shorter VOTs than adults for voiceless initial stops into the early school years.

The second question asked in the present study was whether children become more skilled at achieving voiceless VOT targets during the ages studied here. Measuring variability for each speaker individually provides an estimate of how consistent speakers are in their productions. We predicted that variability would be greater in children's productions of words with voiceless initial stops, rather than with voiced initial stops, but we did not know if that variability would diminish over the course of the study.

The third question asked was whether children increase the number of stops intended as voiceless as they gain experience in producing their native language. The specific hypothesis to be tested was that there would be a steady increase in the number of stops intended as voiceless over time, demonstrating that children are willing to produce more stops intended to be voiceless as they become more experienced. In other words, evidence of this finding would support the notion of avoidance.

The speech samples that were analyzed for this study were obtained from children in unstructured play situations with one of their parents. The speech samples used in these analyses were collected with the purpose of tracking the emergence of gestural organization for many kinds of phonetic sequences in young children as they acquire their first words. Alternative methods of collecting speech samples include imitation and targeting specific utterances, for example, through the use of toy animals that are given names with desired phonetic sequences. While each of these methods is appropriate for certain research goals, the purpose in our method was to obtain a naturalistic sample of what actually happens during that early word acquisition period.

II. METHOD

A. Participants

Participants were all part of a larger study (McGowan *et al.*, 2004) in which seven children (three males and four females) were each followed for almost a year and a half. Nine children were originally recruited, but two left the study before all recording sessions could be completed. The children recruited were typically developing children. All had normal prenatal histories, normal deliveries, and no special medical conditions. None of the children had an immediate family member with speech, language, or hearing disorders. None of the children had any reported history of otitis media with effusion at the start of the study, and no child was treated for more than one episode while the study was being conducted.

Children were taped at 2 month intervals between roughly the ages of 14 and 31 months. Recording sessions began when the parents estimated that the child had ten recognizable words, and were discontinued when the child began stringing three or more words together to form sentences. Children were not all recorded at precisely the same ages because they reached the landmark point of having ten identifiable words at slightly different times. By beginning recordings at the same point in language development, we hoped to obtain samples across children at consistent "language ages," so to speak. For these analyses, children's data were grouped into 2 month intervals across the study (15–16, 17–18, 21–22, 23–24, and 27–28 months), with a few exceptions. Not all children could be recorded at the ages of 19–20 or 25–26 months because of illness or family vacations. Not all children were recorded at 30–31 months because some had passed the three-word sentence criterion for dismissal from the project. Consequently, these three intervals were not included in the current study. Nonetheless, the sessions for which data were analyzed can provide a good overview of developmental trends. Within these intervals, children were judged to be at approximately the same developmental age, as judged by vocabulary size and general utterance length.

B. Equipment and materials

Recordings were obtained using an AKG C-535EB microphone, a Shure model M268 mixer, and a Nakamichi MR-2 cassette deck with metal tapes. This system had a flat frequency response out to 20 kHz. These recordings were subsequently low-pass filtered with a high-frequency cutoff of 10 kHz and digitized with a Soundblaster A/D card using the SPEECH STATION II software at a 22.05 kHz sampling rate and 16 bit digitization.

The same set of toys was available for play for all children at all sessions and consisted of some small stuffed dolls, foam puzzles, and cloth books. All toys used in these sessions were soft to minimize extraneous noises that might interfere with speech recording. The toys were not chosen to elicit specific responses from the children but were rather meant to stimulate communication in general. However, there were toys present that were balanced in terms of the voicing category of the initial stop in their referent lexical items, such as a foam puzzle of a pig (/p/) and a doll of Bert from Sesame Street (/b/).

C. Procedures

Children were recorded in the same sound-proof booth at each session. Our goal was to have 20 min sessions, but sometimes a turn for the worse in the child's behavior forced us to curtail a recording session. However, all sessions were at least 11 min in length. The child sat on a high chair at a table, with one parent across the table. The microphone was suspended roughly 23 cm above the child's mouth. It was suspended rather than table mounted because pilot work showed that children habituated to its presence more rapidly that way. Parents were instructed to play with their children, trying to elicit as much language as possible. Between sessions parents were asked to keep diaries of new vocabulary items (at the younger ages) and new sentence structures (at the older ages) that they heard at home.

D. Measurements

For this study, 10 min worth of speech was generally analyzed. 10 min sections were obtained by finding the tem-

poral middle of each recording session and using 5 min on either side of that point. If the original session length was only 11 min, the entire session was analyzed. For each child, the recordings were analyzed in "backward" order, starting with the one obtained at the oldest age, then the one at the next-to-oldest age, and so on until all recordings were analyzed. This allowed the examiner to acclimate to each child's speech, starting with what should have been the most intelligible sample.

Words heard as starting with an initial stop were extracted and yielded a total sample of 1458 words. Words that closely matched the adult form and those that did not were included. Words that were not in phrase-initial position were included only when there was a clear break in voicing after the end of the previous word. Words were excluded if there was voicing carried through the closure, as it is never clear what portion of that voicing belongs to the closure of the previous word and what portion belongs to the target word (e.g., Allen, 1985). Furthermore, voicing during oral closures of considerable duration in children's speech is often a consequence of nasalization. As such, that voicing could not be counted as "prevoicing." The first author made phonetic transcriptions of each word, and a phonological transcription according to what word was intended. Whenever possible, these transcriptions were based on a conversational context (e.g., parental repetition of the word, what book the child was looking at, or what toy the child was playing with). For example, [bɪg] might be transcribed as "big," or "pig," depending on the context: If the parent and child were discussing something that was big, it was transcribed as "big;" if they were discussing farm animals, it was transcribed as "pig." The context was ambiguous for 24% of words (350), across all sessions. For these instances when conversational context was ambiguous, transcriptions were based on what the transcribers heard and thought the child intended. When context did not clearly signal word identity, two research assistants also transcribed the word. Generally there was little disagreement about the intended voicing of the stop among the three transcribers; slightly more were transcribed as voiceless than as voiced. (Place was more problematic, but VOT according to place was not considered here.) Specifically, the first author and research assistants identified 90% of the words (315) with ambiguous context as starting with the same voicing target. In the 10% of words for which there was disagreement (35), words were transcribed according to the decisions of the first author, who had the most experience listening to the individual children and so knew what their interests and speech patterns generally were. In particular, the first author had the benefit of being familiar with what were actually *later* produced words from the child because of the backward analysis approach. This sort of ambiguity in word identity is unavoidable when speech samples are collected from unstructured play situations, but again, the availability of samples such as these is important to our collective understanding of phonetic development.

Three measurements were made on words from each session for each child. First, the numbers of words with voiced and voiceless initial stops in the 10 min section were counted. Stops were counted as voiced or voiceless depending on what they would be in the adult version of the word. Second, VOT was measured for each stop from a combined wave form and spectrogram using WAVESURFER (Sjölander and Beskow, 2000). Markers were placed at the start of the broadband aperiodic burst in the wave form (correlate of the oral release) and at the onset of a regular periodic signal in the wave form (correlate of voicing onset), as illustrated in Fig. 1. Marker placement was confirmed in the spectrographic display. VOT was computed as the interval between these two markers. VOTs with more than 50 ms of prevoicing were excluded because of the rarity of such long-lead voicing in initial position in English. Moreover, for these particular samples, it was usually clear that these long-lead segments were not true oral vocalizations, but rather nasal productions. This observation is what would be expected for children, given their small oral cavities (which would make it hard to continue airflow through the glottis for very long before sub- and supraglottal pressures equalized, causing vocalization to cease). Sequences with more than a 300 ms lag were also excluded. Gaps that are long between an oral release and a vocalic segment exceeded the average syllable length for most of these children, and so it was reasonable to judge that the two elements (release and vocalic production) could not be viewed as part of a single integrated syllable. For each child, mean VOT was calculated for voiceless and voiced target initial stops separately.

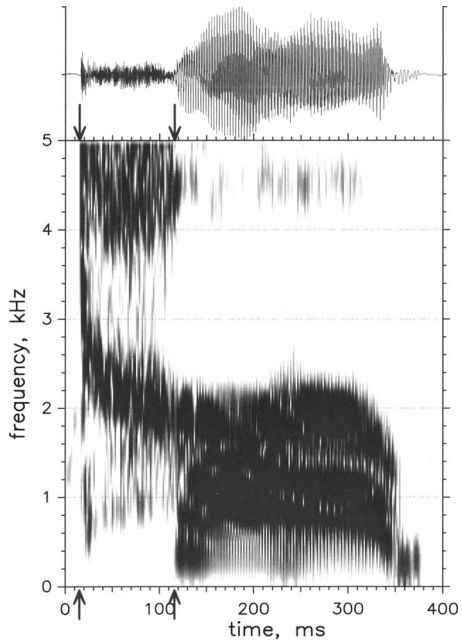


FIG. 1. Wave form and spectrogram of a 24-month-old male's production of /tub/. Arrows indicate VOT measurement boundaries.

As the avoidance hypothesis would suggest, some children within some recording sessions did not produce many words with particular initial stops, making it hard to obtain a sample large enough to provide an accurate estimate of VOT for that particular stop. Therefore, if fewer than five tokens of a specific stop (/b/, /d/, /g/, /p/, /t/, or /k/) were present in the 10 min segment identified for use, VOT for that stop was analyzed from the entire recording session. This was done in order to obtain more reliable estimates of VOT across place.

TABLE I. Number of children (out of seven) who produced fewer than five stops in a category during the middle 10 min of recording, by session.

Age (month)	Stop					
	/b/	/d/	/g/	/p/	/t/	/k/
15–16	4	4	6	7	7	7
17–18	3	2	6	5	7	7
21–22	0	3	4	7	5	5
23–24	0	2	6	5	4	3
27–28	0	1	4	3	1	1

Table I lists the numbers of children (out of seven) who produced fewer than five stops in each category during the 10 min segment at each age. These additional tokens resulted in total sample sizes of 140 (15–16 months), 209 (17–18 months), 325 (21–22 months), 364 (23–24 months), and 420 (27–28 months) being used for the computation of VOTs.

The third measurement made was variability for each child's mean VOT for voiced and voiceless target initial stops separately. Standard deviations (SDs) were used to index variability. Coefficients of variation, computed by dividing SD by the mean as a way to normalize SD, are sometimes used to index variability in temporal speech measures (e.g., Smith, *et al.* 1983). However, that method is generally reserved only for situations in which measures have an absolute limit of 0 ms (such as syllable length). Because initial stops can be prevoiced, that situation does not exist in this case. Within-child standard deviations (WCSDs) were calculated for each child's voiced and voiceless initial stops.

Statistical analyses were performed using the BMDP statistical software (1990). The magnitude of the voicing effect on measures was computed using both η^2 and Cohen's d. η^2 varies between 0 and 1 and measures the proportion of variance accounted for by a single factor. Cohen's d measures the relative magnitude of difference between two means in terms of standard deviation and is therefore a normalized index of effect size (Cohen, 1988). Generally speaking, $d > 0.8$ represents a large effect size.

III. RESULTS

A. Description of children's productions

At 15–16 months, all seven children were producing primarily one- or two-syllable utterances, with one or two identifiable words produced in a session. At 17–18 months, two of the children were mostly producing identifiable words, while the other five were producing combinations of babbled sequences and real words. At 21–22 months, three children were producing only words, and four were producing mostly words. At 23–24 months, four children were producing only words, and three were producing mostly words, generally in a mix of one-word and two-or-more-word utterances. At 27–28 months, all of the children were producing only identifiable words, generally in two-or-more-word utterances. According to parent reports, all of the children in this study were producing ten words at their first recording session (15–16 months). The children reached the point where they were producing 50 words at sometime between 18 and 26 months (mean 22.6 months, median 24 months).

B. VOT

Figure 2 presents VOT averaged across all children and voicing categories (i.e., all tokens from all children were averaged for each time period), with error bars representing standard deviation across all tokens from all children. The number of tokens included in the calculations at each age is indicated at the bottom of the graph. This analysis provides a picture of the overall changes in VOT regardless of voicing category assignment. Mean VOT was 12 ms at the youngest age examined and increased to 40 ms at 23–24 months, where it remained. Standard deviations for VOT also increased over time. Looking at the standard deviation, it is clear that VOTs at 15–16 months fell into a relatively narrow range (−10 to+30 ms) that roughly corresponds to the voiced category in English. This range also includes prevoiced stops, which are relatively uncommon in the initial position in American English. At 23–24 and 27–28 months, children's VOTs span the range of values for adult speakers of American English. A one-way analysis of variance (ANOVA), with age as the categorical variable, was performed on VOT, revealing a significant difference in VOT across age groups: $F(4, 1453)=31.475$, $p<0.001$, $\eta^2=0.80$. Thus, age accounted for a large proportion of variance in VOT (80%).

Figure 3 presents mean VOT for stops judged to be intended as voiced and voiceless, separately. To investigate developmental changes in VOT for voiced and voiceless stops separately, a simple effect analysis was done, investigating age-related changes in VOT for each voicing category

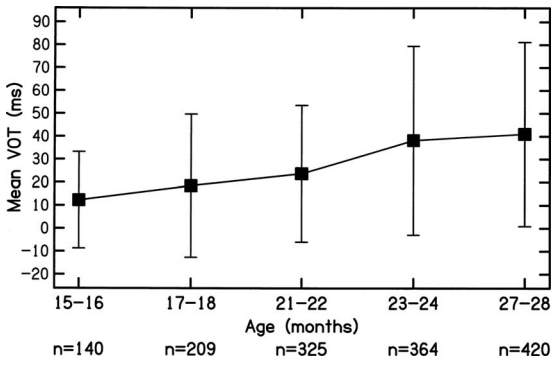


FIG. 2. Mean VOT values for children's stop productions, by age. Error bars represent standard deviations. Total sample sizes for each age are indicated.

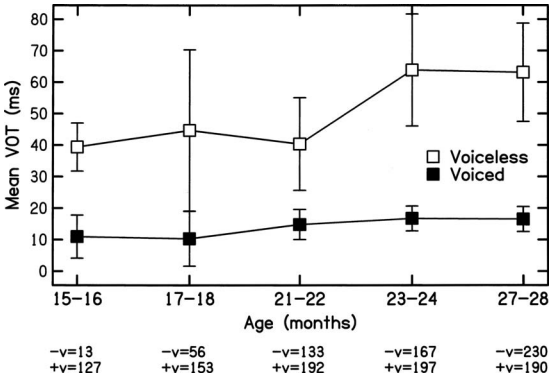


FIG. 3. Mean VOT values for children's voiced and voiceless stop productions, by age. Error bars represent standard deviations. Total sample sizes for voiced and voiceless stops for each age are indicated.

(i.e., voicing was held constant). A simple effect analysis is an appropriate selection of statistical tests for experiments with several independent factors because it permits the examination of effects for one or more of those factors at each level of another factor while using the overall estimate of error variance. This procedure provides a more sensitive test than would be obtained by doing separate ANOVAs at each level of the factor. Significant results were obtained for both the voiced $[F(4,24)=2.88,\ p=0.045]$ and voiceless $[F(4,\ 24)=4.26,\ p<0.01]$ stops. Children's mean voiced VOTs started at 11 ms, and stabilized at 16 ms at 21–22 months. Mean voiceless VOTs were constant at roughly 40 ms through 21–22 months. Then, at 23–24 months, the mean VOT jumped to above 60 ms. The relative magnitudes of the age differences on VOT can be indexed by computing the difference in mean VOT at each of these times and dividing by the pooled standard deviation (Cohen's d). Comparing children's initial voiced VOTs to their stable voiced VOTs (at 21–22 months) gives a d of 0.65 (a medium effect size), while comparing children's initial voiceless VOTs to their stable voiceless VOTs (at 23–24 months) gives a d of 1.79 (a large effect size). The error bars on Fig. 3 indicate standard deviation across children. It is clear that voiceless stops were produced with more variability in VOT (greater standard deviations) than voiced stops. It is also interesting that standard deviations for voiced stops decreased over the course of the study, while standard deviations for voiceless stops increased after the first recording session and remained high. It appears that these children were producing voiced stops more accurately with experience (resulting in lower variability) but were not yet able to perfect the coordination of gestures for producing voiceless aspirated stops (resulting in higher variability). These standard deviations were larger than those found for both adults and older children in previous studies (Koenig, 2000; Nittrouer, 1993).

C. Within-child standard deviations (WCSDs)

Very few syllable-initial stops meeting the criterion of "voiceless" were found in the samples of 15–16 month olds. With so few tokens from each child, it would not be appropriate to compute WCSDs. Consequently, statistics for

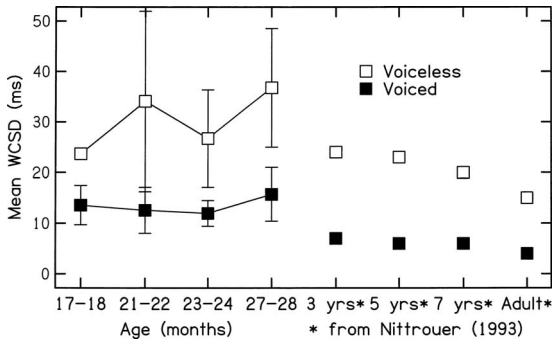


FIG. 4. Mean VOT WCSD, by age, including data from Nittrouer (1993). Error bars represent standard deviations.

WCSD were calculated beginning with the 17–18 month session. The left half of Fig. 4 displays mean WCSDs for each intended voicing category, across recording sessions, with error bars representing WCSD standard deviations. It appears that variability in VOTs for children's voiced initial stops was low and relatively consistent, with WCSDs of 12–16 ms. Variability VOTs for their voiceless initial stops, on the other hand, was greater, with mean WCSDs of 24–37 ms. A factorial analysis was performed on WCSD, examining the effects of age and voicing. Only the main effect of voicing was found to be significant, $F(1,\ 6)=44.23$, $p<0.001$, $\eta^2=0.82$, supporting the conclusion that WCSDs were greater for the voiceless than for the voiced stops. WCSD did not change significantly over time, for either the voiced or voiceless stops. As a comparison, note that Koenig (2000) reported mean within-subject standard deviations for voiceless stops of 11 ms for adults and 21 ms for 5-year-old children, and both of these standard deviations are shorter than those found for voiceless stops for our subjects at any of the ages we studied. We conclude that at least over the age range examined in this study, children did not improve in their abilities to coordinate the vocal-tract release and the onset of voicing.

D. Counts of words with initial voiceless and voiced stops

Figure 5 displays mean numbers of voiced and voiceless target stops, as counted from each 10 min segment; that is, stops outside of this time window that were examined to help provide an estimate of VOTs were not included in this analy-

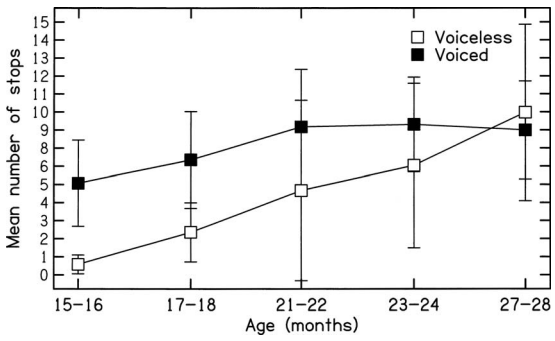


FIG. 5. Mean number of voiced and voiceless stops, by age. Error bars represent standard deviations.

sis. In examining Fig. 5, it is clear that the numbers of target voiced stops produced remained relatively constant over time, while the growth in target voiceless stops was nearly linear. Children started out attempting less than one voiceless stop on average during the 10 min analysis segment at 15–16 months and were producing similar numbers of voiced and voiceless stops by the end of this series of recordings.

To examine the developmental changes in stop count for voiced and voiceless stops separately, a simple effects analysis was done, investigating age-related changes in count for each target voicing category. There was no significant age effect for the numbers of voiced stops, but there was for the numbers of voiceless stops, $F(4,24)=6.55$, $p=0.001$. This effect reflects the steady growth in the numbers of words with intended initial voiceless stops produced by children across recording sessions. The relative magnitude of the change in the count of voiceless stops between 15–16 and 27–28 months was calculated using Cohen's d, resulting in a d of 2.42. This is clearly a very large effect.

E. Individual developmental patterns

Group averages inform us about general developmental trends across children, but cannot tell us if all children follow a similar developmental pattern. The most important information that longitudinal studies such as this one provides concerns patterns for individual children. It may be that the group trends reported here do not hold for all children. Perhaps one or a few of the children began exhibiting adultlike VOT patterns as soon as they began producing their first words. In fact, when individual patterns of development were examined, it was found that no child in this study exhibited an adultlike behavior in VOT production from the start of first words. In particular, all of the children produced very few target initial stops that could be considered to be intended as voiceless during the earliest two test sessions (note, however, that six out of seven children produced at least one). In other words, there was no evidence that the apparent avoidance of word-initial voiceless stops was characteristic only of a subset of the children. This finding was interesting because many of the items that children this age are typically interested in and, indeed, that these particular children talked about at later ages, begin with voiceless stops, for example, cats and cars. Again, the procedures of data collection for this study did not specifically encourage or dissuade children from using particular words, but the opportunities were present for producing words with both voiced and voiceless stops.

Examining VOTs of individual children showed two distinct patterns of acquisition. Four of the seven children in this study followed the overall pattern described earlier: All VOT measures were within the adult category of voiced during the earliest session, and then VOTs for words with intended initial voiceless stops increased to over 60 ms during the course of the study. The other three children showed a slightly different pattern of acquisition. For these children, VOT measures for words with intended voiceless stops were initially on the low end of the adult range; that is, VOTs were within 40–55 ms for these children. Over the course of the study, two of these children showed increases in VOTs for words intended to have initial voiceless stops. One of these children, however, never showed any change in VOTs for words intended to have initial voiceless stops, even though she did increase the number of words in this category that she attempted. Finally, all children in the study showed the same pattern regarding WCSD: when VOT for words intended to have voiceless initial stops was on the low end of the adult range (i.e., within 40–60 ms) or adultlike (i.e., over 60 ms), variability was high.

In summary, whether or not these children started out with clearly short-lag or ambiguous VOTs for voiceless stops, they all went through some sort of a developmental process. For most of the children, this process involved changes in the number of initial stops intended as voiceless that they attempted and increases in VOT itself. For one child, however, this developmental process demonstrated itself as only a change in the number of stops intended as voiceless. This child never achieved adultlike VOTs for words intended to start with voiceless stops. All children showed a lack of skill in coordinating supralaryngeal and laryngeal gestures for the production of voiceless VOT, as indicated by the finding of great variability through the entire length of this study.

IV. DISCUSSION

The goal of this study was to examine developmental trends in gestural coordination for VOT. We posed three specific questions: The first question was whether children's VOT for syllable-initial stops intended to be voiceless become more like those present in their native language later than VOT for syllable-initial stops intended to be voiced. It is clear that the children in this study followed a gradual acquisition process in their productions of voiceless VOTs. Looking at overall patterns (Fig. 3), mean voiceless VOTs remained in the short-lag or ambiguous range (i.e., below 55 ms) between 15–16 and 21–22 months. At 23–24 months, mean VOT increased so that it was in a range where stops would consistently be categorized as voiceless by native English speakers (i.e., above 60 ms). Looking at individual results, four of the seven children in this study performed according to Macken and Barton's three-stage model (1980a): voiceless stops were produced with short-lag VOTs until 21–22 months; then voiceless stops were produced with ambiguous VOTs; and at 23–24 months, voiceless stops were produced with VOTs closer to mature values. Two of the other three children achieved more mature VOTs at an earlier age, but still started out at 15–16 months producing VOTs in the ambiguous range. The seventh child did not acquire mature voiceless VOTs during the course of this study, even though values were within the ambiguous range from the onset of first words. According to these collective findings, we conclude that children only gradually modify their target voiceless VOTs to be similar to those of adults in their language community.

The second question asked was whether children would become more skilled at achieving VOT targets over the course of this study, as would be indicated by significant reductions in within-speaker VOT variability (i.e., WCSD). Results showed that WCSDs stayed fairly constant during the time course of this study, with voiceless stops showing more variability than the voiced stops (Fig. 4). Thus, variability in the organization of laryngeal and supralaryngeal gestures for the production of long-lag VOT does not diminish until older ages than those we examined in this study. So, while children's target VOTs for voiceless initial stops may have become more adultlike over the course of this study, their skill at achieving those targets did not improve (taking variability as the reciprocal of skill, e.g., Goodell and Studdert-Kennedy, 1993; Hodge, 1990; Kent, 1976). The right half of Fig. 4 presents variability data for children and adults from Nittrouer (1993). This graph shows that variability in the coordination of upper vocal tract and laryngeal gestures decreases with increased experience but that this decrease is not seen until children are older than they were when recordings were made for this study. This result for the coordination of laryngeal and supralaryngeal gestures in the production of *syllable-initial* stops labeled as voiced and voiceless mirrors what was found for the coordination of gestures in the production of *syllable-final* stops labeled as voiced and voiceless by Nittrouer *et al.* (2005): in that study, the duration of the vocalic syllable portion preceding voiced and voiceless final stops was measured. It was found that vocalic duration was more variable in samples from 5- and 7-year-olds than in those from adults, and this age-related trend was found only for syllables with voiceless final stops. As with syllable-initial voiceless stops, the timing of a supralaryngeal gesture (in this case, closing) must be precisely coordinated with a laryngeal gesture (in this case, abduction) in the production of syllable-final voiceless stops. The coordination of these gestures is neither as critical nor as difficult for syllable-final voiced stops as for syllable-final voiceless stops: final stops are categorized as voiced as long as voicing stops after the vocal tract closes, and that happens naturally when the pressure in the closed space of the supraglottal cavity equals the pressure of the subglottal cavity. For voiceless final stops, the speaker must make an explicit abduction gesture at just the right time.

The third specific question asked by us was whether children increase the number of stops intended as voiceless as they gain experience in producing their native language. The highly significant near-linear growth in the count of stops intended as voiceless (Fig. 5) indicates that these children did. Our procedures were sufficiently unstructured to have allowed children to produce as many words with either voiced or voiceless stops as they wanted. Nonetheless they showed a preponderance of words with voiced initial stops, particularly before 23–24 months. This evidence supports the position of others that children may avoid using words that they have difficulty producing during the earliest stage of word production (e.g., Schwartz and Leonard, 1982).

This study examined articulatory gestures in early speech production, focusing on the acquisition of stop consonants because voiceless stops require careful coordination in timing between the release of the oral closure and the onset of phonation. We found evidence that children gradually develop the timing between the larynx and the upper vocal-tract articulators that permit the perception by others of voiceless stops, supporting the concept of gestural reorganization over the first couple of years of life. The findings reported here also speak of the desirability of studying acquisition through the lens of gestural coordination: as children gained language experience, they were found to shift the timing between laryngeal and supralaryngeal gestures as needed to support the perception of voiceless initial stops. If we had not examined interarticulator coordination, all we could have reported would have been that voiced initial stops dominate early productions. There was also evidence of differences in developmental patterns among individual children in this study, but in no case did a child produce perfectly articulated and timed voiceless VOTs from the onset of first words.

The use of speech samples collected in unstructured play settings resulted in some challenges, but also strengthened the conclusions of this study and those of other investigators, as well. The biggest challenge was that decisions about whether specific stops were intended to be voiced or voiceless were not as clear cut as if an imitation task or a task with prenamed stimulus materials had been used. On the other hand, information was gathered about what actually happens in a naturalistic setting. Until now, the avoidance hypothesis has been tested primarily by presenting prenamed stimulus materials to young children and having them speak those names. Therefore, it was not known if children under the age of 2 years actually avoided the use of words with difficult (for them) phonetic sequences in natural settings. The results of this study show that children do appear to engage in such avoidance. That is, in typical conversational exchanges with a parent where no specific target was being solicited, children produced words with less complex articulatory coordination more frequently than those requiring more complex articulatory coordination.

In summary, this study found evidence to support the position that phonetic structure gradually emerges in children's speech production as they gain experience with their native language. The first words of four of these seven children adhered to the everything-moves-at-once principle in that the larynx was adducted at the same time that the upper vocal tract moved away from stop closure. Three of the children were able to delay the onset of voicing for a bit, but their laryngeal and supralaryngeal gestures were clearly not organized the same as those of adult speakers in their language community. All but one of the children showed evidence of reorganizing their speech-related gestures over the course of this study, but their variability remained high, indicating that their consistency in articulatory organization remained poor. Finally, all children initially seemed to avoid words that required the difficult task of delaying the onset of voicing relative to vocal-tract opening.

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC00633 from the National Institute on Deafness and Other

Communication Disorders, the National Institutes of Health, to S.N. We thank Carol Manning and Gina Meyer for their assistance in data collection and digitizing. Portions of this work were presented at the 147th Meeting of the Acoustical Society of America, New York, May 2004.

Allen, G. D. (**1985**). "How the young French child avoids the pre-voicing problem for word-initial voiced stops," J. Child Lang **12**, 37–46.

Barton, D., and Macken, M. A. (**1980**). "An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children," Lang Speech **23**, 159–169.

Bean, C. H. (**1932**). "An unusual opportunity to investigate the psychology of language," J. Genet. Psychol. **40**, 181–202.

BMDP Statistical Software (**1990**). BMDP Statistical Software, Inc., Los Angeles.

Browman, C. P., and Goldstein, L. (**1989**). "Articulatory gestures as phonological units," Phonology **6**, 201–251.

Bruner, J. S. (**1973**). "Organization of early skilled action," Child Dev. **44**, 1–11.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum, Hillsdale, NJ).

de Boysson–Bardies, B., Sagart, L., Halle, P., and Durand, C. (**1986**). "Acoustic investigation of crosslinguistic variability in babbling," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterström (Stockton, New York), pp. 113–126.

Dobrich, W., and Scarborough, H. S. (**1992**). "Phonological characteristics of words young children try to say," J. Child Lang **19**, 597–616.

Eilers, R., Oller, D. K., and Benito-Garcia, C. R. (**1984**). "The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study," J. Child Lang **11**, 313–336.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (**1971**). "Speech perception in infants," Science **171**, 303–306.

Ferguson, C. A., and Farwell, C. B. (**1975**). "Words and sounds in early language acquisition," Language **51**, 419–439.

Gilbert, J. H. (**1977**). "A voice onset time analysis of apical stop production in 3-year-olds," J. Child Lang **4**, 103–110.

Goffman, L., and Smith, A. (**1999**). "Development and phonetic differentiation of speech movement patterns," J. Exp. Psychol. Hum. Percept. Perform. **25**, 649–660.

Goodell, E. W., and Studdert-Kennedy, M. (**1991**). "Articulatory organization of early words: From syllable to phoneme," in Proceedings of the XIIth International Congress of Phonetic Sciences (Aix-en-Provence), pp. 166–169.

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.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (**2000**). "The physiologic development of speech motor coordination: Lip and jaw coordination," J. Speech Lang. Hear. Res. **43**, 239–255.

Hodge, M. M. (**1990**). "Measuring coarticulation in children's speech: Quirks and questions," Paper presented at the Child Phonology Conference, Madison, WI, May.

Jakobson, R. (**1941**). *Kindersprache, Aphasie und Allgemeine Lautgesetze*. Reprinted as *Child Language, Aphasia, and Phonological Universals*, 1968. Mouton, The Hague.

Kehoe, M. M., Lleo, C., and Rakow, M. (**2004**). "Voice onset time in bilingual German-Spanish children," Bilingualism: Lang. Cognit. **7**, 71–88.

Kelso, J. A., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (**1984**). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures," J. Exp. Psychol. Hum. Percept. Perform. **10**, 812–832.

Kent, R. D. (**1976**). "Tutorial: Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," J. Speech Hear. Res. **19**, 421–445.

Kent, R. D. (**1983**). "The segmental organization of speech," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 57–89.

Kewley-Port, D., and Preston, M. S. (**1974**). "Early apical stop production: A voice onset time analysis," J. Phonetics **2**, 195–210.

Koenig, L. L. (**2000**). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," J. Speech Lang. Hear. Res. **43**, 1211–1228.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by J. W. Hardcastle and A. Marchal (Kluwer Academic, The Netherlands), pp. 403–439.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Locke, J. L. (**1980**). "Mechanisms of phonological development in children: Maintenance, learning, and loss," in *Papers from the Sixteenth Regional Meeting of the Chicago Linguistic Society* (Chicago Linguistic Society, Chicago), pp. 220–238.

Locke, J. L. (**1983**). *Phonological Acquisition and Change* (Academic, New York).

Löfqvist, A., McGarr, N. S., and Honda, K. (**1984**). "Laryngeal muscles and articulatory control," J. Acoust. Soc. Am. **76**, 951–954.

Löfqvist, A., and Yoshioka, H. (**1980**). "Laryngeal activity in Swedish obstruent clusters," J. Acoust. Soc. Am. **68**, 792–801.

Macken, M. A., and Barton, D. (**1980a**). "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants," J. Child Lang **7**, 41–74.

Macken, M. A., and Barton, D. (**1980b**). "The acquisition of the voicing contrast in Spanish: a phonetic and phonological study of word-initial stop consonants," J. Child Lang **7**, 433–478.

MacNeilage, P. F., and Davis, B. (**1991**). "Acquisition of speech production: Frames, then content," in *Attention and Performance XIII*, edited by M. Jeannerod (Erlbaum, New York), pp. 453–476.

MacNeilage, P. F., and Davis, B. L. (**1993**). "Motor explanations of babbling and early speech patterns," in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Kluwer Academic, Dordrecht), pp. 341–352.

McGowan, R. S., Nittrouer, S., and Manning, C. J. (**2004**). "Development of [ɹ] in young, midwestern, American children," J. Acoust. Soc. Am. **115**, 871–884.

Menn, L. (**1978**). "Phonological units in beginning speech," in *Syllables and Segments*, edited by A. Bell and J. B. Hooper (North-Holland, Amsterdam), pp. 157–172.

Menn, L. (**1983**). "Development of articulatory, phonetic and phonological capabilities," in *Language Production: Development, Writing and Other Language Processes*, edited by B. Butterworth (Academic, New York), pp. 3–50.

Menyuk, P., and Klatt, M. (**1975**). "Voice onset time in consonant cluster production by children and adults," J. Child Lang **2**, 223–231.

Munhall, K. G. (**1985**). "An examination of intra-articulator relative timing," J. Acoust. Soc. Am. **78**, 1548–1553.

Nittrouer, S. (**1991**). "Phase relations of jaw and tongue tip movements in the production of VCV utterances," J. Acoust. Soc. Am. **90**, 1806–1815.

Nittrouer, S. (**1993**). "The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study," J. Speech Hear. Res. **36**, 959–972.

Nittrouer, S., Estee, S., Lowenstein, J. H., and Smith, J. (**2005**). "The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops," J. Acoust. Soc. Am. **117**, 351–364.

Nittrouer, S., Munhall, K., Kelso, J. A., Tuller, B., and Harris, K. S. (**1988**). "Patterns of interarticulator phasing and their relation to linguistic structure," J. Acoust. Soc. Am. **84**, 1653–1661.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Hear. Res. **32**, 120–132.

Nittrouer, S., Studdert-Kennedy, M., and Neely, S. T. (**1996**). "How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables," J. Speech Hear. Res. **39**, 379–389.

Oller, D. K., and Eilers, R. E. (**1982**). "Similarity of babbling in Spanish- and English-learning babies," J. Child Lang **9**, 565–577.

Osgood, C. E. (**1953**). *Method and Theory in Experimental Psychology* (Oxford University Press, New York).

Prather, E. M., Hedrick, D. L., and Kern, C. A. (**1975**). "Articulation development in children aged two to four years," J. Speech Hear Disord. **40**, 179–191.

Schwartz, R. G., and Leonard, L. B. (**1982**). "Do children pick and choose: An examination of phonological selection and avoidance in early lexical acquisition," J. Child Lang **9**, 319–336.

Schwartz, R. G., Leonard, L. B., Loeb, D. M., and Swanson, L. A. (**1987**). "Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance." J. Child Lang **14**, 411–418.

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., and Fletcher, P. (**2000**). "Co-

vert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*, edited by Michael Broe and Janet Pierrehumbert (Cambridge University Press, Cambridge), pp. 194–207.
Sjölander, K., and Beskow, J. (**2000**). "Wavesurfer: An open source speech tool," Proceedings of the Sixth International Conference on Spoken Language Processing, Vol. **4**, 464–467.
Smith, A., and Goffman, L. (**1998**). "Stability and patterning of speech movement sequences in children and adults," J. Speech Lang. Hear. Res. **41**, 18–30.
Smith, B. L., and McLean-Muse, A. (**1986**). "Articulatory movement characteristics of labial consonant productions by children and adults," J. Acoust. Soc. Am. **80**, 1321–1328.
Smith, B. L., Sugarman, M. D., and Long, S. H. (**1983**). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," J. Acoust. Soc. Am. **74**, 744–749.
Snow, D. (**1997**). "Children's acquisition of speech timing in English: A comparative study of voice onset time and final syllable vowel lengthening," J. Child Lang **24**, 35–56.
Studdert-Kennedy, M. (**1987**). "The phoneme as a perceptuomotor structure," in *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing*, edited by A. Allport, D. G. MacKay, W. Prinz, and E. Scheerer (Academic, Orlando), pp. 67–84.
Tyler, A. A., and Saxman, J. H. (**1991**). "Initial voicing contrast acquisition in normal and phonologically disordered children," Appl. Psycholinguist. **12**, 453–479.
Velten, H. V. (**1943**). "The growth of phonemic and lexical patterns in infant language," Language **19**, 281–292.
Vihman, M. M., and Velleman, S. L. (**1989**). "Phonological reorganization: A case study," Lang Speech **32**, 149–170.
Vihman, M. M., and Velleman, S. L. (**2000**). "The construction of a first phonology," Phonetica **57**, 255–266.
Werker, J. F., and Tees, R. C. (**1999**). "Influences on infant speech processing: Toward a new synthesis," Annu. Rev. Psychol. **50**, 509–535.
Whalen, D. H., Levitt, A. G., and Goldstein, L. M. (**2007**). "VOT in the babbling of French- and English-learning infants," J. Phonetics **35**, 341–352.
Zlatin, M. A., and Koenigsknecht, R. A. (**1976**). "Development of the voicing contrast: A comparison of voice onset time in stop perception and production," J. Speech Hear. Res. **19**, 93–111.

Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children

Lucie Ménard[a)]
Laboratoire de Phonétique, Département de Linguistique et de Didactique des Langues, Université du Québec à Montréal, Case Postale 8888, Succ. Centre-Ville, Montréal, Québec H3C 3P8, Canada

Pascal Perrier
GIPSA-Lab, Département Parole et Cognition, UMR CNRS No. 5009, Université Stendhal, Boîte Postale 25, 38040 Grenoble Cedex 9, France

Jerôme Aubin
Laboratoire de Phonétique, Département de Linguistique et de didactique des Langues, Université du Québec á Montréal, Case Postale 8888, Succ. Centre-Ville, Montréal, Québec H3C 3P8, Canada

Christophe Savariaux
GIPSA-Lab, Département Parole et Cognition, UMR CNRS No. 5009, Université Stendhal, Boîte Postale 25, 38040 Grenoble Cedex 9, France

Mélanie Thibeault
Laboratoire de Phonétique, Département de Linguistique et de Didactique des Langues, Université du Québec à Montréal, Case Postale 8888, Succ. Centre-Ville, Montréal, Québec H3C 3P8, Canada

(Received 6 November 2007; revised 22 May 2008; accepted 28 May 2008)

The relations between production and perception in 4-year-old children were examined in a study of compensation strategies for a lip-tube perturbation. Acoustic and perceptual analyses of the rounded vowel [u] produced by twelve 4-year-old French speakers were conducted under two conditions: normal and with a 15-mm-diam tube inserted between the lips. Recordings of isolated vowels were made in the normal condition before any perturbation (N1), immediately upon insertion of the tube and for the next 19 trials in this perturbed condition, with (P2) or without articulatory instructions (P1), and in the normal condition after the perturbed trials (N2). The results of the acoustic analyses reveal speaker-dependent alterations of F1, F2, and/or F0 in the perturbed conditions and after the removal of the tube. For some subjects, the presence of the tube resulted in very little change; for others, an increase in F2 was observed in P1, which was generally reduced in some of the 20 repetitions, but not systematically and not continuously. The use of articulatory instructions provided in the P2 condition was detrimental to the achievement of a good acoustic target. Perceptual data are used to determine optimal combinations of F0, F1, and F2 (in bark) related to these patterns. The data are compared to a previous study conducted with adults [Savariaux *et al.*, J. Acoust. Soc. Am. **106**, 381–393 (1999)]. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945704]

PACS number(s): 43.70.Ep, 43.70.Mn, 43.70.Fq, 43.70.Bk [BHS] Pages: 1192–1206

I. INTRODUCTION

In recent decades, artificial perturbation of the speech articulators has proven to be a very fruitful experimental paradigm. Indeed, substantial research conducted within this framework has shed light on the nature of the internal representations of vowels and consonants and the role of feedback (auditory, proprioceptive, etc.) in controlling the vocal apparatus (Aasland *et al.*, 2006; Jones and Munhall, 2003; Houde and Jordan, 1998; Guenther *et al.*, 1998; McFarland *et al.*, 1996; Savariaux *et al.*, 1995).

With regard specifically to compensatory abilities in children, perturbation experiments have led to somewhat contradictory results. For instance, in a bite-block experiment conducted on a 4- and an 8-year-old subject, Oller and MacNeilage (1983) concluded that children cannot achieve complete compensation in the spectral domain. Speakers were instructed to repeat productions of /i/ and /æ/ in both free-mandible and fixed-mandible conditions (with a bite-block). Spectrographic analysis revealed differences between the two conditions, suggesting that the children did not fully compensate for the perturbation, in the spectral domain. However, listeners judged the stimuli to be fairly good, which led the researchers to conclude that compensatory strategies could consist in preserving other acoustic parameters (duration, for instance). These results partly confirm Gibson and McPhearson's (1980) study. Those authors instructed 6- and 7-year-old subjects to produce Swedish vowels with and without bite-blocks inserted. Acoustic measure-

[a)]Electronic mail: menard.lucie@uqam.ca

 0001-4966/2008/124(2)/1192/15/$23.00

ments showed partial compensations for the perturbation in the spectral domain, a pattern which was confirmed by perceptual assessment (vowels were less accurately transcribed). However, a later articulatory study suggested that 4-year-old children exhibit adult-like compensatory abilities. Smith and McLean-Muse (1986) studied lip and jaw displacement and velocity in vowels produced in normal and bite-block conditions by three groups of subjects: 4- and 5-year-old children, 7- and 8-year-old children, and adults. Although the children produced more within-speaker articulatory variability than the adults, all three subject groups showed comparable compensatory abilities. Thus, the authors concluded that this ability is acquired early in childhood and requires very limited language experience. Similar results were obtained by Baum and Katz (1988), in a study of five speakers in each of the following age groups: 4- to 5-year-old children and 7- to 8-year-old children. Speakers were instructed to repeat the vowels [i a u] in normal and bite-block conditions. Acoustic measurements of the first two formant frequencies extracted at vowel onset and vowel midpoint revealed no significant differences between perturbed and unperturbed trials, for both groups of children. Furthermore, speakers compensated at vowel onset in the perturbed trials, suggesting that no auditory feedback was required in this process. Campbell (1999) also suggests that children do not rely on auditory feedback to produce compensatory articulatory strategies in speech.

Together, these studies suggest conflicting conclusions regarding 4-year-old children's compensatory abilities when they are instructed to produce vowels while the jaw is fixed by a bite-block. This type of perturbation does not modify the geometry of the vocal tract: It keeps one articulator from contributing to the constriction area and location typical of this vowel category, but without perturbing the geometry. Another type of perturbation consists of a lip tube inserted between the lips. This perturbation not only prevents the speaker from closing the lips while producing the vowel [u], for instance, but it also forces a complete reorganization of the articulatory and geometric strategies used to perform the speech task. This kind of perturbation was studied by Savariaux *et al.* (1995) in an articulatory and acoustic experiment conducted on 11 speakers producing the French vowel [u] in the normal condition and with a 2.5-cm-diam lip tube inserted between the lips (perturbed condition). During the first perturbed trial, upon insertion of the lip tube, seven speakers moved their tongues backwards, presumably to limit the expected deterioration of the acoustic signal. However, none of them achieved full compensation in this trial. In the remaining 19 perturbed trials, those seven speakers used auditory feedback to develop compensation strategies (backward movement of the tongue to shift the constriction location from the velo-palatal to the velo-pharyngeal region). One speaker showed complete compensation in the F1/F2 domain, while four speakers did not compensate at all. The authors proposed that the great between-speaker variability in the compensation strategies reveals the speaker specificity of the internal representation of the articulatory–acoustic relations in the region of the French vowel [u].

A perceptual study of the vowels produced in normal and perturbed conditions (Savariaux *et al.*, 1999) revealed that globally, instances of /u/ produced in perturbed conditions were less intelligible than those produced in the normal condition. Perceived vowels were represented in an acoustic–auditory space defined by linear combinations of F1, F2, and F0, in bark. A reinterpretation of the acoustic data in light of the perceptual space revealed that speakers who produced compensatory maneuvers altered F1, F2, and F0 in order to achieve an acoustic target. Thus, speakers had a good knowledge of the articulatory-to-acoustic relationships related to /u/.

In a follow-up study, Savariaux *et al.* (1997) performed an acoustic experiment aimed at examining the exact nature of F0 alterations in speakers and the role played by articulatory guidance. Subjects were instructed to produce /ogu/ sequences in normal and perturbed conditions (with the lip tube inserted between the lips). The results showed that three subjects achieved complete compensation in this condition. Thus, articulatory instructions inducing a general posterior positioning of the tongue, which is close to the appropriate tongue shape to compensate for the lip-tube perturbation, improved the speakers' ability to recalibrate their articulatory-to-acoustic maps. To our knowledge, no such study has been conducted with child speakers.

The present experiment was designed to extend Savariaux's (1995, 1997, 1999) studies and investigate abilities to compensate for a lip-tube perturbation in 4-year-old French children. If children already exhibit adult-like compensation strategies, it might be suggested that motor equivalence and internal representations of the speech apparatus are acquired early and do not require extended language experience. On the other hand, if children achieve only limited compensation for the perturbation, it would appear that more elaborate internal representations and motor control are required, together with extended language experience, to produce flexible articulatory strategies. The issue of the role of auditory feedback in formulating compensation strategies will also be addressed by comparing acoustic measurements extracted immediately upon insertion of the lip tube to measurements extracted in subsequent perturbed trials. Since the experimental design and data analysis method are similar to those of Savariaux's studies conducted with French adults (Savariaux *et al.*, 1995, 1999), results of the present study will be compared to those obtained with the adult group.

II. EXPERIMENT 1

A. Method

1. Subjects

Twelve French children ranging in age from 3 years 10 months to 4 years 11 months participated in this study. Speakers were recruited in a day-care center and were all native speakers of Continental French. The children were monolingual speakers of French living in the Grenoble (French Alps) area. They had no history of speech or language difficulties (as determined by the day-care center's

professionals). Children were also given a pretest screening by two of the experimenters in order to ensure that they had no oral cavity anomalies.

2. Material and procedure

Acoustic recordings of the French vowel [u] were made using a digital audio tape recorder (TASCAM) and a high-quality microphone. Subjects were instructed to repeat the isolated vowel [u] in four conditions: normal condition before lip tube insertion (hereafter N1), with the lip tube inserted between the lips (hereafter P1), with the lip tube inserted between the lips and the instructions to start from [o] (hereafter P2), and in the normal condition after removal of the lip-tube (N2). The P2 condition was added following Savariaux *et al.* (1997), who showed that this condition provided articulatory instructions that improved the subject's compensatory response. The instructions were as similar as possible to those provided to adult subjects in the studies of Savariaux *et al.* (1995, 1997). Throughout the perturbed phrases, the experimenters reminded the subjects that the target was the vowel /u/. The order of the two perturbed conditions was counterbalanced across subjects. Subjects produced 20 repetitions of [u] in each condition. For one subject, 12 repetitions were produced in P2 and 15 trials were recorded in P1. Six repetitions of each of the vowels [i a o ɔ œ] in normal condition were also recorded.

A 1.5-cm-long lip tube was built of Plexiglas™. The length was chosen so as to avoid lengthening the labial constriction. The diameter was chosen using simulations with an articulatory-to-acoustic model integrating nonuniform vocal tract growth [variable linear articulatory model (VLAM), Boë and Maeda, 1997]. This model is a scaled version of an adult model (Maeda, 1989), based on Goldstein's (1980) anatomical data from birth to adulthood. This model generates realistic vocal-tract shapes and has been used by our group in various studies (Ménard *et al.*, 2002, 2004, 2007). The model is controlled by seven articulators, which represent functional articulatory blocks: lip (height and protrusion), jaw height, tongue position (tip, body, back), and larynx height. The model provides a sagittal view of the vocal tract shape, the corresponding area function [based on Heinz and Stevens (1965)], and the transfer function (Badin and Fant, 1984). VLAM allowed us to determine the optimal diameter for the lip tube. The model was set at the 4-year-old stage, and the lip height parameter was increased so that the lip area value increased as well. A resulting value of 1.77 cm^2 was chosen, which corresponds to a 1.5 cm diameter. Figure 1 shows, as predicted by the model, the percentage variation in F1 and F2 resulting from the perturbation, relative to the prototypical [u] for that stage (N1). This condition, representing the formant values with the lip tube inserted but without any compensatory maneuver, will be referred to as "Pert." Figure 1 reveals that the insertion of the lip tube, without any compensatory maneuvers, results in a 42% increase in F1 and a 32% increase in F2. Those simulated changes are in the range of those reported in Savariaux *et al.* (1995, 1999), and the perturbation can thus be considered equivalent, considering the children's smaller vocal tract. Because /u/'s F1 and F2 are affiliated to Helmholtz

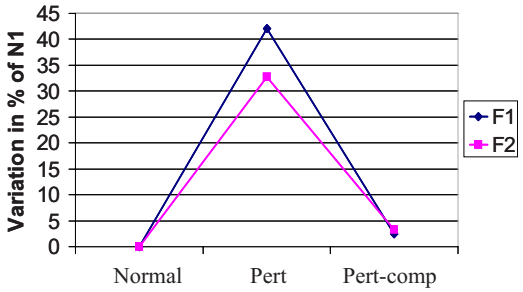


FIG. 1. (Color online) Predicted values of F1 and F2 in the normal condition (Normal), with the lip tube inserted but with no compensation (Pert), and with the lip tube inserted and compensation strategies (Pert-comp: Backward and downward movement of the tongue). Values are in percentage variation compared to the normal condition.

resonators, those formant frequencies are sensitive to various changes in constriction area and constriction length. Indeed, a Helmholtz's resonance frequency can be calculated by the following formula: $F=(c/2\pi k)\sqrt{(A_{co}/L_{co}V_{ca})}$, where c is sound velocity, k is a factor accounting for variation in the cross-section shape, A_{co} is constriction area, L_{co} is constriction length, and V_{ca} is cavity volume. In order to decrease formant values, several strategies can be used, possibly in combination: decrease constriction area, increase constriction length, or increase cavity volume. Local changes can be made in any of these dimensions, which results in small alterations in formant values. However, simulations carried out with VLAM suggest that the best compensatory strategy resulting in minimal formant alterations compared to the nonperturbed condition N1 (labeled "Pert-comp" in Fig. 1) involves a backward and downward movement of the tongue body, which increases constriction length and reduces constriction area (because of the shape of the palate, the downward movement of the tongue is necessary to avoid complete occlusion). Those strategies decrease the frequency of the affiliated formant (F2) and thus counterbalance the increase in frequency related to the decrease in cavity volume of the back cavity. The backward movement of the tongue increases the volume of the front cavity, thus decreasing F1. This displacement results in full compensation for the perturbation. The articulatory compensation strategy (Pert-comp) lowers F1 and F2 so that they almost reach the N1 value.

3. Data analysis

All vowels were digitized at a rate of 44 100 Hz with 16 bit quantization. Data were downsampled at 22 050 Hz after low-pass filtering at a cut-off frequency of 11 025 Hz, in order to obtain a more accurate formant detection in the [0, 4000 Hz] range. The first two formant frequencies were then extracted for each vowel, using the linear predictive coding (LPC) algorithm integrated in the PRAAT speech analysis program (Boersma and Weenink, 2007), at vowel onset (first glottal pulse) and at vowel midpoint. The number of poles varied from 10 to 14 (number of LPC coefficients from 20 to 28), which is in the range of parameters used by Lee *et al.* (1999) and Hillenbrand *et al.* (1995). A 14 ms Hamming window was used, with a pre-emphasis factor of 0.98 (pre-emphasis from 50 Hz for a sampling frequency of 22 050 Hz). It is well known that formant measurements are

particularly difficult to extract in high-pitched voices, due to the large distance between adjacent harmonics, leading to undersampled spectra. This is especially important for LPC analyses, in which formant measures are greatly influenced by the closest harmonic (Atal and Schroeder, 1974). However, LPC analysis is the procedure used by recent acoustic studies of child speech (Lee *et al.*, 1999; Hillenbrand *et al.*, 1995). Thus, we tried to avoid formant measurement errors by comparing, for each vowel, the automatically extracted formant values overlaid on a wide-band spectrogram with a spectral slice obtained by a fast fourier transform (FFT) computation with a Hamming window. When large discrepancies were observed either (i) between the overlaid formant values and the spectrogram or (ii) between the overlaid formant values and the spectral slice, the prediction order of the automatic detection algorithm was readjusted and the analysis was performed again.

4. Statistical analysis

For the production part of the study, three separate repeated-measure analyses of variance (ANOVAs) were carried out with experimental condition (N1, P1, P2, and N2) and measurement point (vowel onset, vowel midpoint) as the within-subject factors, using F1, F2, or F0 as the dependent variable. Data from the 12 subjects were included in the analysis. Another set of three repeated-measure ANOVAs was conducted on spectral measurements for the first and last trials only in each of the four conditions. Thus, the independent variables were experimental condition (N1, P1, P2, and N2) and trial number (first and last). For each analysis, interaction effects were further explored by planned comparisons with the alpha level set to 0.05.[1] Results for which p-levels are lower than 0.05 will be reported. This design was chosen to reveal global trends in speech production data, following Max and Onghena (1999). However, because between-speaker variability is often reported in such studies, individual behaviors will also be described. It should be noted that the order of the perturbed conditions (P1 followed by P2 or P2 followed by P1) did not significantly influence the data produced in each condition. Indeed, the results of T-tests carried out on F1, F2, and F0 values did not reveal a significant difference between the data for the six subjects who performed P1 before P2 and the data for the remaining six subjects who performed P2 before P1. Thus, in the following analyses, the order of the conditions will be presented but this factor was not included in the design of the ANOVAs.

B. Results and discussion

Mean F1 and F2 values for each subject and condition, at each measurement time (first glottal pulse or midpoint) are shown in Tables I and II. The measurements extracted at the vowel midpoint will be presented first, since we considered this data point to be representative of the vowel's target value. The measurement extracted at the vowel onset will be presented in order to evaluate whether compensation occurred immediately. Standard deviation values are presented in square brackets. For the sake of clarity, the percentage variation relative to the normal preperturbed condition (N1) is also presented. For each speaker, the order of elicitation of the perturbed conditions P1 and P2 is shown in subscript. Recall, however, that since no significant effect of this factor on the formant and F0 values was found, data in each of the perturbed conditions were pooled together and the elicitation order was ignored in the analysis. Graphic representations of the mean values and the standard deviations of the mean F1 and F2 values are provided in Fig. 2.

1. Mean spectral measures across conditions

F1 values. The data presented in Table 1, for the vowel midpoint in the P1 and P2 conditions, display considerable between-speaker variability. Some speakers produced the [u] vowel in the P1 condition with a minimal increase in F1 relative to N1 (such as S1, S5, and S11), whereas some others produced a large F1 increase in P1 relative to N1 (such as S2 and S8). When provided with articulatory cues (P2 condition), most subjects did not produce F1 values that were any closer to the values in the N1 condition. Thus, the difference in F1 values between the P2 and N1 conditions was no smaller than the difference in F1 between the P1 and N1 conditions; in fact, it was greater. An examination of the onset and the midpoint measurement values suggests a slight decrease in F1 throughout vowel duration for most speakers. In order to assess the effects of condition (N1, P1, P2, N2) and measurement point (vowel onset or midpoint), a two-way repeated-measure ANOVA was conducted on F1 values. The results revealed a significant main effect of experimental condition on F1 values [$F(3,33)=33.38; p<0.01$]. Post hoc tests revealed that F1 values were significantly higher in both perturbed conditions than in the normal preperturbed condition N1 [$F(1,11)=12.45; p<0.01$ for P1 and $F(1,11)=42.23; p<0.01$ for P2]. Values produced in the P2 condition were significantly higher than those produced in the P1 condition [$F(1,11)=6.00; p<0.05$]. In the postperturbed condition N2, F1 was significantly lower than in the normal preperturbed condition N1 [$F(1,11)=32.80; p<0.01$] and in both perturbed conditions: P1 [$F(1,11)=29.96; p<0.01$] and P2 [$F(1,11)=73.51; p<0.01$], suggesting the existence of a robust aftereffect.

Concerning the effects of vowel measurement point, the ANOVA revealed that F1 values were significantly higher at the vowel onset than the vowel midpoint [$F(1,11)=25.52; p<0.01$]. However, no significant effect of the interaction between measurement point and condition was observed, suggesting that F1 variations throughout the duration of the vowel do not arise from an adjustment in response to the perturbation.

F2 values. For F2 values [Table II and Fig. 2 (right-hand panel)], as was the case for F1, the measurements display significant between-speaker variability. In the P1 condition, at the vowel midpoint, all subjects but one (S7) produced the vowel [u] with an increase in mean F2 value relative to the N1 condition. However, for four subjects (S1, S4, S5, and S12), the mean percentage increase in F2 value is less than 10%. Turning now to the variation in F2 in the N2 condition, Table II shows that all subjects but one (S6) produced lower F2 values in the normal postperturbed condition N2, compared to the preperturbed condition N1, confirming a robust aftereffect, as was the case for F1 values

TABLE I. Mean F1 values (in hertz) for each subject, in the four experimental conditions and at two measurement points (onset and midpoint). Standard deviations are presented in square brackets. Percentage variation relative to the N1 condition is presented in parentheses.

	Onset				Midpoint			
	N1	P1	P2	N2	N1	P1	P2	N2
$S1_{P1\text{-}P2}$	434	395	571	410	396	414	444	423
	[41]	[15]	[32]	[11]	[21]	[15]	[25]	[14]
		(−9%)	(+32%)	(−5%)		(+5%)	(+12%)	(+7%)
$S2_{P1\text{-}P2}$	443	553	573	356	394	513	484	348
	[58]	[108]	[63]	[0]	[37]	[77]	[52]	[18]
		(+25%)	(+29%)	(−20%)		(+30%)	(+23%)	(−12%)
$S3_{P1\text{-}P2}$	472	556	618	439	444	439	520	420
	[46]	[98]	[78]	[40]	[35]	[34]	[66]	[46]
		(+18%)	(+31%)	(−7%)		(−1%)	(+17%)	(−5%)
$S4_{P2\text{-}P1}$	518	611	681	491	490	566	623	427
	[74]	[71]	[90]	[55]	[110]	[70]	[82]	[90]
		(+18%)	(+31%)	(−5%)		(+15%)	(+27%)	(−13%)
$S5_{P2\text{-}P1}$	465	581	638	396	387	401	561	371
	[131]	[100]	[86]	[24]	[10]	[24]	[69]	[32]
		(+25%)	(+37%)	(−15%)		(+4%)	(+45%)	(−4%)
$S6_{P2\text{-}P1}$	449	629	619	508	445	506	523	413
	[55]	[194]	[96]	[63]	[28]	[111]	[114]	[25]
		(+40%)	(+38%)	(+13%)		(+14%)	(+17%)	(−7%)
$S7_{P1\text{-}P2}$	360	530	540	407	534	476	550	426
	[15]	[128]	[151]	[31]	[65]	[77]	[34]	[70]
		(+47%)	(+50%)	(+13%)		(−11%)	(+3%)	(−20%)
$S8_{P1\text{-}P2}$	447	639	633	403	414	527	502	355
	[57]	[111]	[86]	[27]	[42]	[70]	[36]	[27]
		(+43%)	(+42%)	(−10%)		(+27%)	(+21%)	(−14%)
$S9_{P1\text{-}P2}$	362	556	650	374	491	549	533	390
	[83]	[124]	[136]	[15]	[104]	[55]	[56]	[79]
		(+54%)	(+80%)	(+3%)		(+12%)	(+9%)	(−20%)
$S10_{P2\text{-}P1}$	472	587	640	401	501	468	399	415
	[38]	[153]	[190]	[65]	[55]	[46]	[95]	[43]
		(+24%)	(+36%)	(−15%)		(−7%)	(−20%)	(−17%)
$S11_{P2\text{-}P1}$	419	556	588	429	571	573	551	505
	[28]	[70]	[69]	[48]	[41]	[80]	[78]	[36]
		(+33%)	(+40%)	(+2%)		(0%)	(−3%)	(−12%)
$S12_{P2\text{-}P1}$	427	555	546	411	382	425	379	353
	[54]	[132]	[71]	[32]	[17]	[68]	[47]	[25]
		(+30%)	(+28%)	(−4%)		(+11%)	(−1%)	(−8%)

[Table I and Fig. 2 (left-hand panel)]. However, the mean percentage decrease in F2 values ranges from −36% to −3%, suggesting variability across speakers.

A repeated-measure ANOVA with measurement point (vowel onset and vowel midpoint) and experimental condition (N1, P1, P2, N2) as the within-subject factors was carried out on the 12 subjects' mean F2 values. As shown in Fig. 2 (right-hand panel), a significant effect of measurement point was observed $[F(1,11)=51.32;p<0.05]$. Indeed, F2 was higher at vowel onset than at vowel midpoint. This difference is found in both perturbed conditions and in both normal conditions, as revealed by the lack of a significant interaction effect between measurement point and experimental condition. The effect of the condition factor was significant $[F(3,33)=48.78;p<0.05]$. Post hoc tests showed that F2 values in N1 were significantly lower than in P1 $[F(1,11)=32.99;p<0.01]$ and in P2 $[F(1,11)=22.81;p<0.01]$. F2 values observed in P1 did not differ from those measured in P2. Values in the N2 condition were significantly lower than values in N1 $[F(1,11)=22.69;p<0.01]$, suggesting a robust aftereffect, and they were lower than in both perturbed conditions P1 $[F(1,11)=93.47;p<0.01]$ and P2 $[F(1,11)=101.13;p<0.01]$.

F0 values. Mean F0 values are presented in Table III for each subject and each condition. Standard deviations are provided in square brackets and mean percentage increases in the P1, P2, and N2 conditions relative to the N1 condition are shown in parentheses. It is noticeable in Table III that the evolution of F0 values over the four experimental conditions varies among the 12 subjects. A repeated-measure ANOVA was computed on the mean F0 values for the 12 subjects with measurement point (vowel onset and vowel midpoint) and experimental condition as the within-subject factors. No significant effects of the factors, either as main effects or in interaction, were observed.

To summarize, an analysis of the trials produced in all conditions reveals that, overall, speakers were significantly affected by the lip tube. Indeed, F1 and F2 values were higher in the P1 and P2 conditions compared to N1. In the normal postperturbed condition N2, formant values were

TABLE II. Mean F2 values (in hertz) for each subject, in the four experimental conditions and at two measurement points (onset and midpoint). Standard deviations are presented in square brackets. Percentage variation relative to the N1 condition is presented in parentheses.

	Onset				Midpoint			
	N1	P1	P2	N2	N1	P1	P2	N2
$S1_{P1\text{-}P2}$	951 [54]	1118 [70] (+18%)	1313 [78] (+38%)	904 [51] (−5%)	1075 [39]	1175 [48] (+9%)	1292 [115] (+20%)	991 [22] (−8%)
$S2_{P1\text{-}P2}$	808 [61]	1390 [232] (+72%)	1348 [141] (+67%)	695 [24] (−14%)	839 [89]	1255 [152] (+50%)	1240 [74] (+48%)	726 [71] (−13)
$S3_{P1\text{-}P2}$	944 [85]	1237 [152] (+31%)	1265 [139] (+34%)	703 [113] (−26%)	982 [127]	1105 [113] (+12%)	1123 [85] (+14%)	709 [49] (−28%)
$S4_{P2\text{-}P1}$	1005 [112]	1350 [196] (+34%)	1244 [235] (+24%)	888 [101] (−12%)	1024 [171]	1110 [55] (+8%)	1099 [66] (+7%)	929 [83] (−9%)
$S5_{P2\text{-}P1}$	997 [26]	1250 [154] (+25%)	1185 [179] (+19%)	817 [50] (−18%)	1035 [64]	1074 [47] (+4%)	1105 [121] (+7%)	809 [87] (−22%)
$S6_{P2\text{-}P1}$	859 [115]	1434 [216] (+67%)	1350 [154] (+57%)	1009 [106] (+17%)	971 [35]	1176 [147] (+21%)	1212 [162] (+25%)	1006 [96] (+4%)
$S7_{P1\text{-}P2}$	752 [47]	1495 [485] (+99%)	1176 [217] (+56%)	693 [39] (−8%)	1204 [134]	1110 [136] (−8%)	1105 [74] (−8%)	775 [100] (−36%)
$S8_{P1\text{-}P2}$	790 [98]	1362 [390] (+72%)	1251 [280] (+58%)	739 [60] (−6%)	838 [38]	1147 [137] (+37%)	1043 [75] (+25%)	814 [105] (−3%)
$S9_{P1\text{-}P2}$	691 [84]	1220 [255] (+77%)	1278 [282] (+85%)	701 [50] (+1%)	884 [189]	1124 [121] (+27%)	1041 [91] (+18%)	720 [92] (−19%)
$S10_{P2\text{-}P1}$	794 [84]	1352 [445] (+70%)	1379 [547] (+74%)	788 [119] (−1%)	857 [103]	981 [91] (+15%)	1005 [108] (+17%)	792 [65] (−8%)
$S11_{P2\text{-}P1}$	820 [66]	1246 [204] (+52%)	1215 [345] (+48%)	725 [42] (−12%)	1050 [85]	1164 [100] (+11%)	1037 [71] (−1%)	872 [97] (−17%)
$S12_{P2\text{-}P1}$	929 [69]	1327 [429] (+43%)	1160 [332] (+25%)	810 [45] (−13%)	917 [48]	956 [127] (+4%)	915 [73] (0%)	827 [38] (−10%)

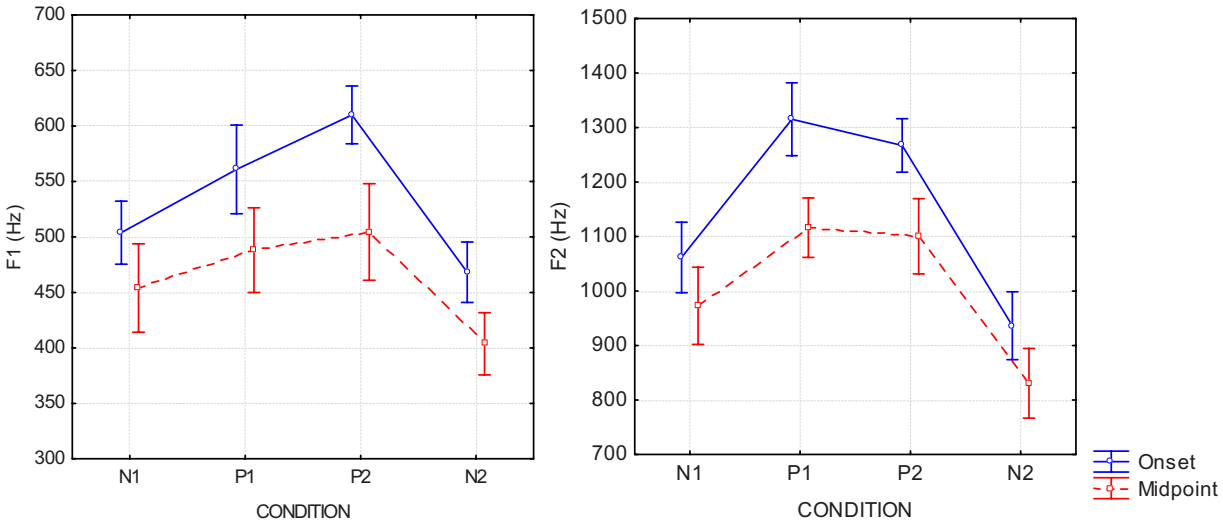


FIG. 2. (Color online) Mean and standard deviation values of F1 (left-hand panel) and F2 (right-hand panel) across subjects, in the four experimental conditions (N1, P1, P2, N2). Values measured at vowel onset correspond to the solid line, and values measured at vowel midpoint are depicted by the dashed line. All values are in hertz.

TABLE III. Mean F0 values (in hertz) for each subject, in the four experimental conditions and at two measurement points (onset and midpoint). Standard deviations are presented in square brackets. Percentage variation relative to the N1 condition is presented in parentheses.

	Onset				Midpoint			
	N1	P1	P2	N2	N1	P1	P2	N2
$S1_{P1\text{-}P2}$	323 [46]	362 [13] (+12%)	313 [60] (−3%)	378 [8] (+17%)	364 [16]	384 [19] (+5%)	428 [34] (+18%)	416 [13] (+14%)
$S2_{P1\text{-}P2}$	262 [34]	317 [29] (+21%)	256 [35] (−2%)	292 [27] (+12%)	308 [9]	323 [12] (+5%)	267 [16] (−13%)	316 [5] (+2%)
$S3_{P1\text{-}P2}$	277 [31]	260 [20] (−6%)	244 [25] (−12%)	231 [10] (−17%)	236 [3]	240 [7] (+2%)	241 [14] (+2%)	238 [5] (+1%)
$S4_{P2\text{-}P1}$	286 [46]	298 [38] (+4%)	285 [41] (0%)	260 [28] (−9%)	294 [15]	296 [6] (+1%)	283 [12] (−4%)	295 [9] (+1%)
$S5_{P2\text{-}P1}$	366 [22]	356 [33] (−3%)	323 [19] (−12%)	335 [28] (−8%)	374 [10]	357 [12] (−5%)	315 [15] (−16%)	335 [16] (−11%)
$S6_{P2\text{-}P1}$	386 [14]	371 [14] (−4%)	374 [16] (−3%)	352 [59] (−9%)	348 [11]	364 [14] (+5%)	356 [19] (+2%)	356 [19] (+2%)
$S7_{P1\text{-}P2}$	313 [19]	301 [21] (−4%)	295 [36] (−6%)	304 [45] (−3%)	295 [7]	305 [10] (+3%)	288 [8] (−2%)	290 [13] (−2%)
$S8_{P1\text{-}P2}$	322 [27]	302 [24] (−6%)	285 [33] (−12%)	316 [17] (−2%)	314 [4]	315 [14] (0%)	268 [10] (−15%)	303 [11] (−4%)
$S9_{P1\text{-}P2}$	301 [62]	312 [41] (+4%)	298 [37] (−1%)	324 [26] (+8%)	311 [31]	321 [22] (+3%)	271 [21] (−13%)	297 [17] (−5%)
$S10_{P2\text{-}P1}$	255 [15]	252 [19] (−1%)	237 [31] (−7%)	232 [15] (−9%)	279 [20]	250 [17] (−10%)	215 [13] (−23%)	244 [20] (−13%)
$S11_{P2\text{-}P1}$	342 [26]	319 [53] (−7%)	297 [75] (−13%)	291 [25] (−15%)	316 [6]	330 [18] (+5%)	330 [16] (+4%)	287 [7] (−9%)
$S12_{P2\text{-}P1}$	340 [37]	328 [46] (−4%)	320 [32] (−6%)	341 [33] (0%)	312 [16]	335 [26] (+8%)	313 [15] (0%)	314 [19] (+1%)

significantly lower than in the normal preperturbed condition N1. Considerable between-speaker variability was observed in the extent to which those acoustic parameters were affected by the perturbation.

2. Learning effects over the perturbed trials

In order to evaluate the variation in formant values within a given condition across the 20 trials, Fig. 3 presents

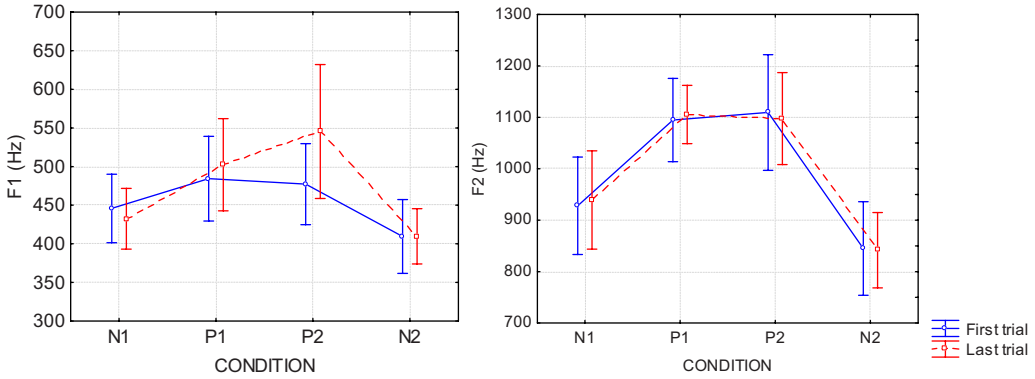


FIG. 3. (Color online) Mean and standard deviation values of F1 (left-hand panel) and F2 (right-hand panel) across subjects, in the four experimental conditions (N1, P1, P2, N2), for the first (solid line) and last trials (dashed line). All values are in hertz.

F1 (left-hand panel) and F2 values (right-hand panel) for the first and last trials in the four experimental conditions. Values measured at vowel midpoint are considered here. Mean values and standard deviations are represented for the 12 subjects. The solid line corresponds to the first trial, whereas the dashed line depicts the last trial. Two separate repeated-measure ANOVAs carried out on F1 and F2 values with trial number (first and last) and experimental condition (N1, P1, P2, N2) as the within-subject factors did not reveal any significant effect on the variation in formant values. Thus, overall, speakers did not show any tendency to alter their formant values from the first to the last trials in the F1 and F2 dimensions separately, suggesting that there was no learning effect. Furthermore, no significant within-subject correlations were found between trial number (from 1 to 20) and any of the acoustic parameters. It must be remembered, however, that those data include all subjects. As was observed in previous perturbation experiments (Savariaux *et al.*, 1995, for instance), subjects vary in the extent to which they respond to the perturbation.

In summary, in spite of considerable interspeaker variability, clear trends emerge from the analysis of the F1, F2, and F0 values produced by the subjects in normal and perturbed conditions: For the majority of the subjects, F1 and F2 are significantly modified by the insertion of the lip tube in the P1 and P2 conditions; none of the subjects systematically improved his/her (F1, F2) patterns during the perturbed phase in either the P1 condition or the P2 condition. These two observations suggest that, in general, subjects were not able to develop a robust compensation strategy within the 20 trials of the perturbed phase. However, a robust aftereffect is observed for all the subjects, suggesting the possibility that they may have learned new articulatory-to-auditory internal representations. Hence, speakers may have demonstrated abilities in the acoustic–auditory space, defined by linear combinations of spectral measures (Ménard *et al.*, 2002; Savariaux *et al.*, 1999). Section III presents a perceptual description of the stimuli produced by the subjects, to determine whether strategies to compensate for the presence of the lip tube were successful.

III. EXPERIMENT 2

Previous experiments have shown that invariant correlates of perceived vowels can be found in linear combinations of spectral parameters. The identification of openness, for instance, has been found to be related to the difference between F0 and F1 in various languages (Traunmüller, 1981; Syrdal and Gopal, 1986; Hoemeke and Diehl, 1994; Ménard *et al.*, 2002). Prior to the computation of this difference, hertz values are converted into a semilogarithmic scale, usually the "critical band units" (bark) scale, based on psychoacoustic experiments (Potter and Steinberg, 1950; Traunmüller, 1981). In French, high vowels (like /u/) would be related to a distance between F1 and F0 of less than 2 barks (Ménard *et al.*, 2002). Concerning place of articulation, the F3−F2 difference would account for the perception of this feature in English (Syrdal and Gopal, 1986), whereas F2−F1 would be involved for the same feature in Swedish (Fant, 1983). Hirahara and Kato (1992) propose that F2−F0 is related to the perception of place of articulation. Similarly, Savariaux *et al.* (1999) found that the perception of /u/ in French was associated with a low value for the F2−F0 difference and/or with a low F1 value. Thus, for /u/, combinations of formant frequencies and F0 appear to be involved in the perceptual identification of the speech target. The goal of this second experiment is twofold. First, the results will be used to determine to what extent subjects were able to achieve a good compensation strategy to produce the perceptual target related to /u/. Second, the acoustic parameters related to the perceptual identification of /u/ will be determined and used to reinterpret the production data presented in Sec. I.

A. Method

1. Subjects and stimuli

A subset of the produced vowels was used as stimuli for a perceptual test, as in Savariaux *et al.* (1999) and McFarland *et al.* (1996). For each speaker, the first five vowels produced in the normal preperturbed (N1) and postperturbed condition (N2) and 10 vowels in each of the perturbed conditions (P1 and P2) were selected. The selected vowels in the perturbed conditions were those from the first and last trials, as well as the odd-numbered trials 3, 5, 7, 9, 13, 15, 17, and 19. This selection was representative of the whole set of stimuli produced in each condition, while keeping the number of stimuli reasonably low. Separate T-tests with F0, F1, and F2 as the variables performed between the ten selected vowels for the perceptual test and the remaining ten vowels produced in the first experiment (N1) by each speaker did not reveal any significant differences between the two data sets. One repetition of each of the vowels [i a o] was also included. As a result, a total of 396 vowels (33 vowels $\times$ 12 speakers) constituted the set of stimuli. The total duration of the test was 40 min. Fifteen native speakers of French, ranging in age from 22 to 35 years old, served as subjects for the experiment. The participants did not report any history of auditory abnormality or speech production disorder. The test took place in a quiet room, on a Toshiba portable computer, using the perceptual experiment procedure implemented in PRAAT. Vowels were presented once binaurally via high-quality headphones. The tests consisted of an identification task and a quality-rating task. Participants had to (i) identify the vowel they heard from among the French oral vowels /i y u a o œ ɔ/, and (ii) rate the quality of the vowel they heard, if the vowel was /u/. For the latter task, seven choices were available: excellent /u/, very good /u/, good /u/, average /u/, bad /u/, very bad /u/, not a /u/. The participants had to select an icon displayed on the computer screen using the mouse.

2. Data analysis

In a first analysis of the perceptual responses, global identification scores were calculated. This parameter corresponds to the percentage of /u/ vowels correctly identified by the listeners. Then, in a subsequent analysis, only stimuli

TABLE IV. Mean percentage of produced [u] perceived as [u] for each subject, in the four experimental conditions. Percentage variation relative to the N1 condition is presented in parentheses.

	N1	P1	P2	N2
$S1_{P1\text{-}P2}$	91	79 (−13%)	86 (−5%)	91 (0%)
$S2_{P1\text{-}P2}$	73	55 (−25%)	15 (−80%)	97 (+33%)
$S3_{P1\text{-}P2}$	60	23 (−61%)	1 (−98%)	77 (+29%)
$S4_{P2\text{-}P1}$	47	19 (−60%)	1 (−99%)	100 (+114%)
$S5_{P2\text{-}P1}$	96	87 (−9%)	18 (−81%)	93 (−3%)
$S6_{P2\text{-}P1}$	73	48 (−35%)	42 (−43%)	92 (+25%)
$S7_{P1\text{-}P2}$	61	45 (−26%)	3 (−96%)	79 (+28%)
$S8_{P1\text{-}P2}$	88	34 (−61%)	7 (−92%)	97 (+11%)
$S9_{P1\text{-}P2}$	53	24 (−55%)	11 (−79%)	80 (+50%)
$S10_{P2\text{-}P1}$	73	36 (−51%)	8 (−89%)	69 (−5%)
$S11_{P2\text{-}P1}$	43	35 (−17%)	36 (−16%)	80 (+88%)
$S12_{P2\text{-}P1}$	61	68 (+11%)	59 (−3%)	100 (+63%)

which were perceived as /u/ by at least 50% of the listeners were included in the analysis. These stimuli will be referred to as the dominantly perceived vowels.

3. Statistical analysis

For the identification task of the perceptual test, identification scores were computed by dividing, for each vowel produced, the number of responses from the 15 listeners for which the perceived vowel corresponded to the produced vowel. For quality-rating responses, among the vowels perceived as /u/, three categories were considered, based on the average quality rating task: Vowels rated "very bad" or "bad" were pooled into one category (referred to as "bad"), vowels rated "average" were labeled "average," and vowels rated "good," "very good," or "excellent" were pooled in a third category labeled "good." Vowels rated "not an /u/" constituted a fourth category. Two separate repeated-measure ANOVAs were carried out on identification scores and mean goodness-rating scores with experimental condition as the within-subject factor (N1, P1, P2, and N2).

B. Results

1. Mean identification scores

The results of the perceptual test were used to determine to what extent subjects were able to achieve a good compensation strategy to produce the acoustic–auditory target related to /u/. Table IV presents the mean percentage of produced /u/ vowels perceived as /u/ in each condition, for each subject. In the P1, P2, and N2 conditions, the percentage variation relative to the N1 condition is presented in parentheses. Mean perceptual scores across the four experimental conditions are depicted in Fig. 4. First, Table IV clearly shows that, even in the N1 condition, the perceptual scores vary among speakers. Indeed, some speakers (such as S1 and S5) produced /u/ that were generally correctly identified by listeners (over 90% correct), whereas some others (such as S4 and S11) produced /u/ that were associated with low scores (47% and 43%).

A repeated-measure ANOVA was carried out on those values with condition (N1, P1, P2, N2) as the within-subject factor. A significant effect of condition was found $[F(3,11)=37.88; p<0.05]$. Planned comparisons revealed that the percentage of correct identification was higher in the normal preperturbed condition N1 than in the two perturbed conditions, P1 and P2 $[F(1,11)=37.88; p<0.05]$. Identification scores in N1 were in turn lower than in the normal postperturbed condition N2 $[F(1,11)=14.85; p<0.05]$, confirming the robust aftereffect found in the F1 versus F2 domain, and corresponding to a more canonical production than in the N1 condition. Furthermore, scores in P2 were lower than in P1 $[F(1,11)=13.25; p<0.05]$, suggesting that the use of articulatory instructions was actually detrimental to the achievement of a good acoustic–auditory target. On the basis of those perceptual criteria, then, compensation was better in the P1 condition than in the P2 condition, which was not observed for the F1, F2, and F0 dimensions independently.

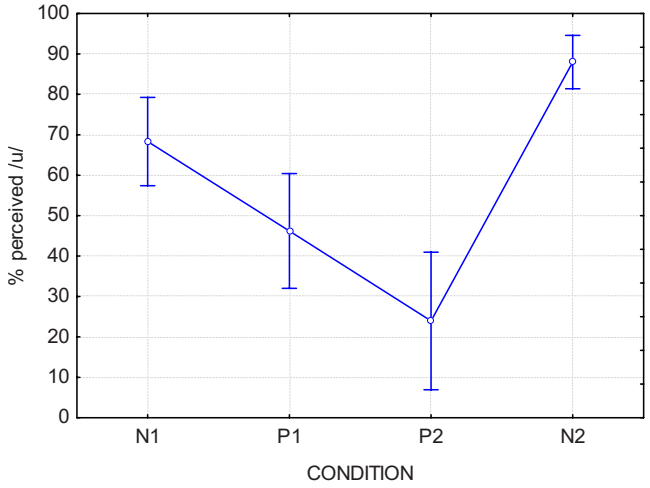


FIG. 4. (Color online) Mean and standard deviation values of identification scores across subjects, in the four experimental conditions (N1, P1, P2, N2).

2. Learning effects over the perturbed trials

In order to determine the possible learning strategies applied in the development of compensatory maneuvers over the course of the 20 trials, the identification of the first and last trials in P1 and P2 was examined. Data are plotted in Fig. 5 (P1 in the left-hand panel and P2 in the right-hand panel). Since considerable between-speaker variability is observed, the identification scores for each of the 12 subjects are presented separately in the graphs. Data corresponding to the first trial are depicted by the solid line, whereas values corresponding to the last trial are represented by the dashed line. The mean identification scores for the N1 condition correspond to the bars. It can be observed that, in the P1 condition, in the first trial, 2 speakers (S6 and S12) reached a perceptual score greater than or equal to their mean identification score in N1. Seven speakers improved the perceptual score from the first to the last trials. However, in P2, almost all speakers produced less intelligible vowels in the last trial compared to the first trial.

To evaluate to what extent the variation in identification from the first to the last trials reflects learning mechanisms, simple regression analyses were carried out between trial

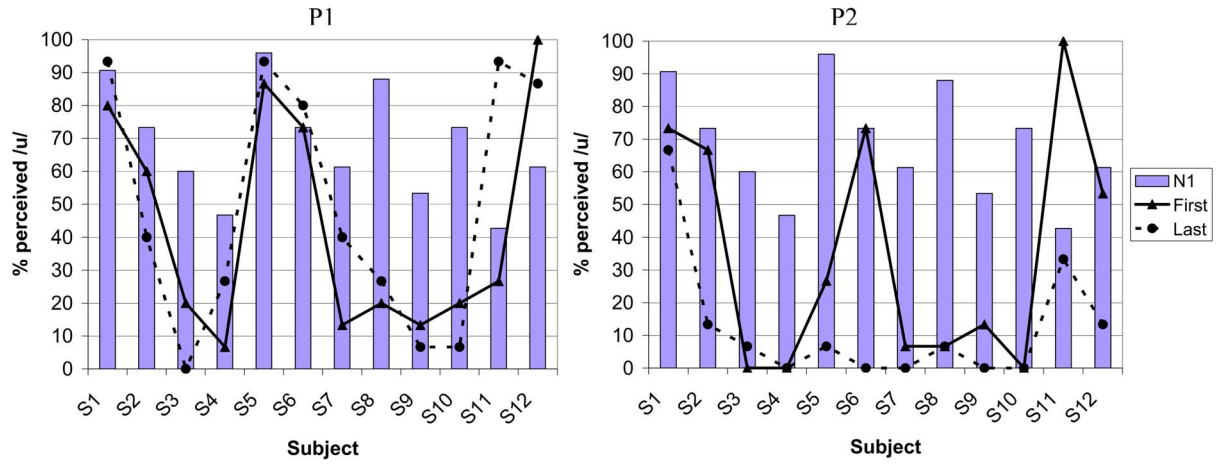


FIG. 5. (Color online) Identification scores of first (solid line) and last (dashed line) trials in P1 (left-hand panel) and P2 (right-hand panel), for each subject. Mean scores in the N1 condition are provided and correspond to the bars.

numbers and identification scores in P1 and P2, for each subject. No significant correlation was found. The variability observed for each speaker within the perturbed conditions might reflect the incapacity of the children to reproduce identically a vocal tract configuration, rather than their attempt to improve their production. However, as can be noticed in Tables I and II, the standard deviations measured for the acoustic parameters are higher in the two perturbed conditions P1 and P2 than in the two normal conditions N1 and N2. Two separate repeated-measure ANOVAs were carried out on F1 and F2, with measurement points and conditions as the within-subject variables. Both ANOVAs revealed a significant effect of the condition factor on the standard deviations [for F1: $F(3,33)=12.89, p<0.05$; for F2: $F(3,33)=13.99, p<0.05$], with higher values in P1 and P2 than in N1 and N2 [for F1: $F(1,11)=20.25, p<0.05$; for F2: $F(1,11)=23.71, p<0.05$]. The larger variability observed in both perturbed conditions compared to the normal preperturbed and postperturbed conditions supports the hypothesis of a search for a compensatory strategy to improve the /u/ production in presence of the perturbation. The absence of any evidence for a learning effect suggests that this search is based on a trial-and-error basis, probably guided by auditory feedback. Furthermore, those perceptual results suggest that compensatory strategies guided by articulatory information (P2) were not successful, since they did not allow the speakers to improve the identification scores of their vowels.

3. Best perceived /u/ in perturbed conditions

Since our results suggest that learning does not take place during the 20 trials of the training phase in the perturbed conditions (P1 and P2), it can be assumed that the last trial is not necessarily the best one in terms of perceptual efficiency. Hence, seeking out the best-perceived vowel within the trials produced in perturbed conditions could inform one about each subject's inherent articulatory capability to compensate for the perturbation, independently of the capacity to learn and memorize the best strategy. The percentage of produced /u/ correctly identified as /u/ for this best-perceived trial in P1 and P2 is depicted in Fig. 6 for each subject. Values in the P1 condition correspond to the solid line, and values in the P2 condition are depicted by the dashed line. For the sake of clarity, the mean identification scores for the normal trials in the N1 condition are also displayed. Figure 6 reveals that, in the P1 condition, all speakers produced at least one vowel associated with a higher or equal identification score than the mean identification score in the N1 condition. If compensation proficiency is evaluated through the ability to produce a vowel in the perturbed condition associated with an identification score equal to or greater than the mean identification score in the normal condition N1, then all subjects were able to achieve complete compensation in at least one stimulus, during the 10 trials. Seven speakers even had identification scores of greater than 80% and can thus be considered to be the best compensators: S1, S5, S6, S8, S10, S11, and S12. Hence, it can be concluded that all speakers had the articulatory skills to fully compensate for the perturbation that was introduced with the lip tube chosen for the experiment. The reasons why they did not do it permanently after having done it once do not originate from articulatory skills.

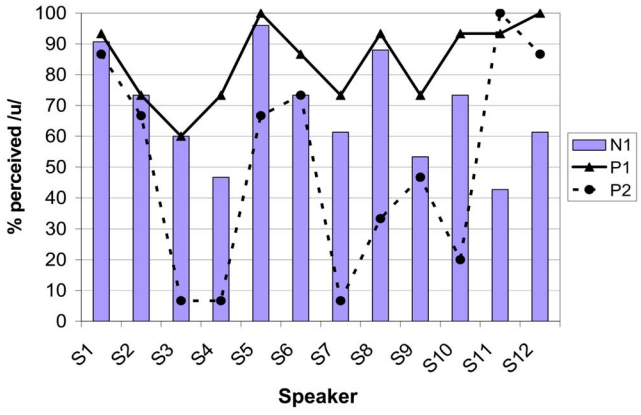


FIG. 6. (Color online) Identification scores of the best perceived stimulus in P1 (solid line) and P2 (dashed line), for each subject. Mean scores in the N1 condition are provided and correspond to the bars.

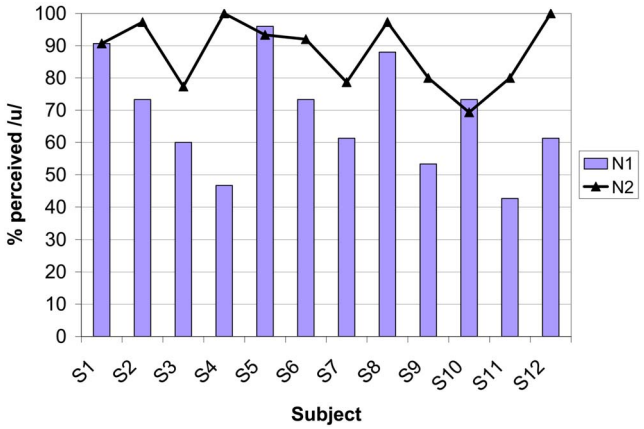


FIG. 7. (Color online) Mean identification scores in N1 and N2 conditions for each subject. Values for N1 correspond to the bars; values for N2 correspond to the solid line.

As could be expected from the results presented in Sec. III B 2, in the P2 condition, the best-perceived vowel is associated with a lower identification score than in the N1 condition for 9 speakers out of the 12 (S1, S2, S3, S4, S5, S7, S8, S9, and S10), suggesting that compensation was poorly achieved by these speakers in this condition. This result confirms that the use of the proposed articulatory instructions can be detrimental to 4-year-old subjects. For two subjects (S11 and S12), the best-perceived stimulus in P2 reached an identification score higher than in the N1 condition and greater than 80%. Thus, overall, the articulatory instructions provided in the P2 condition did not improve identification compared to P1.

4. Comparison of normal preperturbed (N1) and postperturbed (N2) conditions

The mean identification scores for the trials produced in the normal preperturbed condition N1 and the normal postperturbed condition N2 are depicted for each subject in Fig. 7. The scores obtained for the N1 condition correspond to the bars, whereas the scores obtained for the N2 condition correspond to the solid line. Comparing both values for each subject, it can be seen that, in agreement with the across-subject comparisons presented in Fig. 4, a robust aftereffect is observed. All speakers but three (S1, S5, and S10) improved their identification scores in the N2 condition compared to the N1 condition. Two of the speakers who did not show an increase in identification score from N1 to N2 (S1 and S5) did, however, produce near-perfect scores (over 90%), suggesting a ceiling effect.

To summarize, the analysis of the production and perception data in perturbed conditions show that (1) for the majority of the subjects, the insertion of the lip tube induces a perturbation which cannot immediately be compensated for; (2) all subjects are able to compensate for the perturbation at least once during the perturbed phase; (3) compensation strategies do not seem to be learned since no consistent improvement in production is observed from the beginning to the end of the learning phase; rather, compensation seems to be reached on a trial-and-error basis; (4) an aftereffect, which suggests the existence of some kind of learning process, is observed for all subjects including those for whom our measures do not show any evidence that they were perturbed by the insertion of the lip tube. This apparent contradiction between points (3) and (4) will be discussed in Sec. IV.

5. Acoustic parameters related to the target /u/

In order to characterize the spectral parameters corresponding to the perception of the target /u/ in French and to relate those parameters to the acoustic analysis, the stimuli used in the perception experiment were plotted in various spaces determined by the spectral parameters or by combinations of these parameters in order to look for the best clustering between perceived categories. Following Savariaux *et al.* (1999) and Ménard *et al.* (2002), stimuli for which at least 50% of the listeners perceived a given vowel category and quality (referred to as dominantly perceived vowels) were plotted in spaces consisting of various combinations of F1, F2, and/or F0. Graphic representations of two acoustic–

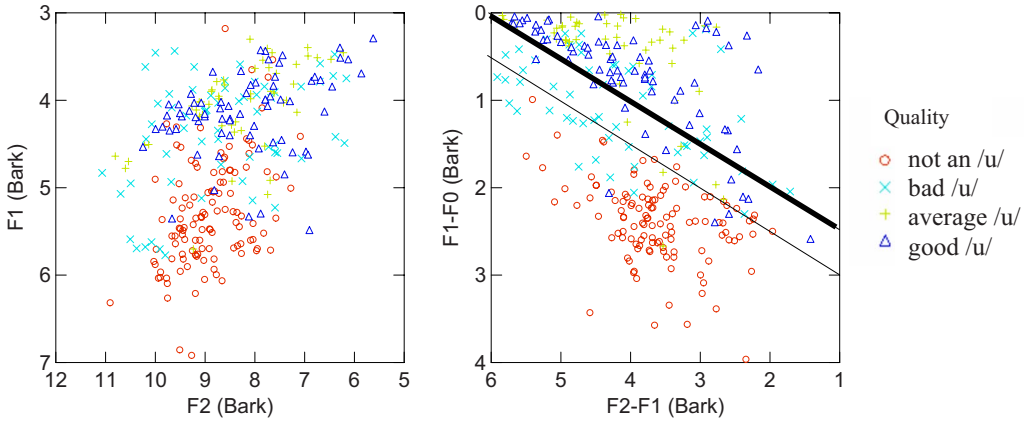


FIG. 8. (Color online) Dominantly perceived stimuli in the F1 vs F2 space (left-hand panel) and in the F2−F1 vs F1−F0 space (right-hand panel). All values are in bark. Circles correspond to stimuli rated "not an /u/," crosses correspond to stimuli rated "bad," plus signs correspond to stimuli rated "average," triangles correspond to stimuli rated "good." The thin solid line (right-hand panel) corresponds to the category boundary between perceived "not a /u/" and "bad /u/," for which (F2+F1)/2−F0=3.5 bark. The thick solid line (right-hand panel) corresponds to the category boundary between perceived "bad /u/" and "perceived average or good /u/," for which (F2+F1)/2−F0=3 bark.

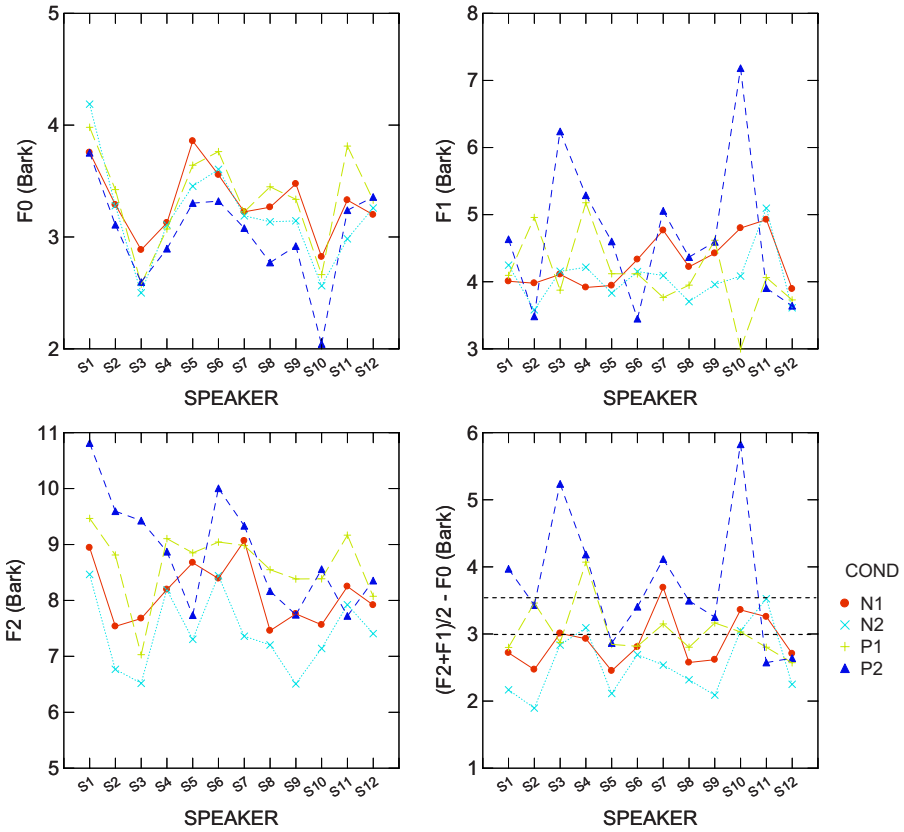


FIG. 9. (Color online) Values of F0 (upper-left-hand panel), F1 (upper-right-hand panel), F2 (lower left-hand panel), and (F2+F1)/2−F0 (lower right-hand panel) of the best perceived stimulus in each of the perturbed conditions P1 and P2, for each speaker. Mean values in N1 and N2 conditions are also displayed. All values in are bark.

auditory spaces are provided in Fig. 8. In the left-hand panel, data were plotted in the standard F1 versus F2 space (in bark). It can be seen that the four categories greatly overlap. In the right panel, the F1−F0 and F2−F1 space (in bark) is presented. These two parameters result in a much better classification of the categories. The thin solid line superimposed on the graph represents the boundary between the "not a /u/" and "bad /u/" categories. This boundary corresponds to the equation (F2+F1)/2−F0=3.5. Vowels for which this acoustic parameter is greater than 3.5 are perceived, in their large majority, as "not an /u/." The thick solid line in Fig. 8 (right-hand panel) corresponds to the category boundary between perceived bad /u/ and "good /u/ or average /u/." For this boundary, the value of the parameter (F2+F1)/2−F0 is 3 bark. Vowels for which this parameter is lower than 3 are identified as good /u/ or average /u/. Vowels for which (F2+F1)/2−F0 is greater than 3 bark and lower than 3.5 bark are perceived as bad /u/." A discriminant analysis performed on those values with (F1−F0) and (F2−F1) as the classification parameters and the three categories bad /u/, not a /u/, and average or good /u/ as the grouping factor revealed a percentage of correct classification of 85%. A similar discriminant analysis run with F1 and F2 as the classification factors and perceived category as the grouping factor resulted in a much lower percentage correct classification (70%). Thus, the addition of F0 as an acoustic parameter related to the perception of /u/ improves the classification of the data. These results are in line with previous work done on French (Savariaux *et al.*, 1999; Ménard *et al.*, 2002).

6. An interpretation of the compensation strategies in light of the perceptual data

The results of the perceptual test led us to propose that the parameters (F2+F1)/2 and F0, in bark, are related to the perceived target region associated with /u/ in the acoustic–auditory space. Those parameters can therefore be used to reanalyze acoustic data and better describe speakers' compensatory strategies. Figure 9 represents each speaker's best-perceived stimulus in the P1 and P2 conditions along four acoustic dimensions, in bark: F0, F1, F2, and ((F2+F1)/2−F0). Mean values in the normal conditions N1 and N2 are also displayed. Two category boundaries are represented by the dotted lines. Figure 9 thus represents the same data as those presented in Fig. 6, but in acoustical dimensions.

According to the data presented in Fig. 6, in the P1 condition, seven speakers produced one perturbed /u/ with an identification score of greater than 80%: S1, S5, S6, S8, S10, S11, and S12. Those speakers can be considered as the best compensators. An examination of the acoustic values for the corresponding speakers in Fig. 9 reveals various strategies. First, some speakers (S5 and S12) were able to strongly limit the impact of the perturbation with very few changes along the (F2+F1)/2−F0 dimension, with minimal alteration in F1 and F2 (less than 0.2 bark). Thus, these speakers were able to achieve a complete compensation for the perturbation for at least one /u/ among the 20 trials by producing almost identical F1 and F2 values as in N1. For some other speakers (S8, S10, and S11), F2 was increased by values ranging from 0.7 to 1.2 bark in P1 compared to N1, but F1 was decreased

by values ranging from 0.5 to 1.7 bark, thus maintaining the (F2+F1)/2−F0 complex below 3.5 bark. Four other subjects increased F0 values, hypothetically to counterbalance the increase of F2. The increase in F0 varies from 0.3 to 0.6 bark (see S11, in Fig. 9, upper left-hand panel), values in the range of those found by Savariaux *et al.* (1999). Although the interpretation of an active control of F1, F2, and F0 to maintain a low value of the (F2+F1)/2−F0 parameter is highly speculative, the fact that compensation can sometimes be observed in the (F2+F1)/2−F0 dimension but not in the F1, F2, and F0 dimensions separately reveals that the degrees of freedom required to produce an acoustic–auditory target are known and controlled early in life.

IV. GENERAL DISCUSSION

A. Compensatory mechanisms in children

The results presented so far demonstrated that compensatory abilities in 4-year-old French children are speaker dependent, a pattern also found in adults (Savariaux *et al.*, 1995, 1999). Perceptual data showed that upon insertion of the lip tube, in the first perturbed trial, two speakers (S6 and S12) were immediately able to compensate in the P1 condition. Indeed, those speakers produced /u/ with an identification score greater than, or equal to their mean identification scores in the N1 condition. Two speakers (S6 and S11) also achieved immediate compensation in the first trial in the P2 condition, as revealed by the identification scores. The remaining speakers, however, did not demonstrate compensatory abilities at the first perturbed trial, based on the variation in identification scores in the perturbed conditions compared to those in the normal preperturbed condition N1. Interestingly, no systematic improvement was observed during the perturbed phase, but in the P1 condition, all speakers have the articulatory skills to produce at least one good compensation (see Fig. 6). This shows (1) that all the subjects were able to compensate for the perturbation, and (2) that the motor control of speech is still too immature in 4-year-old children to allow generalization and learning in such a short time.

Turning now to the P2 condition, we found that only one speaker increased his identification score over the 20 trials (but to a value of 8%). Thus, for almost all speakers, the articulatory instructions provided in this condition were detrimental to the creation of successful compensatory strategies. One could object that speakers may not have understood the instructions related to the P2 condition. However, this is very unlikely since children were reminded during this perturbed phase by the experimenters that they had to go from /o/ and reach /u/.

It should be noted that none of the speakers demonstrated a real learning mechanism during the perturbed trials (either P1 or P2). Indeed, no significant correlations between trial number and identification scores were obtained, suggesting that the compensation strategies were not obtained by a gradual error correction mechanism operating on articulatory maneuvers. Rather, our 4-year-old subjects seem to have produced a good target in the presence of the lip tube on a trial-and-error basis, as revealed by the larger variability (measured by standard deviation) found in P1 and P2 compared to N1 and N2. Even though most of the children produced alterations along the acoustic dimensions that can be interpreted as strategies allowing them to achieve better compensation in the acoustic–auditory space, they could not store those strategies. This can be explained by the immaturity of their internal model of the articulatory–acoustic relations. This pattern contrasts with the adult data presented in Savariaux *et al.* (1995, 1999), in which speakers showed evidence of gradual improvement in the articulatory–acoustic domain. In this respect, it appears that 4-year-old children have a limited knowledge of the relationships between articulatory maneuvers and their acoustic consequences. Nevertheless, a robust aftereffect is observed in P1 and P2 conditions, even for the subjects who did not compensate. This contradiction deserves further analysis.

B. The detrimental effect of articulatory guidance

Identification scores in the P1 condition suggest that, for almost all subjects, articulatory changes were produced in order to reach the acoustic–auditory target. Changes in tongue shape and position, however, were likely local and did not involve movements as large as those required to fully compensate. As discussed in Sec. II A 2, simulations with VLAM have revealed that /u/'s formant frequencies are affiliated with Helmholtz resonators. Thus, local changes in constriction area and/or constriction length may have been done in order to modify formant frequencies while preserving, as much as possible, tongue position related to the original /u/. Thus, when lip area is increased, small changes in tongue position can alter formant values and enhance identification scores, even though those strategies are not optimal (Perkell *et al.*, 1993). For some speakers (because of the morphology of their vocal tract, for instance), however, this task may have been more difficult because of the changes of articulatory position these compensatory maneuvers required.

Another compensatory strategy predicted by our simulations with VLAM involves a large displacement of the tongue to a position closer to that in /o/. The articulatory cues provided in the P2 condition were intended to guide the speakers to this latter configuration. However, in the P2 condition, it can be hypothesized that the compensatory maneuvers were less successful because of the large distance between the produced somatosensory target and the intended somatosensory target (induced by the articulatory instructions concerning tongue position), the articulatory alterations induced by the instructions being too far from the somatosensory target associated with /u/. Thus, large changes in lingual articulatory configurations were not produced. This hypothesis is currently tested with articulatory measurements.

C. The nature of the speech task for 4-year-old children

The fact that over the course of the 20 trials, each subject has achieved a good compensation in the spectral domain associated with good perception scores supports the

hypothesis that these subjects were trying to reach a target in the auditory domain. The observation that none of the subjects kept producing the vocal tract configuration associated with a good auditory product certainly suggests that the auditory objective could have been not satisfactory *per se*. In addition, the inability of all subjects to achieve compensation in the perceptual domain in P2 condition suggests that somatosensory aspects contribute also to the specification of the task. In this context, one cannot completely discard the possibility that the discomfort associated with an unusual speaking condition could have been responsible for the larger intraspeaker variability observed in P1 and P2 conditions, simply because it forced the subjects to move their articulators around the canonical vocal tract configuration without any specific auditory goal. However, in such conditions the probability for every speaker to reach by coincidence the right auditory target over the course of a reduced number of trials would have been quite low. This is why, in line with the conclusions of Savariaux *et al.* (1995, 1999) for adults, we strongly favor the hypothesis of an auditory target for 4-year-old children as well (Perrier, 2005). In this context, the absence of systematic and continuous improvement over the course of the 20 trials in the P1 condition is explained by the fact that children would not integrate the auditory feedback in a form that can be processed in terms of learning because of immature internal models. This result points to the very distinct nature of the speech representations in 4-year-old children and adults. In the present study, children did not demonstrate a good knowledge of their articulatory capabilities in relation to their impact on acoustics. A closed-loop correction of articulatory positions on the basis of the minimization of the difference between the intended auditory feedback and the auditory feedback actually produced did not occur. This pattern is very different from the one obtained in adults by Savariaux *et al.* (1995, 1999). In the latter study, even though almost all speakers failed to achieve complete compensation by the end of the perturbed condition, all of them had produced some articulatory–acoustic maneuvers leading to a minimization of the discrepancy between intended and produced /u/.

The presence of the robust aftereffect is nonetheless intriguing. How could speakers use different strategies and improve their identification scores after removal of the tube if no learning mechanism had occurred in the perturbed conditions? We hypothesize that children establish associative links between multisensory representations (phonemic) and articulatory maneuvers, but that those links are not yet internalized in the form of an internal model of speech control. Children aim to produce a good acoustic–auditory target /u/ during the perturbed trials. After removal of the tube, their goal is still to produce a good target. In all likelihood, then, the observed aftereffect reflects the speakers' efforts to produce the canonical values of /u/ and increase their identification scores, rather than the maintenance of any compensatory maneuvers they used earlier. Further studies designed to investigate tongue shape and position in such perturbed conditions are currently in progress, with the hope that they will shed more light on this issue.

ACKNOWLEDGMENTS

This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Fonds Québécois de Recherche sur la Société et la Culture (FQRSC). The authors would like to thank David H. McFarland for insightful discussions. They are also grateful to Zofia Laubitz for copyediting the paper. They also thank the day care center and the subjects for their patience.

[1]The Bonferonni correction, which would have resulted in adapting the probability level to 0.0125, was not applied here, in order to follow as strictly as possible the method previously used in Savariaux *et al.* (1995, 1999) and Baum and Katz (1988).

Aasland, W. A., Baum, S. R., and McFarland, D. H. (**2006**). "Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation," J. Acoust. Soc. Am. **119**, 2372–2381.

Atal, B. S., and Schroeder, M. R. (**1974**). "Recent advances in predictive coding—Applications to speech synthesis," in *Speech Communication*, edited by G. Fant, (Almqvist and Wiksell, Stockholm, Sweden), Vol. **1**, pp. 27–31.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computation," STL-QPSR **2-3**, 53–108.

Baum, S. R., and Katz, W. F. K. (**1988**). "Acoustic analysis of compensatory articulation in children," J. Acoust. Soc. Am. **84**, 1662–1668.

Boë, L.-J., and Maeda, S. (**1997**). "Modélisation de la croissance du conduit vocal. Espace vocalique des nouveau-nés et des adultes. Conséquences pour l'ontogenése et la phylogenése" [Modelling vocal tract growth: Vowel space for newborns and adults. Consequences for ontogenesis and phylogenesis], *Journées d'Études Linguistiques. La Voyelle dans Tous ces États*, Nantes, pp. 98–105.

Boersma, P., and Weenink, D. (**2007**). PRAAT, Version 4.4.07, www.praat.org. Last viewed on 12 February 2007.

Campbell, M. (**1999**). "Articulatory compensation for a biteblock, with and without auditory feedback in hearing and hearing-impaired speakers," Doctoral dissertation, City University of New York, New York.

Fant, G. (**1983**). "Feature analysis of Swedish vowels—A revisit," STL-QPSR **2-3**, 1–19.

Gibson, A., and McPhearson, L. (**1980**). "Production of bite-block vowels by children," Phonetic Exp. Res. Inst. Linguistics Univ. Stockholm **II**, 26–43.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tract of the growing children," thesis of Doctor of Science, MIT, Cambridge, MA.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

Heinz, J. M., and Stevens, K. N. (**1965**). "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech," *Proceedings of the Fifth International Congress on Acoustics*, Liège, September 7–14, p. A44.

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. Tokhura, E. Vatikiotis-Bateson, and Y. Sagisaka, (Ohmsha/IOS Press, Tokyo), pp. 89–112.

Hoemeke, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: The role of F1−F0 distance," J. Acoust. Soc. Am. **96**, 661–674.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Jones, J. A., and Munhall, K. G. (**2003**). "Learning to produce speech with an altered vocal tract: The role of auditory feedback," J. Acoust. Soc. Am. **113**, 532–543.

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.

Maeda, S. (**1989**). "Compensatory articulation during speech," in *Speech*

Production and Modelling, edited by W. J. Hardcastle, and A. Marchal, (Kluwer Academic, Dordrecht), pp. 131–149.
Max, L., and Onghena, P. (**1999**). "Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research," J. Speech Lang. Hear. Res. **42**, 261–270.
McFarland, D. H., Baum, S. R., and Chabot, C. (**1996**). "Speech compensations to structural modifications of the oral cavity," J. Acoust. Soc. Am. **100**, 1093–1104.
Ménard, L., Schwartz, J.-L., and Boë, L.-J. (**2004**). "Role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for synthesized French vowels from birth to adulthood," J. Speech Lang. Hear. Res. **47**, 1059–1080.
Ménard, L., Schwartz, J.-L., Boë, L.-J., and Aubin, J. (**2007**). "Production-perception relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model," J. Phonetics **35**, 1–19.
Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., and Vallée, N. (**2002**). "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood," J. Acoust. Soc. Am. **111**, 1892–1905.
Oller, D. K., and MacNeilage, P. F. (**1983**). "Development of speech production. Perspectives from natural and perturbed speech," in *The Production of Speech*, edited by P. F. MacNeilage, (Springer, New York), pp. 91–108.
Perkell, J. S., Matthies, M. L., Svirsky, M. A., and Jordan, M. I. (**1993**). "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study," J. Acoust. Soc. Am. **93**, 2948–2961.
Perrier, P. (**2005**). "Control and representations in speech production," ZAS Pap. Linguist. **40**, 109–132.
Potter, R. K., and Steinberg, J. C. (**1950**). "Toward the specification of speech," J. Acoust. Soc. Am. **22**, 807–820.
Savariaux, C., Boë, L. J., and Perrier, P. (**1997**). "How can the control of the vocal tract limit the speaker's capability to produce the ultimate perceptive objectives of speech?," *Proceedings EUROSPEECH'97*, Rhodes, Greece, September 22–25, Vol. **2**, pp. 1063–1066.
Savariaux, C., Perrier, P., and Orliaguet, J.-P. (**1995**). "Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production," J. Acoust. Soc. Am. **98**, 2428–2442.
Savariaux, C., Perrier, P., Orliaguet, J.-P., and Schwartz, J.-L. (**1999**). "Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis," J. Acoust. Soc. Am. **106**, 381–393.
Smith, B. L., and McLean-Muse, A. (**1986**). "Articulatory movement characteristics of labial consonant productions by children and adults," J. Acoust. Soc. Am. **80**, 1321–1328.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
Traunmüller, H. (**1981**). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.

A simple-shear rheometer for linear viscoelastic characterization of vocal fold tissues at phonatory frequencies

Roger W. Chan[a)]
Department of Otolaryngology-Head and Neck Surgery and Graduate Program in Biomedical Engineering, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9035

Maritza L. Rodriguez
Graduate Program in Biomedical Engineering, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9035

(Received 27 September 2007; revised 27 May 2008; accepted 29 May 2008)

Previous studies reporting the linear viscoelastic shear properties of the human vocal fold cover or mucosa have been based on torsional rheometry, with measurements limited to low audio frequencies, up to around 80 Hz. This paper describes the design and validation of a custom-built, controlled-strain, linear, simple-shear rheometer system capable of direct empirical measurements of viscoelastic shear properties at phonatory frequencies. A tissue specimen was subjected to simple shear between two parallel, rigid acrylic plates, with a linear motor creating a translational sinusoidal displacement of the specimen via the upper plate, and the lower plate transmitting the harmonic shear force resulting from the viscoelastic response of the specimen. The displacement of the specimen was measured by a linear variable differential transformer whereas the shear force was detected by a piezoelectric transducer. The frequency response characteristics of these system components were assessed by vibration experiments with accelerometers. Measurements of the viscoelastic shear moduli (G' and G'') of a standard ANSI S2.21 polyurethane material and those of human vocal fold cover specimens were made, along with estimation of the system signal and noise levels. Preliminary results showed that the rheometer can provide valid and reliable rheometric data of vocal fold lamina propria specimens at frequencies of up to around 250 Hz, well into the phonatory range. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2946715]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.35.Mr [AL] Pages: 1207–1219

I. INTRODUCTION

Phonation is characterized by a flow-induced self-sustained oscillation of the vocal fold mucosa, involving primarily the vocal fold cover, or superficial layer of the lamina propria. The dynamics and energy transfer of vocal fold oscillation are dictated by the interactions between the aerodynamic stresses acting on the vocal fold surface and the mechanical response of the vocal fold tissue (Chan and Titze, 2006; Zhang *et al.*, 2007). In particular, the viscoelastic shear properties of the vocal fold cover are critical, because the mucosal wave that propagates on the cover during oscillation is a shear wave (Chan and Titze, 1999; Gray *et al.*, 1999). The viscoelasticity of the vocal fold cover under shear deformation contributes to the determination of the key parameters of phonation, such as the fundamental frequency, the amplitude of oscillation, and phonation threshold pressure (Chan and Titze, 2000, 2006; Titze, 2006; Zhang *et al.*, 2007).

Many previous studies for measuring the viscoelastic shear properties of the vocal fold cover have used parallel-plate torsional rheometry at small-strain amplitudes (around 0.01 rad) and small gap sizes (around 0.2–0.3 mm) to quantify the elastic shear modulus (G'), viscous shear modulus (G''), dynamic viscosity (η'), and damping ratio (ζ) within the linear viscoelastic region of the tissue (e.g., Chan and Titze, 1999, 2000; Chan, 2004; Klemuk and Titze, 2004; Titze *et al.*, 2004). The frequency range of viscoelastic measurement varied according to the rheometer used in the specific studies. Chan and Titze (1999) used a controlled-stress torsional rheometer (Bohlin CS-50) with a parallel-plate geometry to determine the linear viscoelastic shear properties of 15 excised human vocal fold mucosa (cover) specimens, across a frequency range of 0.01–15 Hz. Titze *et al.* (2004) used an improved controlled-stress torsional rheometer (Bohlin CVO-120), with higher resolutions and a more reliable controlled-stress mode. The system had a torque range of 0.0001–150 mN m and a resolution of 1 nN m, compared to the 0.001–10 mN m torque range and the 0.2 μN m resolution of the CS-50 model. Valid viscoelastic data were obtained at frequencies of up to 80 Hz, and when the gap size was decreased from 0.2 to 0.1 mm, the data could be valid for up to around 100 Hz.

Chan (2004) introduced the use of controlled-strain torsional rheometry, with a torsional shear strain applied to a tissue specimen and the shear stress response measured. Valid viscoelastic data of 17 canine vocal fold mucosa specimens were obtained at frequencies of up to around 50 Hz. Further measurements were made by Klemuk and Titze (2004) using the CVO-120 rheometer of Titze *et al.* (2004),

[a)]Author to whom correspondence should be addressed. Tel.: (214) 648-0386. FAX: (214) 648-9122. Electronic mail: roger.chan@utsouthwestern.edu

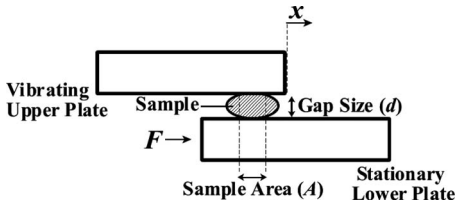


FIG. 1. The principle of simple-shear rheometry. A linear, simple-shear deformation is applied to a tissue or material specimen by the upper plate with a small-amplitude translational sinusoidal displacement x. A harmonic shear force F due to the viscoelastic response of the specimen is transmitted to the lower plate, separated from the upper plate by a gap size d. The contact area A between the specimen and the upper plate can be visualized from directly above through the transparent upper plate.

but in controlled-strain mode. Phonosurgical biomaterials such as collagen (Zyderm), a thiolated hyaluronic acid (HA) hydrogel (HA-DTPH), and micronized alloderm (Cymetra) were tested at frequencies of up to around 80 Hz. The highest frequencies at which valid data were obtained were found to be around 40 Hz for HA-DTPH, and around 80 Hz for Zyderm and Cymetra (Klemuk and Titze, 2004).

The typical fundamental frequency (F_0) range of speech is around 100–200 Hz for male and around 200–300 Hz for female. Although the frequencies of viscoelastic measurement had improved in previous studies, they were nonetheless still at the low end of the phonatory range (Chan, 2004; Klemuk and Titze, 2004; Titze *et al.*, 2004). Viscoelastic characterization of vocal fold tissues should be done at phonatory frequencies in order that the rheometric data become directly applicable to phonation, without having to rely on extrapolations or theoretical predictions (Chan, 2001, 2004; Chan and Titze, 2000). This paper reports the design and validation of a custom-built, controlled-strain, linear, simple-shear rheometer system capable of direct empirical measurements of viscoelastic properties at frequencies in the phonatory range.

II. SIMPLE-SHEAR RHEOMETRY

First of all, it is of interest to review the physics of oscillatory shear rheometry as applied to simple-shear deformation according to the theory of linear viscoelasticity (Ferry, 1980), as opposed to torsional rheometry (Chan and Titze, 1999; Titze *et al.*, 2004). Figure 1 shows the principle of linear, simple-shear rheometry, where a small-amplitude, translational sinusoidal displacement x is applied to a viscoelastic tissue or material specimen through a vibrating upper plate. The resulting shear force F due to the viscoelastic response of the specimen is transmitted to the lower plate, and can then be detected by a transducer. Consider the illustration in Fig. 1 as a single degree-of-freedom system, with lumped elements of mass, stiffness, and damping, the displacement of the upper plate x can be represented in complex notation as

$$x^* = x_0 e^{i\omega t}, \tag{1}$$

where x_0 is the amplitude of displacement, i is the imaginary number $\sqrt{-1}$, ω is the angular frequency, and t is time. The equation of motion for the system is

$$m\frac{d^2x^*}{dt^2} + c\frac{dx^*}{dt} + kx^* = F^*, \tag{2}$$

where m is mass, c is damping, k is stiffness of the system, and F^* is the resulting shear force in complex notation. For linear, small-amplitude shear, once steady state is reached, the applied sinusoidal displacement would result in a harmonic shear force at the same frequency, with a phase shift of δ according to the theory of linear viscoelasticity (Chan and Titze, 1999) as follows:

$$F^* = F_0 e^{i(\omega t+\delta)}, \tag{3}$$

where F_0 is the amplitude of F^*. For linear viscoelasticity, the phase shift δ will be independent of the displacement amplitude and the force amplitude (x_0 and F_0). Substituting x^* and F^* into the equation of motion yields

$$(k - m\omega^2)x_0 + ic\omega x_0 = F_0 e^{i\delta}. \tag{4}$$

The complex frequency response of the system $H(\omega)$ can be defined as the displacement divided by the force as follows:

$$H(\omega) = \frac{x_0 e^{i\omega t}}{F_0 e^{i\omega t} e^{i\delta}} = \frac{x_0}{F_0 e^{i\delta}}, \tag{5}$$

which results in the following expression from Eq. (4):

$$H(\omega) = \frac{1}{(k - m\omega^2) + ic\omega}. \tag{6}$$

One could define the undamped resonant frequency $\omega_n = \sqrt{k/m}$ and the damping ratio $\zeta = c/2\sqrt{km}$; the frequency response is then given by

$$H(\omega) = \frac{\frac{1}{k}}{1 - \left(\frac{\omega}{\omega_n}\right)^2 + i2\zeta\frac{\omega}{\omega_n}}. \tag{7}$$

By considering the stiffness factor $(1/k)$ together with the force term F_0, and by multiplying and dividing Eq. (7) by the complex conjugate of the denominator, a nondimensional expression can be derived for the complex system response $H(\omega)$ as follows:

$$H(\omega) = \frac{1 - \left(\frac{\omega}{\omega_n}\right)^2}{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2\zeta\frac{\omega}{\omega_n}\right]^2} - i\frac{2\zeta\frac{\omega}{\omega_n}}{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2\zeta\frac{\omega}{\omega_n}\right]^2}, \tag{8}$$

where the magnitude response $|H(\omega)|$ and the phase response $\delta(\omega)$ are given by

$$|H(\omega)| = \frac{1}{\sqrt{\left[1-\left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2\zeta\frac{\omega}{\omega_n}\right]^2}}, \tag{9}$$

$$\delta(\omega) = -\tan^{-1}\left(\frac{c\omega}{k-m\omega^2}\right) = -\tan^{-1}\left(\frac{2\zeta\frac{\omega}{\omega_n}}{1-\left(\frac{\omega}{\omega_n}\right)^2}\right). \tag{10}$$

By definition, the damped resonant frequency ω_0 is the frequency at which the magnitude of the frequency response $|H(\omega)|$ is maximum. It can also be deduced from Eq. (9) that ω_0 is related to the undamped resonant frequency ω_n by

$$\omega_0 = \omega_n\sqrt{1-\zeta^2} \quad \text{for } \zeta > 1/\sqrt{2}, \tag{11}$$

$$\omega_0 = \omega_n\sqrt{1-2\zeta^2} \quad \text{for } \zeta < 1/\sqrt{2}. \tag{12}$$

In order to compute the viscoelastic shear properties of the specimen, shear strain and shear stress of the specimen can be defined based on the displacement and force given in Eqs. (1) and (3). Shear strain can be defined as the displacement x divided by the distance between the plates, or the gap size d (Fig. 1) as follows:

$$\gamma^* = \tan^{-1}\frac{x^*}{d} \approx \frac{x^*}{d} \quad \text{for } x_0 \ll d, \tag{13}$$

whereas shear stress is defined as

$$\tau^* = \frac{F^*}{A}, \tag{14}$$

where A is the area of the specimen experiencing the strain, as indicated by the area of contact between the specimen and the upper plate (Fig. 1). Hence,

$$\gamma^* = \gamma_0 e^{i\omega t} = \frac{x_0}{d}e^{i\omega t}, \tag{15}$$

$$\tau^* = \tau_0 e^{i(\omega t+\delta)} = \frac{F_0}{A}e^{i(\omega t+\delta)}, \tag{16}$$

where γ_0 is the amplitude of γ^* and τ_0 is the amplitude of τ^*. The corresponding linear constitutive equation relating shear stress to shear strain is

$$\tau^* = G^*\gamma^*, \tag{17}$$

where G^* is the complex shear modulus. By definition, G^* is composed of a real part and an imaginary part as follows:

$$G^* = G' + iG''. \tag{18}$$

The real part G' is the elastic shear modulus, and the imaginary part G'' is the viscous shear modulus. Hence, for the linear theory of viscoelasticity (Ferry, 1980), the constitutive equation can be expressed as

$$\tau^* = G'\gamma^* + \frac{G''}{\omega}\frac{d\gamma^*}{dt}. \tag{19}$$

The elastic and viscous shear moduli G' and G'' are related to the strain amplitude and the stress amplitude (γ_0 and τ_0), and the phase shift (δ). Based on Eqs. (15) and (16), they are expressed in terms of the displacement amplitude and the force amplitude (x_0 and F_0) as follows:

$$G' = \frac{F_0 d\cos\delta}{Ax_0}, \tag{20}$$

$$G'' = \frac{F_0 d\sin\delta}{Ax_0}. \tag{21}$$

The dynamic viscosity η' is related to the viscous shear modulus G'', and the damping ratio (also called loss factor or damping factor) ζ is the ratio of the viscous to the elastic moduli as follows:

$$\eta' = \frac{G''}{\omega}, \tag{22}$$

$$\zeta = \frac{G''}{G'}. \tag{23}$$

The amplitude of the displacement x_0 can be detected by a displacement transducer at the upper plate, and the amplitude of the shear force response F_0 of the specimen can be detected by a force transducer at the lower plate. The phase shift δ can be measured with a temporal analysis of the displacement and force signals, where it is shown as (δ/ω) on the time axis. The area of contact between the specimen and the plates (A) can be estimated by analysis of scaled images taken from directly above the upper plate. With x_0, F_0, δ, and A determined, and given a known gap size d, the viscoelastic functions can be calculated from Eqs. (20)–(23). The design of the rheometer to quantify these viscoelastic properties based on this principle is illustrated next.

III. METHOD

A. Design of the simple-shear rheometer

The EnduraTEC ElectroForce (ELF) 3200 mechanical testing system (Bose Corporation, ElectroForce Systems Group, Eden Prairie, MN) was chosen as a base system to provide several key design criteria that are critical to the development of a controlled-strain rheometer for the measurements of tissue viscoelasticity at high frequencies. First, the system is capable of prescribing a precise oscillatory deformation through a linear motor. The linear motor design is modified from that of Bose subwoofers, involving a lightweight permanent magnet suspending in a controlled electromagnetic field, producing translational, oscillatory motion of the permanent magnet upon alternations of the electromagnetic field. The design does not include any mechanical seals or bearings, but the moving permanent magnet is attached directly to an actuator and a fixture through which a specimen is mounted and deformed. As a result, frictional energy loss in the driving mechanism of the motor is minimized, enabling the motor to maintain the same sinusoidal displacement at various amplitudes over a wide range of frequency. This is facilitated by displacement feedback control, which monitors the actual displacement of the actuator real time through a linear variable differential transformer (LVDT),

such that the target prescribed displacement is closely approximated. These features allow precise oscillatory shear deformation to be performed on specimens of varying stiffness both within and beyond the linear viscoelastic region, i.e., over a wide range of displacement amplitudes (up to around ± 6.50 mm) covering both small-strain (linear) and large-strain (nonlinear) oscillations.

Second, it is crucial that the shear force resulting from the viscoelastic response of the specimen upon deformation is detected by a force transducer capable of accurate and reliable measurements over a wide range of frequency. This is especially challenging as the magnitude of the force response depends on the elastic modulus of the specimen, and is typically in the millinewton range or smaller for soft tissue specimens. Piezoelectric force transducers designed with quartz crystals generating an electrical potential proportional to an applied force have the natural advantage of detecting dynamic forces at high frequencies. The output impedance of such transducers must be low ($<100\ \Omega$) so as to minimize signal degradation and the output noise level, especially for the low magnitudes of force for our application.

Third, inertial effects due to the system and the specimen should be minimized, whereas the frequency of system resonance should be maximized, in order to minimize measurement errors and time-dependent artifacts caused by system inertia, sample inertia, and system resonance (Chan, 2004; Titze *et al.*, 2004). The moving parts of the ELF 3200 system (shaft of the linear motor, the actuator, and the fixtures or plates in contact with the specimen) can be fabricated with lightweight acrylic material, leading to minimal inertial time delays and contributing to increase the system resonant frequency. The linear motor design enabling minimal friction in the moving parts over a wide range of frequency will contribute to the minimization of the system inertial effects during oscillation. Also, a piezoelectric force transducer with significant stiffness against deformation will also facilitate a higher system resonant frequency.

Taking into account these design criteria, a controlled-strain, simple-shear rheometer system was custom built based on the EnduraTEC ELF 3200 system (Fig. 2). As illustrated in Fig. 2, a tissue specimen is subjected to a linear, simple shear between two parallel, rectangular acrylic tissue plates according to the principle in Fig. 1. The upper plate is attached to the shaft of the linear motor through an actuator, applying a translational displacement x to the specimen at a specified magnitude and frequency. The linear motor is capable of a force range of ± 225 N (peak amplitude), an acceleration of up to 100 G, a displacement range of ± 6.50 mm, and a frequency range of 0.000 01–400 Hz. The motor is under displacement feedback control, with displacement of the upper plate detected by a lightweight LVDT (Schaevitz MHR 250; Measurement Specialties Inc., Hampton, VA) with minimal friction, in order to minimize system inertial errors. The shear force resulting from the viscoelastic response of the specimen to the applied strain is detected by a piezoelectric quartz force transducer (PCB Model 209C12; PCB Piezotronics, Depew, NY) rigidly attached to a stationary lower plate. The piezoelectric transducer has a force range of ± 4.45 N, a force resolution of 90 μN, and a frequency range of 0.05 Hz–30 kHz. The output impedance of the transducer is very low ($<100\ \Omega$), which minimizes output signal distortion and noise. The stiffness of 3.5×10^8 N/m is very high, which can minimize the system inertia. The discharge time constant of the piezoelectric quartz crystal is longer than 10 s, implying that DC signal drifting inherent in all piezoelectric transducers is not a significant source of error for dynamic measurements at 1–400 Hz.

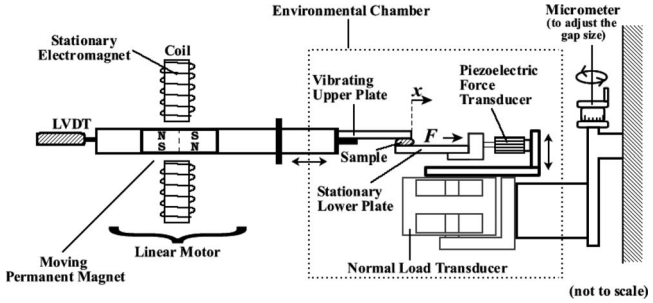


FIG. 2. Schematic of the custom-built, controlled-strain, linear, simple-shear rheometer system. A LVDT displacement transducer (Schaevitz MHR 250) is attached to the shaft of the linear motor through an actuator, for estimation of shear strain of the specimen. The resulting shear force (F) at the lower plate is detected by a piezoelectric force transducer (PCB Model 209C12). A normal load transducer measures the compressive force between the specimen and the plates. A micrometer allows one to adjust the gap size (d) between the two plates to accommodate specimens of varying dimensions. Mechanical testing is performed in an environmental chamber at controlled temperature and humidity.

The lower plate in contact with the tissue specimen is capable of displacement in the vertical (perpendicular) direction for adjustment of the gap size through a micrometer. The nominal gap size is 1.0 mm, although it can be set to anywhere from 0 to around 5.0 mm. There is also a normal load transducer with a force range of ± 2.45 N for detecting the compressive force between the specimen and the upper and lower plates. One purpose of the normal load transducer is to facilitate establishing a zero gap reference, when the normal force changes from zero to nonzero upon contact between the two tissue plates. The monitoring of normal load is also helpful for testing specimens of vastly different dimensions and volumes.

Rheometric measurements of viscoelastic properties are conducted in a transparent acrylic environmental chamber, where water at the base of the chamber is heated to a specified temperature ($\sim 50\ ^\circ$C) to maintain the ambient air temperature in the chamber at around 37 °C, with a relative humidity of close to 100% in order to minimize tissue dehydration. A digital camera is mounted directly above the chamber and photos of the specimen mounted between the tissue plates are taken. Scaled images of the specimen in the chamber are later examined with an image analysis software (IMAGE J, National Institutes of Health, Bethesda, MD), in order to determine the area of contact (A) between the specimen and the upper plate (Fig. 1).

The rheometer is controlled by the WINTEST software (Bose Corporation, ElectroForce Systems Group, Eden Prairie, MN), which allows one to specify a displacement control signal to be applied to the linear motor, with specific target

amplitude, frequency, and number of cycles prescribed and achieved through displacement feedback control. Data collection is performed on the displacement signal output of the LVDT and the force signal output of the piezoelectric force transducer, digitized at a rate of 8196 samples/s. The digitized signals are processed by the WINTEST software after the experiments, for calculating x_0, F_0, and δ of the two signals.

For the viscoelastic measurements to truly reach into the phonatory range, it is critical to validate the rheometer by establishing its frequency response characteristics, as well as the reliability and validity of the measurements. Two key system components, i.e., the LVDT displacement transducer and the piezoelectric force transducer, were validated by an assessment of their frequency response over a frequency range of up to 400 Hz, as detailed below.

B. Frequency response of the displacement transducer (LVDT)

The Schaevitz MHR 250 LVDT for the detection of displacement of the actuator has a displacement range of ±6.35 mm, a sensitivity of 68 mV/mm, and a nonlinearity (hysteresis) error of ±0.25%. It has a lightweight core with minimal friction that contributes to minimize system inertial errors. In order to assess the accuracy of displacement measurements of the LVDT as a function of frequency, an accelerometer (PCB Model 353B18; PCB Piezotronics, Depew, NY) was attached to the actuator of the ELF 3200 system. The procedure to establish the frequency response of the LVDT involved prescribing specific translational displacement amplitudes (0.1 and 0.05 mm) over a frequency range of 1–350 Hz. The target displacement was achieved by displacement feedback control, and was detected by the LVDT as the nominal displacement. The actual displacement of the actuator was derived from the acceleration measured by the accelerometer. The PCB 353B18 accelerometer has a range of 500 G and a sensitivity of 10.99 mV/G at 100 Hz. The frequency response of the accelerometer was assessed by the manufacturer at 23 °C and at a relative humidity of 20%, over a frequency range of 10–10 000 Hz. A flat response with little deviation (<1%) was found between 10 and 3000 Hz.

The nominal displacement amplitude given by the output of the LVDT was compared to the actual displacement amplitude of the actuator estimated from the accelerometer output. For sinusoidal signals, the amplitude of displacement x_0 is related to the amplitude of acceleration a_0 as follows:

$$x_0 = \frac{a_0}{\omega^2} = \frac{a_0}{(2\pi f)^2}, \tag{24}$$

where f is frequency in Hz. The ratio of the nominal displacement amplitude to the actual displacement amplitude from Eq. (24) is an indication of the frequency response of the LVDT.

C. Frequency response of the piezoelectric force transducer

The PCB 209C12 piezoelectric force transducer has been calibrated by the manufacturer for sensitivity statically using standard weights (average sensitivity of 517.7 mV/N) according to certified standards traceable to the National Institute of Standards and Technology (NIST) (ISO 9001, ISO 10012-1, ANSI/NCSL Z540-1, and ISO 17025), but not for frequency response. Two separate procedures were conducted to assess the accuracy of the force measurements at low frequency, as well as the transducer response at higher frequencies. Figure 3 illustrates the setup for low-frequency calibration, where a semiconductor strain gauge (Sensotec Model 31; Honeywell International Inc., Columbus, OH) previously calibrated at 1.0 Hz was used to calibrate the force output of the transducer, i.e., gain of the PCB signal conditioning interface between the transducer and the WINTEST software. Strain gauges are generally used for the accurate measurements of static and low-frequency forces due to the inherent stability of DC signals generated by semiconductors. As shown in Fig. 3, the shaft of the piezoelectric force transducer was rigidly attached to the actuator of the ELF 3200 via an acrylic tissue clamp, whereas the body of the piezoelectric transducer was attached to the shaft of the Sensotec Model 31 strain gauge, rigidly fixed on the ELF 3200. The strain gauge has a force range of ±4.9 N, a nonlinearity (hysteresis) error of ±0.15%, and a frequency range of DC to 740 Hz. It was calibrated by the manufacturer at 1.0 Hz according to NIST traceable standards (ANSI/NCSL Z540-1, ISO 9002 and ISO Guide 25). For calibration, translational sinusoidal deformation at 1.0 Hz was applied and the gain of the PCB signal conditioning interface was adjusted to match the piezoelectric transducer force output to that of the strain gauge. The linearity is verified by varying the amplitude of deformation. Results of the calibration showed that the force outputs of the two transducers were identical within a range of ±4.32 N, with error < ± 0.25% at 1.0 Hz.

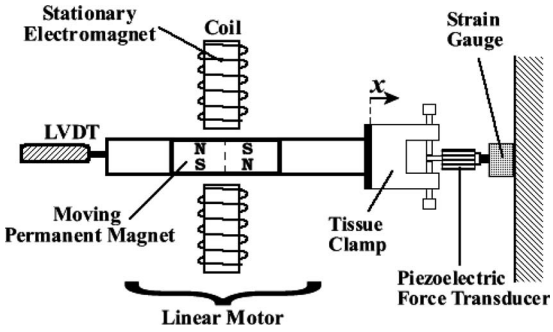


FIG. 3. Schematic of the setup for low-frequency calibration of the piezoelectric force transducer. The shaft of the piezoelectric force transducer is attached to the actuator through a tissue clamp. The shaft of a rigidly fixed, calibrated strain gauge (Sensotec Model 31) is attached to the body of the piezoelectric force transducer. Under sinusoidal oscillation at 1.0 Hz, the gain of the piezoelectric force transducer is adjusted to achieve identical dynamic output voltages from both the piezoelectric transducer and the strain gauge.

The frequency response of the piezoelectric transducer was assessed over a frequency range of 25–400 Hz. The piezoelectric transducer and an accelerometer (PCB Model 353B18) were both attached to the actuator of the ELF 3200, as shown in Fig. 4. As noted above, the frequency response of the accelerometer used was flat over a wide range of frequency (10–3000 Hz). Three different levels of mass were

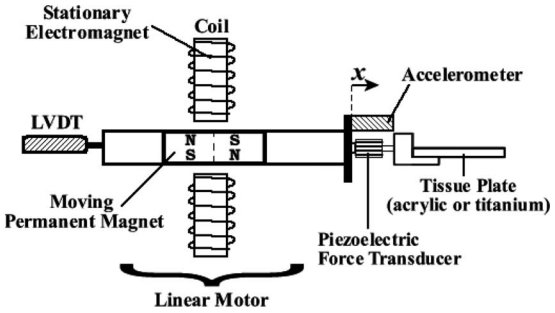


FIG. 4. Schematic of the setup for establishing the frequency response of the piezoelectric force transducer. The body of the piezoelectric force transducer is attached to the actuator, and varying mass (acrylic tissue plate, titanium tissue plate, or no tissue plate) is attached to the shaft of the transducer through an adapter. An accelerometer (PCB Model 353B18) is also attached to the actuator simultaneously for the measurement of acceleration.

attached to the shaft of the piezoelectric transducer, in order to vary the effective mass subjected to oscillation. They included a titanium tissue plate, an acrylic tissue plate, and no tissue plate (with the adapter only). In order to assess the frequency response of the piezoelectric transducer, varying sinusoidal displacement amplitudes were applied to the three conditions of masses, resulting in an acceleration amplitude of around ± 5 G for each mass condition, over a frequency range of 25–400 Hz. At each frequency, the dynamic force output was measured by the piezoelectric transducer while the acceleration of the system was measured by the accelerometer. Based on Newton's second law, the effective mass of vibration representing the total mass of all vibrating parts across different frequencies was obtained from dividing the dynamic force amplitude by the acceleration amplitude (F_0/a_0).

D. Estimation of system noise level

After establishing the accuracy and the frequency response of the LVDT displacement transducer and the piezoelectric force transducer, it was also important to determine the system signal-to-noise level, to reflect the functional frequency range within which measurements can be made with minimum errors. The system signal level was defined as the magnitude of the piezoelectric transducer force output resulting from the testing of specific specimens, whereas the system noise level was defined as the force output amplitude due to the electrical noise inherent in the system. The goal of this assessment was to compare the system noise level with the signal level, so as to determine the frequencies at which noise is or is not acceptable. For the measurement of system signal level, oscillatory shear deformation was conducted on an ANSI S2.21 standard polymer material and on seven human vocal fold cover specimens at frequencies of 1–300 Hz (see next two sections). For system noise, oscillatory shear deformation was conducted with no specimens mounted in the rheometer, under the exact same set of conditions as above.

E. Validation with a standard polymer material

The ANSI S2.21 standard material with known viscoelastic properties was used to validate the viscoelastic measurements of the rheometer system (American National Standards Institute, ANSI S2.21, 1998). This standard material was made for calibrating equipment that measures the dynamic mechanical properties of viscoelastic materials. The material is a polyurethane composed of polytetramethylene ether glycol (molecular weight 2000), 4,4′-diphenylmethane-diisocyanate, and a chain extender blend of 2,2-dimethyl-1,3-propanediol and 1,4-butanediol (ANSI, 1998). Three samples of the standard material were fabricated according to the standard procedure of ANSI S2.21, with the molar concentrations of the prepolymer and the chain extender reduced such that the shear modulus of the material is 1000 times lower (on the order of kPa rather than MPa) to be a closer match of the human vocal fold. The material samples were subjected to oscillatory shear in the rheometer, with a gap size of 1.0 mm, a displacement amplitude of 0.01 mm (1.0% strain), and over a frequency range of 1–300 Hz at 25 °C.

TABLE I. Human subjects characteristics.

Subject	Age	Gender	Race	No. of hours postmortem[a]
1	53	Female	African American	3
2	58	Female	Hispanic	4
3	60	Male	Caucasian	9
4	64	Male	Caucasian	15
5	69	Male	Caucasian	20
6	79	Male	Caucasian	11
7	88	Male	Caucasian	10

[a]The postmortem time was the time lapsed between a subject's death and the rheometric experiments, except for subjects 1 and 2 who were total laryngectomy patients.

F. Measurements of human vocal fold specimens

Five human larynges were excised from cadavers, obtained from the Willed Body Program of our institution within 20 h postmortem. Laryngeal specimens were also obtained from two additional subjects who underwent total laryngectomy due to supraglottic or thyroid cancer that did not involve the true vocal folds. Table I shows some basic characteristics of the subjects. Most of the subjects were Caucasians, but race was not a factor in the tissue procurement procedure. All of the subjects were nonsmokers, with no history of laryngeal disease and pathology. Physical examination of the laryngeal specimens revealed that the true vocal folds of all subjects were normal. The tissue procurement protocol and the experimental protocol were approved by the Institutional Review Board of the University of Texas Southwestern Medical Center. The cadaveric larynges were acceptable since previous research showed that the viscoelastic shear properties of vocal fold tissues are not significantly altered after 24 h of postmortem storage in saline at room temperature (Chan and Titze, 2003). Vocal fold specimens were dissected from the larynges with instruments for phonomicrosurgery, similar to the procedure of Chan and Titze (1999). Briefly, an incision was first made on the superior surface of the vocal fold epithelium with a surgical blade (No. 11), so that the vocal fold cover (epithelium and the

superficial layer of the lamina propria) could be separated from the vocal ligament (middle layer and deep layer of the lamina propria) through blunt dissection with a spatula similar to the Bouchayer spatula. This separation was facilitated by the natural plane of dissection between the superficial layer and the middle layer of the lamina propria. The vocal fold cover was then isolated from the larynx and kept in phosphate-buffered saline solution at a *p*H of 7.4 at room temperature prior to rheometric measurements. The volume of each vocal fold cover specimen was around 0.1–0.2 cm^3, requiring the gap size (d) of the rheometer to be set to between 0.5 and 1.0 mm so that there was complete contact between the specimen and the tissue plates with an area of contact smaller than the overlapping area of the tissue plates. Two testing protocols were conducted with the rheometer. First, a strain sweep was performed, which involved oscillatory shear deformation of the vocal fold cover specimens over a range of displacement amplitude (0.004–1.0 mm) (corresponding to 0.4%–100% strain for a gap size of 1.0 mm), at a constant frequency of 100 Hz. This served to identify the linear viscoelastic region of the specimens, i.e., the range of strain amplitude within which small-amplitude, linear, simple-shear deformation is ensured (Chan and Titze, 1999). Next, a frequency sweep was performed, which involved oscillatory shear deformation of the vocal fold cover specimens at a small-strain amplitude (1.0% strain), over a frequency range of 1–300 Hz.

G. Viscoelastic data analysis

The WINTEST program of the rheometer system was used to examine and process the digitized displacement signal and force signal at each specific frequency obtained from any particular test. Figure 5 shows the typical sinusoidal signals obtained from the testing of (a) no specimen (for estimating the system noise level), (b) a sample of the ANSI S2.21 standard material, and (c) a human vocal fold cover specimen (79-year-old male). The gap size was 1.0 mm and the displacement amplitude was 0.01 mm for all three cases (i.e., 1.0% strain). A window of at least 20 cycles of the digitized displacement signal and force signal was chosen based on the actual displacement amplitude reaching the target prescribed amplitude (±5%), which usually occurred a few seconds after the onset of oscillation. The positive and negative peak amplitudes and their corresponding time points of each cycle were manually (visually) selected and recorded for both signals. The data points were then processed by MATLAB for calculating the displacement amplitude x_0, the force amplitude F_0, and the phase shift δ between the two signals. With the area of specimen (A) measured and the gap size d given, linear viscoelastic shear properties of the specimen can be calculated according to Eqs. (20)–(23).

The reliability of the manual procedure for the selection of peak amplitudes and their corresponding time points of the digitized sinusoidal displacement and force signals was examined. Two raters completed the analysis of the same signals of a sample of the ANSI S2.21 standard material independently, each of them for three trials. The intrarater reliability was determined as the percentage difference in the viscoelastic functions obtained from the three trials of analysis performed by the same rater, and the inter-rater reliability was determined as the percentage difference in the viscoelastic functions obtained from the two raters. Results showed that the average percentage differences in G' and in G'' among the three trials completed by the same rater were 5.9571% (±4.2807%) and 6.0834% (±2.6233%), respectively. The average percentage differences in G' and in G'' among the trials completed by the two raters were 1.8413% (±1.7544%) and 4.1445% (±4.0660%), respectively.

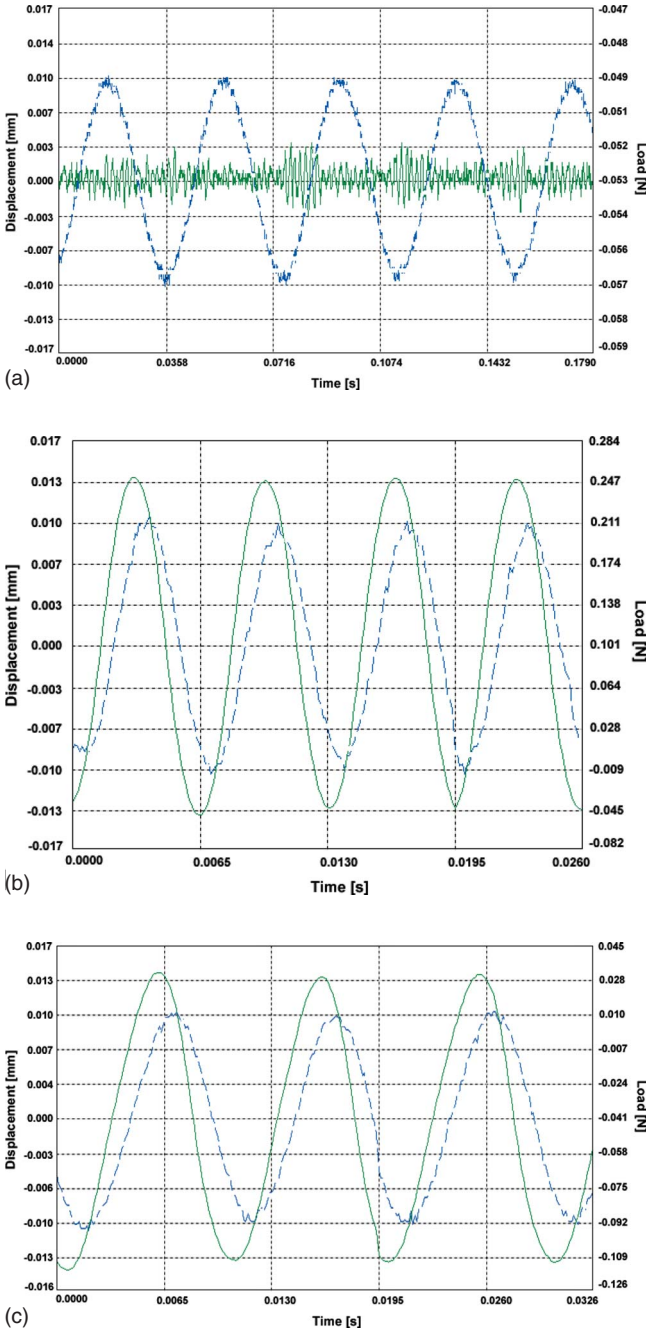


FIG. 5. (Color online) Typical sinusoidal waveforms of the displacement (x) and shear force (F) signals detected by the rheometer for the following: (a) no specimen, for the estimation of system noise level (frequency=25 Hz); (b) the ANSI S2.21 standard polymer material (frequency=150 Hz); and (c) a vocal fold cover specimen from the 79-year-old male (frequency =100 Hz). The displacement amplitude was 0.01 mm in all cases (dotted line=displacement; solid line=force).

IV. RESULTS AND DISCUSSION

A. Complex system response

The complex frequency response, or displacement-force response of the rheometer system $H(\omega)$ can be summarized by the magnitude response $|H(\omega)|$ and the phase response $\delta(\omega)$ as given by Eqs. (9) and (10). The undamped resonant frequency ω_n was given by $\sqrt{k/m}$, and the damping ratio $\zeta = c/2\sqrt{km}$. The effective mass m, stiffness k, and damping c of the system were determined from the setup for establishing the frequency response of the piezoelectric force transducer in Fig. 4. The effective mass m was computed from Newton's second law as the force amplitude divided by the acceleration amplitude (F_0/a_0), yielding an average of 0.000 645 5 kg. The effective stiffness k was determined as 1009.45 N/m when the phase shift between displacement and force was small, in which case $F_0 \approx kx_0$. Hence, the undamped resonant frequency ω_n was found to be 1250 rad/s (or 199 Hz). Next, the effective damping of the system was $c=0.802\ 512$ N s/m, as determined from $F_0 \approx c\omega x_0$ when the phase shift was close to $\pi/2$. This resulted in a damping ratio of $\zeta=0.4971$. With ω_n and ζ identified, the magnitude response $|H(\omega)|$ and the phase response $\delta(\omega)$ of the system can then be determined from Eqs. (9) and (10).

Figure 6 shows the system response as a function of frequency. The damped resonant frequency ω_0 was determined as the frequency at which the magnitude response $|H(\omega)|$ was maximum. Figure 6(a) shows the magnitude response, and it is clear that $|H(\omega)|$ reached a peak value of 1.1592 at ~142 Hz (or 892 rad/s) (as indicated by the dotted line), close to what is given by Eq. (12) ($\omega_0=889$ rad/s or 141 Hz). Figure 6(b) shows the phase response. As expected, the phase $\delta(\omega)$ became more negative with frequency. At the undamped resonant frequency of 199 Hz (dotted line), by definition the phase response δ should be at $-\pi/2$ (or $\pi/2$). It was found to be at −1.5714 rad, or close to $-\pi/2$ (−90°).

Given the damped resonant frequency of the system at around 142 Hz, it is reasonable to expect that valid viscoelastic data could be obtained with the rheometer system at frequencies of up to around 280 Hz (i.e., twice the resonant frequency) (Titze *et al.*, 2004). The data presented in the following sections on the frequency response characteristics of the displacement transducer and the force transducer, as well as the system noise level will verify whether meaningful data could indeed be obtained at such frequencies, well into the phonatory range.

B. Frequency response of the LVDT

The frequency response of the LVDT displacement transducer was assessed by vibration studies with an accelerometer attached to the actuator. Target displacement amplitudes were prescribed for the linear motor, with the nominal displacement amplitude detected by the LVDT compared to the actual displacement amplitude derived from the acceleration of the actuator according to Eq. (24). Figure 7 shows the displacement amplitudes over a frequency range of 1–350 Hz, for the target displacements of 0.10 and 0.05 mm. Results showed that the nominal displacement amplitude measured by the LVDT [Fig. 7(a)] was steady and close to the prescribed target amplitudes across the entire frequency range, with lower than 10% error at all frequencies.

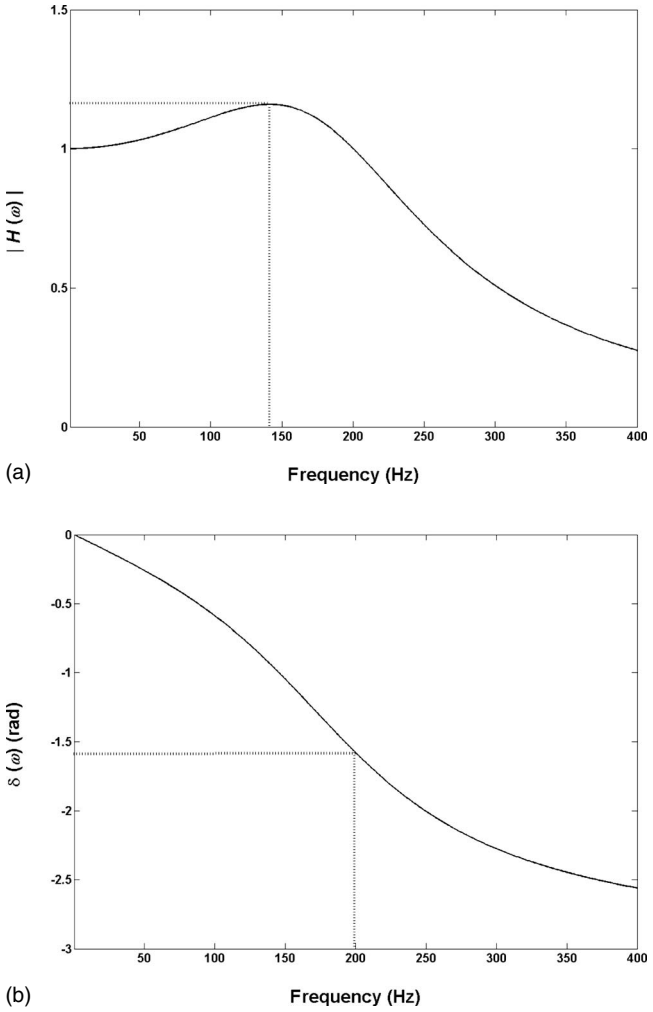
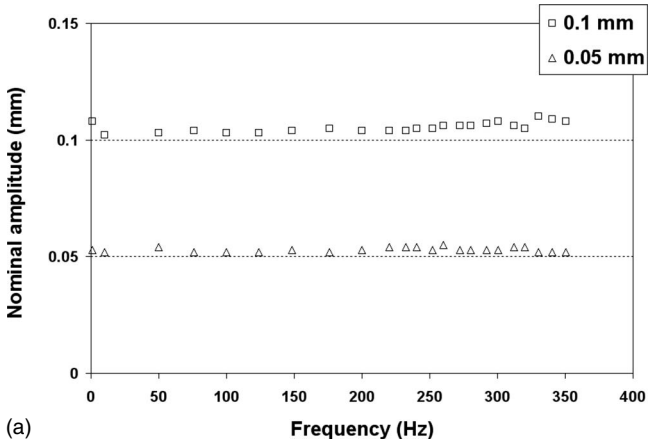
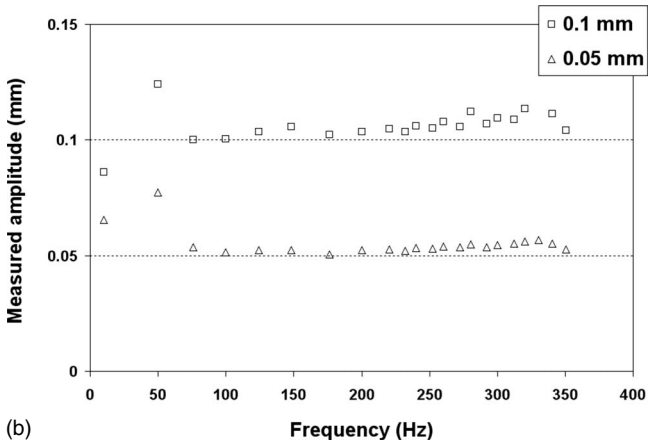


FIG. 6. Complex frequency response of the linear rheometer system over a frequency range of 1–400 Hz: (a) magnitude response $|H(\omega)|$ (dimensionless), and (b) phase response $\delta(\omega)$ in radians. In (a), the damped resonant frequency of the system is observed to be around 142 Hz, where $|H(\omega)|$ is maximum (dotted line). In (b), at the undamped resonant frequency (199 Hz), the phase is at around $-\pi/2$ (dotted line).

The actual displacement amplitude derived from the accelerometer [Fig. 7(b)] was as steady across most frequencies, except at low frequencies (50 Hz and below) where the magnitude of error in displacement ranged from 16% to 55%. This error was likely due to the inaccuracy of the peak-picking algorithm of the WINTEST software in selecting the peak amplitudes for low-frequency and low-magnitude acceleration signals, especially observed for the target amplitude of 0.05 mm. It should be noted that this error would not be incurred in the analysis of the displacement signal and the force signal for the viscoelastic measurement of specimens, because the selection of the signal peak amplitudes and their corresponding time points was performed manually, without the use of any peak-picking software algorithm. As stated in Sec. III the percentage errors in G' and in G'' resulting from the manual peak selection procedure during viscoelastic data analysis were at or below 6%.





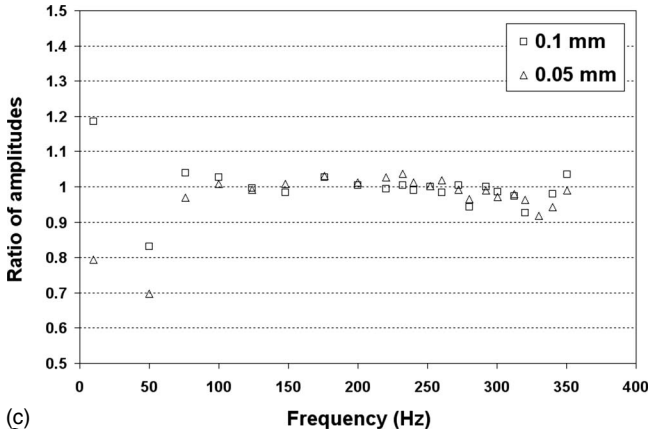


FIG. 7. (a) Nominal displacement amplitude of the LVDT as a function of frequency, with target amplitudes of 0.05 and 0.1 mm. (b) Measured displacement amplitude of the LVDT as estimated by the accelerometer as a function of frequency. (c) Ratio of the nominal displacement amplitude to the measured amplitude of the LVDT. This ratio indicates the frequency response of the LVDT, which is seen to be flat between around 75 and 275 Hz.

Figure 7(c) shows the ratio of the nominal displacement amplitude to the actual displacement amplitude as a function of frequency. This ratio was an indication of the frequency response of the LVDT, and it was found to be very close to unity at frequencies of 75–300 Hz (within 4% of 1.0, except for one data point at 280 Hz, which was 5.7% smaller than 1.0). Hence, it was determined that the LVDT displacement transducer is accurate within this frequency range. The ratio

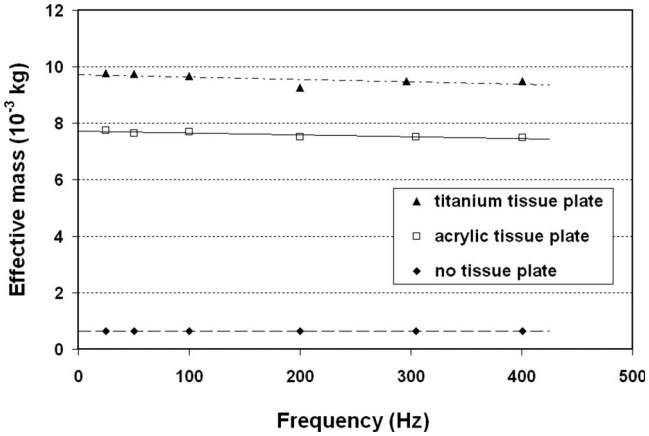


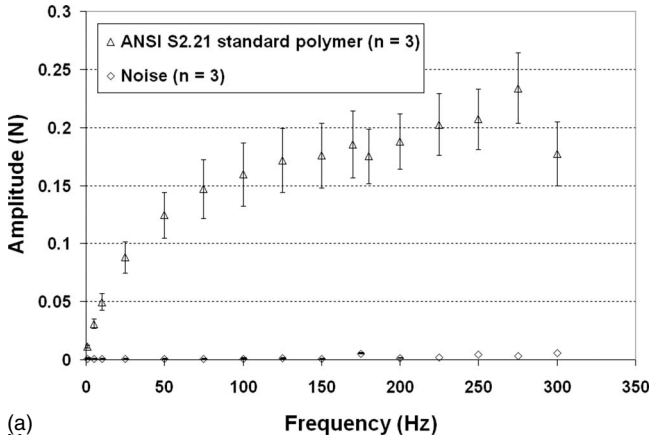
FIG. 8. Frequency response of the piezoelectric force transducer as indicated by the effective mass of vibration over a frequency range of 25–400 Hz. Three levels of mass are shown corresponding to a titanium tissue plate, an acrylic tissue plate, and no tissue plate mounted to the adapter (Fig. 4). A flat frequency response can be seen for the piezoelectric transducer over the entire frequency range examined.

at low frequencies (at or below 50 Hz), however, was not as close to unity due to errors in the measurement of acceleration amplitudes, as described above. Since the LVDT demonstrated a steady output in the nominal displacement amplitude [Fig. 7(a)], the errors at low frequency may only reflect those of the accelerometer response. In order to confirm that the LVDT transducer is also accurate at frequencies below 50 Hz, further vibration experiments involving accelerometers with higher output magnitudes are warranted.

C. Frequency response of the piezoelectric force transducer

The frequency response of the piezoelectric force transducer was examined by applying an acceleration amplitude of around ±5 G to three levels of mass attached to the actuator of the ELF 3200 over a frequency range of 25–400 Hz, as shown in the setup in Fig. 4. The force amplitude was measured by the piezoelectric transducer and the acceleration amplitude was measured by the accelerometer, allowing for the calculation of the effective mass of vibration as an indication of the frequency response of the piezoelectric transducer.

Figure 8 shows the results of this assessment. As expected, it is clear that the effective mass of vibration was the highest for the titanium tissue plate condition, followed by the acrylic tissue plate condition, and the condition of no tissue plate (with only the adapter attached to the actuator of the ELF 3200). The average difference in effective mass across all frequencies between the titanium tissue plate condition and the no tissue plate condition was 8.9229 g, whereas the average difference between the acrylic tissue plate condition and the no tissue plate condition was 6.9542 g. These differences were within 2.4% of the actual physical mass of the two tissue plates (9.155 and 7.124 g, respectively), validating the current approach for the assessment of frequency response. It is clear that the effective mass did not vary significantly across all frequencies for all of the



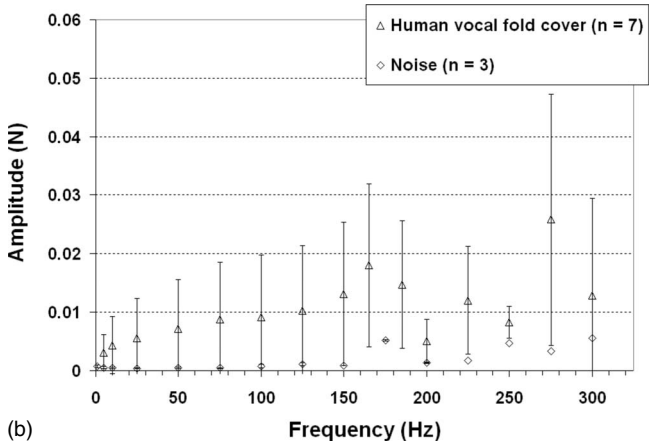


FIG. 9. (a) Comparisons of the sinusoidal force amplitude (system signal level) for the ANSI S2.21 standard polymer material ($n=3$) to the force amplitude without any specimen (system noise level from three trials) in the frequency range of 1–300 Hz (means ± standard deviations). The system noise level remains much lower than the signal level across all frequencies. (b) Comparisons of the sinusoidal force amplitude (system signal level) for the human vocal fold cover ($n=7$) to the system noise level (from three trials) in the frequency range of 1–300 Hz (means ± standard deviations). The system noise level remains about one standard deviation lower than the signal level at frequencies of up to around 275 Hz.

three levels of mass, suggesting that the frequency response of the piezoelectric transducer was flat and steady, up to a frequency of 400 Hz (Fig. 8).

D. System noise level

The system noise level was measured as the piezoelectric transducer output force amplitude without the presence of a specimen, and was compared with the system signal level, which was the force amplitude for specific specimens, in order to determine the frequencies at which the system noise was at an acceptably low level. The system signal level was obtained with (1) three samples of the ANSI S2.21 standard polymer material and (2) seven human vocal fold cover specimens. Three trials of oscillatory shear deformation were performed without any specimen to determine the system noise at frequencies of 1–300 Hz. Figure 9(a) shows the force amplitude for the ANSI S2.21 standard material samples, in comparison with the system noise level as a function of frequency. The mean values and the standard deviations are shown for both, although the standard deviations of the noise trials were too small for the error bars to be visible (average SD=0.000 254 8 N as compared to an average noise of 0.001 849 N). The results showed that the system noise level was consistently lower than the signal level for the standard material, with the difference increasing with frequency. At phonatory frequencies (>100 Hz), the system noise was around two orders of magnitude below the signal level.

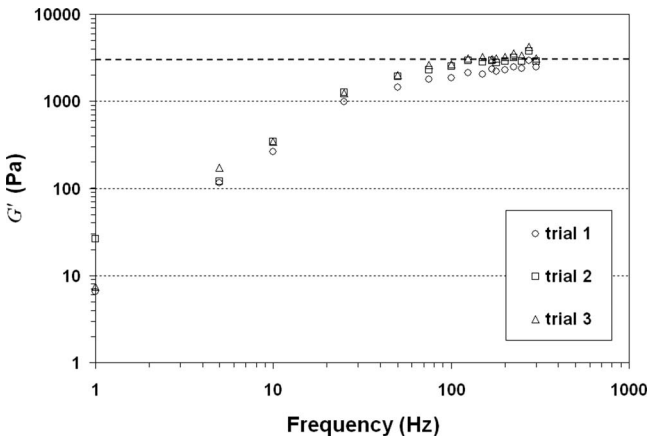


FIG. 10. Elastic shear modulus (G') of the three samples of the ANSI S2.21 standard polymer material as a function of frequency. The Dotted line represents the target elastic modulus of the standard material (3.0 kPa at 25 °C at around 170 Hz).

The system noise level was also consistently lower than the signal level for the average human vocal fold cover ($n=7$) across all frequencies [Fig. 9(b)]. The differences between the mean piezoelectric transducer output force amplitudes and the noise amplitudes were not as large as those in Fig. 9(a), but the signal level remained around one standard deviation higher than the system noise level over the entire frequency range, except at 300 Hz where the difference was less than one standard deviation. The signal level at or above 275 Hz was highly variable, with large standard deviations that introduce uncertainties into the validity of the data. These findings suggested that the frequency range within which valid viscoelastic measurements of the human vocal fold cover can be made was up to around 250 Hz, with data collected at or above 275 Hz considered not as meaningful due to the overlap between the signal and the noise of the system.

E. Validation with ANSI S2.21 standard material

Figure 10 shows the elastic shear modulus G' of three samples of the ANSI S2.21 standard polymer material, which is a viscoelastic standard for the calibration of equipment for viscoelastic characterization. The dependence of G' on frequency was found to be consistent with that of published standards (American National Standards Institute, 1998). The target elastic shear modulus of the material was 3.0 kPa at 25 °C at around 170 Hz (dotted line in Fig. 10), and G' obtained from the rheometer was found to be within 9% of the target modulus value. This indicated that the experimental error of measurements made with the rheometer was also likely within 9% of the measured values. This magnitude of

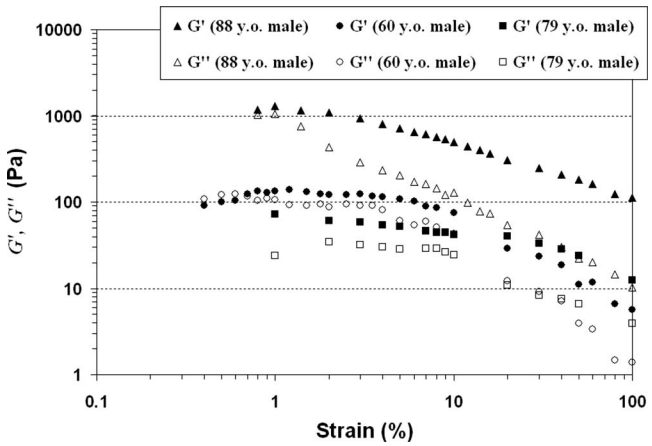


FIG. 11. Elastic shear modulus (G') and viscous shear modulus (G'') of three human vocal fold cover specimens as a function of shear strain (i.e., strain sweep; frequency=100 Hz). Based on the data of G', the small-strain linear region of viscoelasticity can be identified with the strain amplitude up to around 3%–7%.

error was similar to that of torsional rheometers (5%–10%) in previous studies (Chan and Titze, 1999; Chan, 2004; Titze *et al.*, 2004).

F. Viscoelasticity of the human vocal fold cover

The results of the strain sweep experiments of the human vocal fold cover are shown in Fig. 11, where the elastic and viscous shear moduli (G' and G'') of three vocal fold cover specimens are plotted as a function of shear strain amplitude at a frequency of 100 Hz. For a linear stress-strain relationship, the elastic modulus of the specimen would be independent of the displacement amplitude and the force amplitude (x_0 and F_0) (Chan and Titze, 1999). It can be seen in Fig. 11 that G' remained nearly constant in a small-strain, linear region of viscoelasticity, up to a strain amplitude of around 3%–7%, which varied slightly among the different specimens. This region of linear viscoelasticity was similar to that reported previously (5% in Chan and Titze, 1999). In order to ensure a linear, small-amplitude, simple-shear deformation, all frequency sweep experiments in the present study were performed at 1.0% strain.

Figure 12 shows the elastic shear modulus (G') and viscous shear modulus (G'') of the human vocal fold cover across the frequency range of 1–250 Hz, since this was identified as the functional range within which valid viscoelastic measurements can be made with the rheometer. The mean values of seven vocal fold cover specimens are shown, together with the standard deviations displayed as error bars (only upper error bars are shown for visual clarity on the plot). The average dynamic viscosity (η') and damping ratio (ζ) of the specimens over the same frequency range are shown in Figs. 13 and 14, respectively (once again with standard deviations as upper error bars). To characterize the empirical data with a parametric model (i.e., curve fitting), the dependence of G', G'', and η' on frequency f can be described by a power law as follows:

$$G' = pf^q \quad \text{or} \quad \log G' = \log p + q \log f, \tag{25}$$

$$G'' = pf^q \quad \text{or} \quad \log G'' = \log p + q \log f, \tag{26}$$

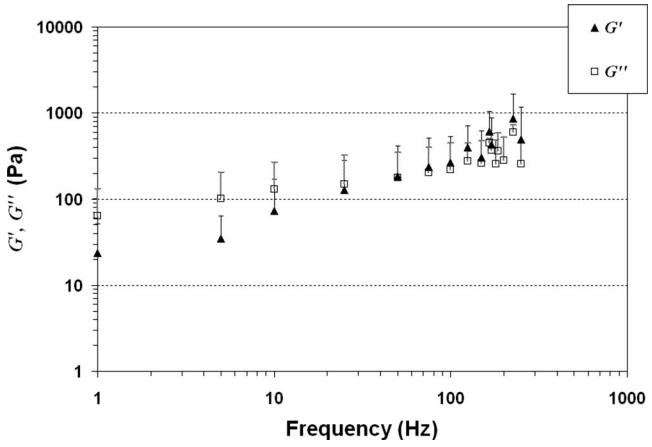


FIG. 12. Elastic shear modulus (G') and viscous shear modulus (G'') of the human vocal fold cover as a function of frequency. The means and standard deviations (upper error bars) of the seven specimens are shown.

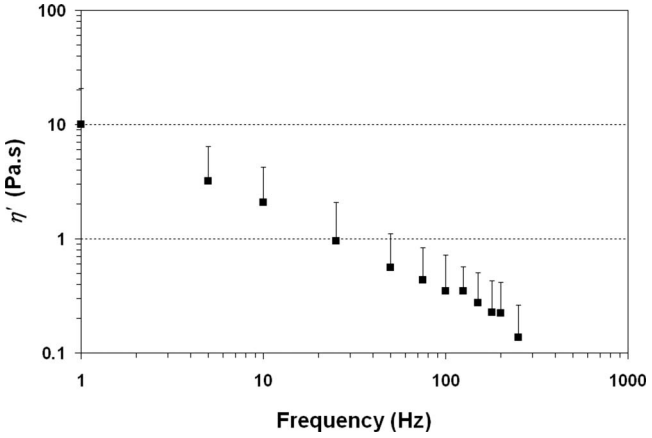


FIG. 13. Dynamic viscosity (η') of the human vocal fold cover as a function of frequency. The means and standard deviations (upper error bars) of the seven specimens are shown.

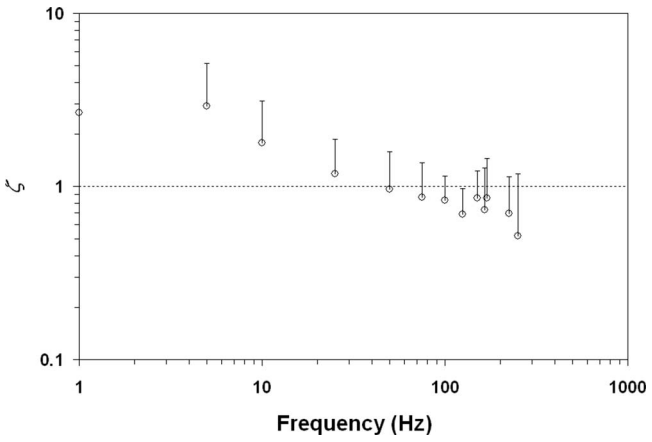


FIG. 14. Damping ratio (ζ) of the human vocal fold cover as a function of frequency. The means and standard deviations (upper error bars) of the seven specimens are shown.

TABLE II. Results of least-squares regressions for the parametric description of elastic shear modulus (G'), viscous shear modulus (G''), and dynamic viscosity (η') of the human vocal fold cover according to Eqs. (25)–(27) ($n=7$). R^2 is the coefficient of determination indicating the goodness of curve fitting.

Viscoelastic function	p	q	R^2
G'	17.246 Pa.s	0.6320	0.9493
G''	58.448 Pa.s	0.3243	0.8363
η'	10.509 Pa.s^2	−0.7391	0.9934

$$\eta' = pf^q \quad \text{or} \quad \log \eta' = \log p + q \log f, \tag{27}$$

where p and q are the parameters (coefficients) of curve fitting. Some previous studies on the viscoelastic properties of vocal fold tissues and phonosurgical biomaterials have used the logarithmic equation ($G'=a \log f+b$) to describe the empirical relationship between G' and frequency (e.g., Klemuk and Titze, 2004). In the present study, however, the power law was found to be a superior model for parametrizing the current data, since it resulted in a much better fit for both G' and G''. Average data for the seven vocal fold cover specimens were fitted to these equations by least-squares regressions. Results of the curve fitting are summarized in Table II, with values of the parameters p and q given, as well as the coefficient of determination R^2, as an estimate of the goodness of fit (Chan and Titze, 1999). It can be seen that the empirical data of G', G'', and η' were well described by the power law [Eqs. (25)–(27)], with the values of $R^2>0.83$.

The magnitudes of G' and G'' at 10 Hz, as shown in Fig. 12, were in a similar range as those in previous studies of the human vocal fold cover (e.g., Chan and Titze, 1999). However, a striking finding from Fig. 12 was that unlike previous studies (Chan, 2004; Chan and Titze, 1999), where the elastic shear modulus G' was consistently higher than the viscous shear modulus G'' at *all* frequencies of oscillation, G' was observed to be lower than G'' at low frequencies, until a "cross over" point at which G' and G'' seemed to converge and then crossed each other, at around 50 Hz. G' did not become higher than G'' until it reached the frequency range above this crossover point. This is a very interesting finding, because such crossover behavior of the viscoelastic moduli has never been observed in previous studies of the human vocal fold cover (Chan and Titze, 1999). In rheology, such viscoelastic behavior is often observed for polymer melts and polymer solutions (Barnes *et al.*, 1989; Ferry, 1980). In the frequency range below the crossover point, commonly called the terminal region, G' was lower than G'', with the specimens demonstrating a liquidlike linear viscoelastic response dominated by their viscous properties. In the frequency range above the crossover point, commonly called the rubbery region or plateau region, G' was higher than G'', indicating predominantly elastic properties in the specimens, characteristic of viscoelastic solids. For polymers, the transition of viscoelastic response between the terminal region and the rubbery region depends on the molecular weight, molecular weight distribution, and the extent of entanglement of the macromolecules (Barnes *et al.*, 1989; Ferry, 1980). For the vocal fold lamina propria, it is likely that such factors also play a key role in determining the viscoelastic response, although it can be expected that the interactions among the fibrous and interstitial proteins in the extracellular matrix are much more complicated than those in synthetic polymers (Chan and Titze, 1999).

The dynamic viscosity η' of the seven specimens decreased monotonically with frequency, indicating the behavior of shear thinning as observed previously (Fig. 13). The magnitude of η' at 10 Hz was consistent with that in previous studies (Chan and Titze, 1999, 2000). Figure 14 shows that the damping ratio ζ of the seven specimens generally decreased with frequency, as opposed to being a relatively flat function in previous studies (Chan and Titze, 1999). An intriguing finding was that ζ was actually above 1.0 at low frequencies (<50 Hz), suggesting that the vocal fold cover was, on average, overdamped below 50 Hz due to the fact of G'' being higher than G' in this terminal region. On the other hand, at higher frequencies in the rubbery (plateau) region (at or above 50 Hz), the damping ratio was always below 1.0, meaning that the vocal fold cover remained underdamped such that oscillation can be readily sustained during phonation. However, the mean value of ζ in the phonatory range (100–250 Hz) was 0.7383 ± 0.1211, considerably higher than those reported before (0.1–0.5 at 10–15 Hz in Chan and Titze, 1999) as well as typical values used in computer models (around 0.1–0.2; Titze, 2006).

V. CONCLUSION

A controlled-strain, linear, simple-shear rheometer system was custom built based on the EnduraTEC ELF 3200 mechanical testing system. The rheometer was designed for direct experimental measurements of the linear viscoelastic shear properties of human vocal fold tissues at frequencies in the phonatory range (above 100 Hz). Vibration experiments were performed to examine the frequency response characteristics of key components of the system, including a LVDT displacement transducer and a piezoelectric force transducer. The system noise level was estimated as the piezoelectric transducer output force amplitude without the mounting of any specimens, and was compared to the signal level of a standard viscoelastic material (ANSI S2.21) and that of the human vocal fold cover. Our findings suggested that the rheometer is capable of valid and reliable measurements of the linear viscoelastic properties of the vocal fold lamina propria, including elastic shear modulus (G'), viscous shear modulus (G''), dynamic viscosity (η'), and damping ratio (ζ) at phonatory frequencies, up to around 250 Hz.

Despite the significance of these findings, the results of the current study should only be considered preliminary due to the small number of vocal fold lamina propria specimens examined. Further studies involving additional specimens are required to corroborate the present findings and conclusions. Viscoelastic shear properties of the vocal fold lamina propria should be examined in the phonatory frequency range with this rheometer, in order to explore important differences in tissue rheological properties due to age, gender, race, and pathology.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and Other Communication Disorders, NIH Grant No. R01 DC006101. The authors wish to thank Min Fu and Bokkyu Lee for their assistance in rheological data collection and data analysis. Special thanks are extended to Alan McMullen, Kirk Biegler, and Troy Nickel of the ElectroForce Systems Group, Bose Corporation for their contributions to the design and validation of the rheometer.

American National Standards Institute, Inc. (**1998**). "Method for preparation of a standard material for dynamic mechanical measurements," *ANSI S2.21-1998* (Acoustical Society of America, New York, NY).

Barnes, H. A., Hutton, J. F., and Walters, K. (**1989**). *An Introduction to Rheology* (Elsevier, Amsterdam, The Netherlands).

Chan, R. W. (**2001**). "Estimation of viscoelastic shear properties of vocal fold tissues based on time-temperature superposition," J. Acoust. Soc. Am. **110**, 1548–1561.

Chan, R. W. (**2004**). "Measurements of vocal fold tissue viscoelasticity: Approaching the male phonatory frequency range," J. Acoust. Soc. Am. **115**, 3161–3170.

Chan, R. W., and Titze, I. R. (**1999**). "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021.

Chan, R. W., and Titze, I. R. (**2000**). "Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling," J. Acoust. Soc. Am. **107**, 565–580.

Chan, R. W., and Titze, I. R. (**2003**). "Effect of postmortem changes and freezing on the viscoelastic properties of vocal fold tissues," Ann. Biomed. Eng. **31**, 482–491.

Chan, R. W., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.

Ferry, J. D. (**1980**). *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley, New York, NY).

Gray, S. D., Titze, I. R., Chan, R., and Hammond, T. H. (**1999**). "Vocal fold proteoglycans and their influence on biomechanics," Laryngoscope **109**, 845–854.

Klemuk, S. A., and Titze, I. R. (**2004**). "Viscoelastic properties of three vocal-fold injectable biomaterials at low audio frequencies," Laryngoscope **114**, 1597–1603.

Titze, I. R. (**2006**). *The Myoelastic-Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City, IA).

Titze, I. R., Klemuk, S. A., and Gray, S. (**2004**). "Methodology for rheological testing of engineered biomaterials at low audio frequencies," J. Acoust. Soc. Am. **115**, 392–401.

Zhang, K., Siegmund, T., and Chan, R. W. (**2007**). "A two-layer composite model of the vocal fold lamina propria for fundamental frequency regulation," J. Acoust. Soc. Am. **122**, 1090–1101.

Consonant confusions in white noise

Sandeep A. Phatak,[a] Andrew Lovitt, and Jont B. Allen
ECE Department and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Illinois 61801

(Received 27 September 2007; revised 24 March 2008; accepted 3 April 2008)

The classic [MN55] confusion matrix experiment (16 consonants, white noise masker) was repeated by using computerized procedures, similar to those of Phatak and Allen (2007). ["Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2316]. The consonant scores in white noise can be categorized in three sets: low-error set {/m/, /n/}, average-error set {/p/, /t/, /k/, /s/, /ʃ/, /d/, /g/, /z/, /ʒ/}, and high-error set {/f/, /θ/, /b/, /v/, /ð/}. The consonant confusions match those from MN55, except for the highly asymmetric voicing confusions of fricatives, biased in favor of voiced consonants. Masking noise cannot only reduce the recognition of a consonant, but also perceptually morph it into another consonant. There is a significant and systematic variability in the scores and confusion patterns of different utterances of the same consonant, which can be characterized as (a) *confusion heterogeneity*, where the competitors in the confusion groups of a consonant vary, and (b) *threshold variability*, where confusion threshold [i.e., signal-to-noise ratio (SNR) and score at which the confusion group is formed] varies. The average consonant error and errors for most of the individual consonants and consonant sets can be approximated as exponential functions of the articulation index (AI). An AI that is based on the peak-to-rms ratios of speech can explain the SNR differences across experiments.
 [DOI: 10.1121/1.2913251]

PACS number(s): 43.71.An, 43.71.Es, 43.66.Dc, 43.72.Dv [MSS] Pages: 1220–1233

I. INTRODUCTION

Masking experiments play a crucial role in understanding the perceptual features of elemental speech sounds. Masking one or more of these features, defined as *events*, leads to a perceptual confusion (Régnier and Allen, 2008). Events are different from, though related to, other commonly used categories of speech features such as articulatory features (place, manner, etc.) or acoustic features (spectrum, temporal modulations, etc.). These features, which are extracted from the signal by auditory system, form the basis for perception of different speech sounds. Events, and their acoustic correlates, can be identified by directly comparing the perceptual confusions with the corresponding masked speech stimuli, on an utterance by utterance basis (Régnier and Allen, 2008). Such comparisons require a quantitative analysis of both the perceptual confusions and the speech stimuli.

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). The classic Miller and Nicely (1955) [MN55] study, which used CMs for measuring consonant confusions for noise-masked and filtered speech, has inspired many subsequent noise-masking CM experiments (Wang and Bilger, 1973); Dubno and Levitt, 1981; Gordon-Salant, 1985; Grant and Walden, 1996; Sroka and Braida, 2005). Phatak and Allen (2007) [denoted here as PA07[1]] used confusion patterns and confusion thresholds, first defined by Allen (2005b), in a quantitative analysis of the CM. PA07 employed large numbers of talkers and listeners to take advantage of the large natural variability in speech production and perception. Many questions raised in PA07 remained open due to large dimensionality ($16C \times 4V \times 18$talkers $\times 10$listeners $\times 6$SNR) and relatively low CM row sums (N). For example, are the differences between PA07 and MN55, such as the asymmetric voicing confusions of PA07 solely due to different noise spectra or due to procedural differences? Are talker and listener variations in perceptual confusions systematic or random? Do these variations, if present, correlate with the variations in speech stimuli?

To answer these and other outstanding questions, a CM experiment was conducted by using procedures similar to PA07, but with a white noise masker, as in MN55. We will refer to this experiment as MN05. One of the main purposes MN05 was to to verify whether the results of the classic MN55 study can be reproduced, which can be considered a validation of the PA07 procedures. To achieve this, the procedures of MN05 were designed to match as close as possible to MN55 procedures by making the least possible changes to PA07 procedures. A three-way comparison among MN55, MN05, and PA07 will let us estimate the effect of the noise spectrum on consonant perception by ruling out the effects of procedural differences, such as the use of recorded speech stimuli, male talkers, digital filters, and computerized presentations. Table I lists the relevant details of the three experiments.

The MN05 data also allows analyses that were not possible with MN55 or PA07 data. For example, unlike MN55 the speech stimuli are now available in MN05 to compare with the consonant confusions in white noise (WN). Such

[a]Authors present address: Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, D.C. 20307

TABLE I. Experimental details for Miller and Nicely (1955) [MN55], Phatak and Allen (2007) [PA07], and the current experiment [MN05]. The details include number of consonants (C), number of vowels (V), number of talkers (T), noise spectrum, and the speech database used.

Experiment	C	V	T	Noise spectrum	Speech stimuli
MN55	16	1	5	White (WN)	Live talkers
PA07	16	4	18	Speech-weighted (SWN)	LDC2005S22
MN05	16	1	18	White (WN)	LDC2005S22

correlations are crucial in establishing the noise-robust acoustic correlates of the perceptual features of speech (Régnier, 2007). Furthermore, the MN55 data were pooled over listeners and talkers, which averages out the possible talker and listener variations that are important for finding noise-robust features. Phatak and Allen (2007) also showed that the articulation index (AI), based on the peak-to-rms ratios of the speech corpus, can be used to parametrize consonant errors. Such analysis can be tested for WN with the present experimental data, but not with the masking data of MN55 due to unavailability of the stimuli.

The long-term goal of our studies is to determine the noise-robust features of basic speech sounds. PA07 and MN05 are the first two experiments in a series of data-collection experiments intended toward achieving this goal. Identifying the perceptual features quantitatively requires comparing the perceptual data collected in these experiments with the corresponding stimuli [Régnier and Allen (2008)]. This paper presents ways to quantify the perceptual data, which is the first step toward such comparisons.

II. METHODS

The testing procedures from the study of PA07 were modified to optimally match the methods of MN55. The speech stimuli were CV syllables with the 16 MN55 consonants followed by vowel /ɑ/, from the LDC2005S22 corpus (Fousek *et al.*, 2004). The syllables used in this study were spoken in isolation by 18 talkers (ten males and eight female). All talkers were native speakers of U.S. English, but three talkers were bilingual and had a part of their upbringing outside the U.S./Canada. MN55 used only female subjects, with one serving as talker, while the other four served as listeners. Since no significant talker-gender differences were observed by PA07, both male and female talkers were used in MN05.

The CV tokens were normalized such that each talker had the same average rms level. Random WN was added to the speech at five different signal-to-noise ratios (SNRs), viz., −12, −6, 0, 6, and 12 dB. When a listener had consonant scores significantly above chance level at −12 dB, then those consonants were presented to that listener at −15 dB, and again at −18 and −21 dB SNRs, if required. Data indicate that all listeners reached −15 and −18 dB SNRs, but rarely reached −21 dB SNR. The SNR was set for each token by using VUSOFT, a software VUmeter (Lobdell and Allen, 2007). The peak value of the VUSOFT output was used to define the speech level for each CV syllable. The speech and noise were filtered to have a bandwidth of 200–6500 Hz to match that in the MN55 experiment. Additionally, the CVs were presented in the quiet condition (i.e., no noise masker) as a control condition.

The stimuli were diotically presented to listeners through headphones (Sennheiser HD280). The listener reported the heard sound by clicking the appropriate choice on a computer screen. Unlike PA07, MN05 did not involve vowel recognition. Therefore, the MATLAB graphic user interface used in PA07 was modified to provide only 16 consonant choices. Consistent with MN55, presentations were randomized over consonants, but not over talkers or SNR. Thus, 18 CVs spoken by the same talker were successively presented at a fixed SNR. The set of 18 CVs for each talker consisted of 16 possible CVs, plus two of those randomly chosen, to limit the possibility of guessing by listener. The talker and the SNR for the next block of trials were then randomly chosen.

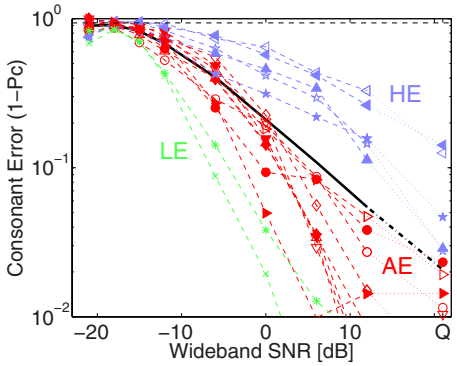
24 listeners (16 males and 8 females) having English as their primary language completed the experiment. The listeners were normal-hearing adults with no history of hearing problems. Three listeners had ages of 36, 45, and 50 years, while the remaining listeners were in the age group of 18–28 years (μ=21.57 yr, σ=2.32 yr). 21 listeners were born and broughtup in U.S. and self-reported to have a mid-western accent. The remaining three listeners had a part of their upbringing in India, South Korea, and China and reported to have South Asian, Southern U.S., and Chinese accents, respectively. However, no significant differences were observed in consonant scores and confusions of these three listeners and those of other listeners, and hence their responses were included. Each listener was trained for about 1 h in the quiet condition with visual feedback.

III. RESULTS

A. Listener and utterance selection

By following the analysis method of PA07, a post-hoc listener and utterance selection were carried out on the data of MN05. Listener selection is necessary to ensure that the listeners are attending to the task and that their scores are comparable to that of an average normal listener. The utterance selection is required to avoid misinterpreting errors due to mislabeled or mispronounced utterances as noise-induced confusions. The error thresholds used for this selection were same as those used in PA07.

23 of the 24 listeners had scores greater than 85% in the quiet condition and formed a homogeneous group (μ=92.4%, σ=3.5%). The responses of the one low-performance listener, who had 78.8% score in quiet, were removed from the dataset.



Consonant	Marker	Consonant	Marker
p	△	b	▲
t	◊	d	♦
k	▽	g	▼
f	☆	v	★
θ	◁	ð	◀
s	○	z	●
ʃ	▷	ʒ	▶
m	✶	n	×

FIG. 1. (Color online) The left panel shows consonant errors $P_e(\mathrm{SNR})=1-P_c(\mathrm{SNR})$ on a log scale, plotted against the wideband SNR in decibels. The solid line shows the average consonant error, while the colored dashed lines with marker symbols are for individual consonants. The three consonant sets: low error (LE), average error (AE), and high error (HE) are color coded. The legend on the right lists the markers used for consonants. The quiet condition is denoted by Q and is plotted at +21 dB. The horizontal dashed line at the top is the chance level error of $1-1/16=15/16$.

Based on the responses of the final set of 23 listeners, the syllable error for each utterance, which is same as the consonant error in this case, was estimated in the quiet condition. 32 of the total of 286 utterances had more than 20% error in quiet and were therefore considered as "ambiguous" utterances. Accordingly, the responses to these utterances were removed from the database. Following this utterance selection, the one low-performance listener had a score of 83.8%, while all other listeners formed a tight group with mean score of 97.9% and standard deviation of 1.2%. Thus, the listener categorization was verified to be unaffected by the utterance selection.

B. Consonant Errors

Figure 1 shows the consonant errors $P_e(\mathrm{SNR})=1-P_c(\mathrm{SNR})$ as a function of SNR. These curves can be categorized into three sets—a low-error (LE) set {/m/, /n/}, an average-error (AE) set {/p/, /t/, /k/, /s/, /ʃ/, /d/, /g/, /z/, /ʒ/}, and a high-error (HE) set {/f/, /θ/, /b/, /v/, /ð/}. These three sets are different from the three consonant sets observed in PA07, due to different noise spectra. The HE set C1 ={/f/, /θ/, /b/, /v/, /ð/, /m/} of PA07 differs from the HE set only by one addition consonant /m/. However, the other two sets are quite different in the two experiments. Such distinct sets were not observed in the $P_e(\mathrm{SNR})$ curves of MN55, except for the nasals /m/, /n/, which had the lowest errors in MN55, consistent with the LE set.

For comparing our scores with the original Miller–Nicely experiment, we find the SNRs required to achieve the same score in the two experiments. These SNRs, i.e., $\mathrm{SNR}_{\mathrm{MN55}}(P_e)$ and $\mathrm{SNR}_{\mathrm{MN05}}(P_e)$, obtained for a range of P_e values, are plotted against each other to obtain the isoperformance SNR contours in Fig. 2(a). For those P_e values which fall between the measures P_e values, the SNRs are estimated by linearly interpolating the $P_e(\mathrm{SNR})$ curves. The dashed curves with markers represent individual consonants, while the thick dash-dotted curve corresponds to the average performance. The thin dashed "reference" line with a slope of 45° corresponds to identical performance in the two experiments. A consonant curve above this dashed line implies that the higher SNR was required in MN05 (ordinate) than in MN55 (abscissa) to achieve the same performance for that consonant. In other words, the consonants that have curves above the reference line have poorer performance in MN05 than in MN55, at a given SNR. A curve below the reference line indicates a better performance in MN05 than in MN55. The proximity of the average performance curve (thick dash dotted) to the reference line in Fig. 2(a) implies that the average consonant performance in MN05 was almost equal to that in MN55. On average, LE and AE consonants performed better, while HE consonants performed slightly worse in MN05 as compared to MN55.

Figure 2(b) shows a similar comparison between MN05 (WN) and PA07 (SWN). There is a 10–12 dB uniform difference between the average scores (dash-dotted line). All consonants have poorer performance in WN relative to SWN, but the precise difference in the performance depends on consonant and varies with SNR. The consonants in set C2

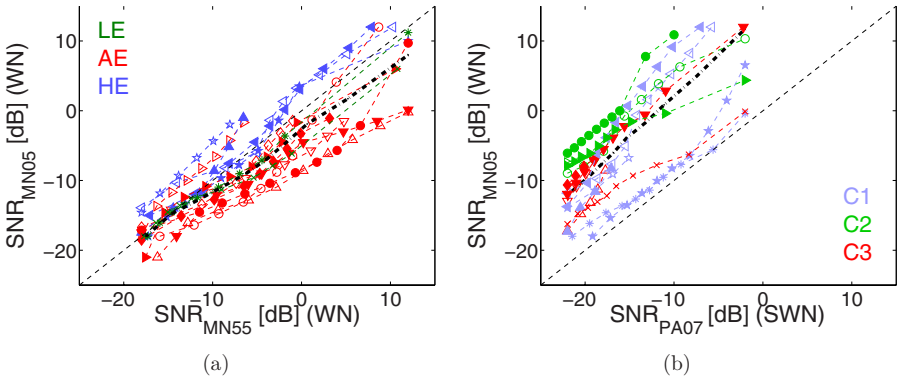


FIG. 2. (Color online) The isoperformance SNR contours for comparing MN05 scores with (a) MN55, and (b) PA07. In both panels, SNRs from MN05 form the ordinate. In (a), the individual consonant contours are color coded according to the three consonant sets LE, AE, and HE. In (b), the color scheme follows the three sets from PA07, i.e., the high-error set C1, the average-error set C3, and the low-error set C2.

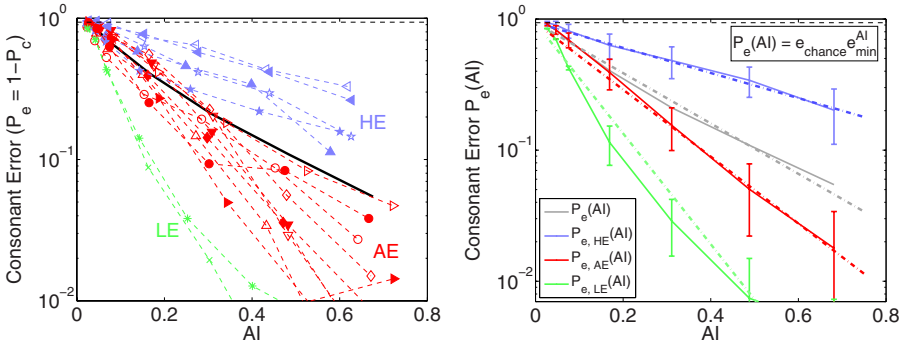


FIG. 3. (Color online) (a) The consonant errors $P_e(\mathrm{AI}) = 1 - P_c(\mathrm{AI})$ on a log scale plotted as a function of AI. The AI values were calculated at each SNR, except for the quiet condition, where the exact SNR was not known. (b) The average consonant error (gray) and the average errors for LE, AE, and HE sets. The solid curves are the empirically measured errors, with errorbars indicating the standard deviations within the sets, while the dash-dotted lines are the predictions of the exponential AI model [Eq. (1)].

from PA07 (i.e., /s/, /ʃ/, /z/, /ʒ/, and /t/) have the largest decrease in performance in MN05 with respect to the PA07 performance. This is expected, because these consonants had highest scores in speech-weighted noise, due to high-frequency energy (PA07). The frequencies above 2 kHz have significantly higher masking in WN than in SWN, resulting in poorer scores for these high-frequency fricatives. On the other hand, consonants /v/, /m/ (both set C1), and /n/ (set C3) have the least decrease in performance, as most of the energy for these consonants is concentrated at low frequencies where the spectra of the two noises are not very different.

1. AI

Allen (2005b) showed that the consonant log errors [i.e., $P_e(\mathrm{AI})$ on log scale] for the MN55 data are linear functions of the AI, following the exponential model:

$$P_e(\mathrm{AI}) = e_{\mathrm{chance}} e^{\mathrm{AI}}_{\mathrm{min}}, \tag{1}$$

from Allen (1994), where e_{min} is minimum error (at AI=1) and e_{chance} is the chance performance error (at AI=0). In this case, $e_{\mathrm{chance}} = 1 - 1/16 = 15/16$. Figure 3(a) shows that the average consonant log error (thick solid line) in MN05 also linearly decreases with AI, in accordance with the model.

Equation (1), which is based on Fletcher's *band-independence* theory, was defined only for average speech [Fletcher and Galt, 1950; Allen (2005a)]. However, the linearity of individual consonant curves in Fig. 3(a) demonstrates that the model also works for individual consonants, as previously observed by Allen (2005b) and PA07. It follows that the log-error curves for consonants can be expressed as

$$P_e(\mathrm{AI}, C_i) \approx e_{\mathrm{chance}} e^{\mathrm{AI}_i}_{\mathrm{min}_i}, \tag{2}$$

where e_{min_i}, and AI_i are the e_{min} and the AI values, respectively, for consonant C_i.

The three consonant sets are more obvious on an AI scale than on a SNR scale (Fig. 3). Figure 3(b) shows that not only average consonant log error (gray) but also the log errors for sets HE and AE are also linear functions of AI. The curvature in the $P_e(\mathrm{AI})$ for set LE is due to only one of the two consonants in that set, viz., /m/. The $\log[P_e(\mathrm{AI})]$ curves for 13 out of the 16 consonants can be matched to straight lines with very LE. This shows that the exponential AI model can be extended beyond the average consonant score, to consonant groups, and even individual consonants.

The average consonant error is a Bayesian sum of individual consonant errors and therefore, according to Eq. (2), can be expressed as a sum of exponential functions of AI, with different bases (i.e., the e_{min_i} values).

$$P_e(\mathrm{AI}) = \sum_{i=1}^{16} P_e(\mathrm{AI}, C_i) \approx e_{\mathrm{chance}} \sum_{i=1}^{16} e^{\mathrm{AI}_i}_{\mathrm{min}_i}. \tag{3}$$

Combining the two expressions for $P_e(\mathrm{AI})$ from Eqs. (1) and (3) results in

$$e^{\mathrm{AI}}_{\mathrm{min}} \approx \sum_{i=1}^{16} e^{\mathrm{AI}_{C_i}}_{\mathrm{min},c_i}. \tag{4}$$

A sum of exponentials cannot be an exponential, unless the bases are equal, but in this case the approximation fits well.

Figure 4 shows the consonant scores for the three different experiments, on SNR, and AI scales. The $P_c(\mathrm{AI})$ curves

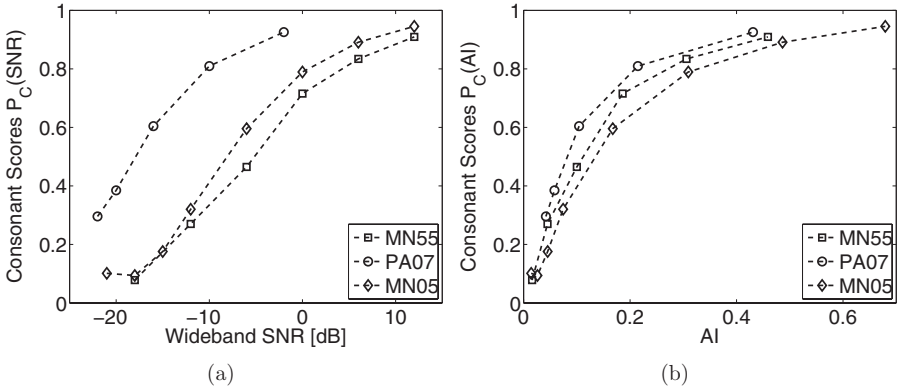


FIG. 4. A comparison of the consonant scores in the current study with those from [MN55] and PA07, plotted as a function of (a) SNR and (b) AI.

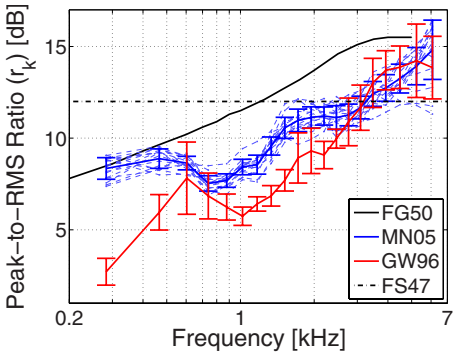


FIG. 5. (Color online) A comparison of the peak-to-rms ratios (r_k) used in different studies. The black solid curve (no errorbars) shows r_k values estimated by Fletcher and Galt (1950) [FG50]. The 16 dashed curves show r_k values for individual consonants, estimated from the CV tokens used in MN05. The means and standard deviations of r_k values for MN05 and Grant and Walden (1996) [GW96] stimuli are shown by the solid curves with error bars. The horizontal dash-dotted line shows the constant 12 dB peak-to-rms ratio used by French and Steinberg (1947) [FS47].

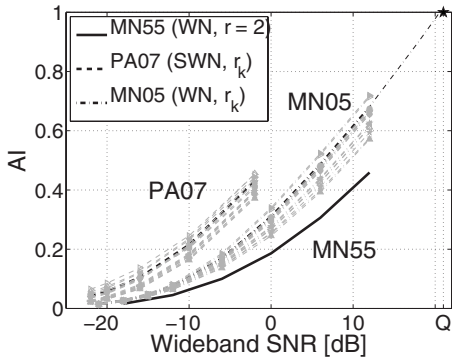


FIG. 6. The relationship between AI and wideband SNR for MN55 (solid line), PA07 (dashed), and MN05 (dash-dotted). The thin grayed-out lines are for individual consonants, while the thick black lines are for the average speech. The AI values for MN55 were estimated using $r=2$, while the frequency-dependent r_k values were estimated for average speech as well as for each consonant in case of PA07 and MN05. The present experiment curve, when extrapolated by using a third order polynomial fit, reaches AI $=1$ at about 21 dB SNR. The quiet condition (Q), which has AI$=1$ by definition, is plotted at this SNR.

are closer to each other than P_c(SNR) curves. This is because the differences in the noise spectra are accounted for by the AI, thus aligning the P_c(AI) curves for speech-weighted and WN. In spite of the same noise type, the AI values for MN55 and MN05 are not identical for a give SNR. The differences are due to the speech spectra and peak-to-rms ratios used in the AI calculation for the two experiments.

The AI values for MN05 were estimated by using the following PA07 formula.

$$\mathrm{AI} = \frac{1}{K}\sum_{k=1}^{K} \min\left[\frac{1}{3}\log_{10}(1 + r_k^2 \mathrm{snr}_k^2), 1\right], \tag{5}$$

where snr_k is the SNR and r_k is the peak-to-rms ratio (both in linear units, not in decibels) in the kth band, out of total $K=20$ articulation bands. The AI values for each consonant are estimated by using the average speech spectrum and average peak-to-rms ratios for that consonant. The details of calculating r_k values can be found in Appendix A of PA07. In this case, the peak-to-rms ratios varied from 2.19 ($\approx$6.89 dB) for /n/ in the 645–795 Hz articulation band to 6.65 ($\approx$16.45 dB) for /d/ in the 5720–7000 Hz band. Figure 5 compares peak-to-rms ratios (r_k) from the current study to those for the VCV syllables from Grant and Walden (1996) [GW96] study and with the r_k values reported by Fletcher and Galt (1950) [FG50]. The r_k values reported by FG50 were derived from the conversational speech data of Dunn and White (1940) and are frequently used as a standard for peak-to-rms correction in calculating AI (Pavlovic, 1984; Rankovic, 2002). The r_k values of MN05 and GW96 are lower than the FG50 values. The peak-to-rms ratios for GW96 are lower than those for MN05 below 3 kHz. This may be because GW96 stimuli (VCV) had two vowels per consonant, while MN05 stimuli (CV) had only one vowel per consonant. The steady and strong vowel formants, which dominate the envelope at these lower frequencies, significantly contribute to the rms value, but not so much to the peak value. The resultant would be lower peak-to-rms ratios for VCVs than for CVs. All three curves shows a significant variation in r_k over frequency, contrary to the claim by French and Steinberg (1947) that $r_k \approx 12$ dB, constant across frequency (horizontal dashed line). They made no claims regarding the variation in r_k across consonants. The dashed curves, which represent individual consonant, show that there is up to 5 dB variation in the r_k values over consonants, especially at higher frequencies.

Figure 6 shows a comparison of AI values, as a function of SNR, for the three experiments. The AI values for MN05 and PA07 were calculated using the spectra and peak-to-rms ratios that were directly estimated from the speech and noise stimuli. Since the speech and noise for MN55 was not available, the AI values for MN55 were calculated by using the straight-line approximation to the Dunn and White (1940) speech spectrum and a constant, frequency-independent peak-to-rms ratio of $r_k=2$ [Allen (2005b)]. At a given SNR, MN05 AI is higher that of the MN55 AI. This difference is predominantly due to the differences in peak-to-rms ratios, rather than the differences in the speech spectra. When the AI values for MN05 are calculated using a constant $r_k=1.7$, the average AI curve for MN05 coincides with the MN55 curve.

C. Consonant confusions

We use *confusion patterns* (CPs) [Allen (2005a)] to analyze the consonant confusions. A CP for a speech sound is obtained by plotting the row of that sound in CM against the SNR. Unlike the tabular form of CM, the formation of consonant groups over a range of SNRs can be directly observed in the CP. The confusion groups are not obvious in a CM table without a specific order of rows and columns, while the CPs do not depend on row and column orders. For example, consider the CP for consonant /t/ from the MN55 data shown in Fig. 7. Each curve corresponds to a particular column entry (h) for the /t/ row, plotted as a function of SNR, namely, $P_{h|/t/}$(SNR). The horizontal dashed line indicates chance, defined as the probability of guessing, which is 1/16. 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

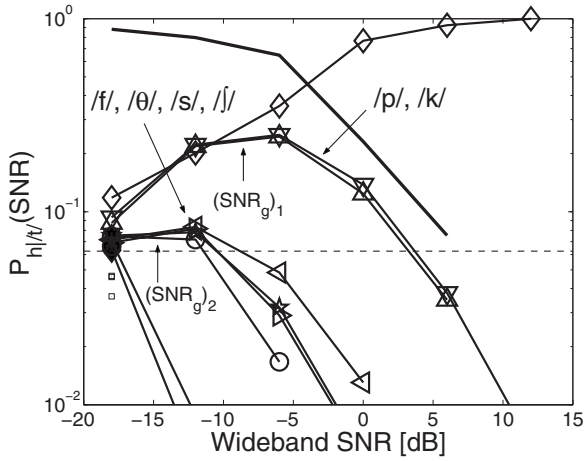
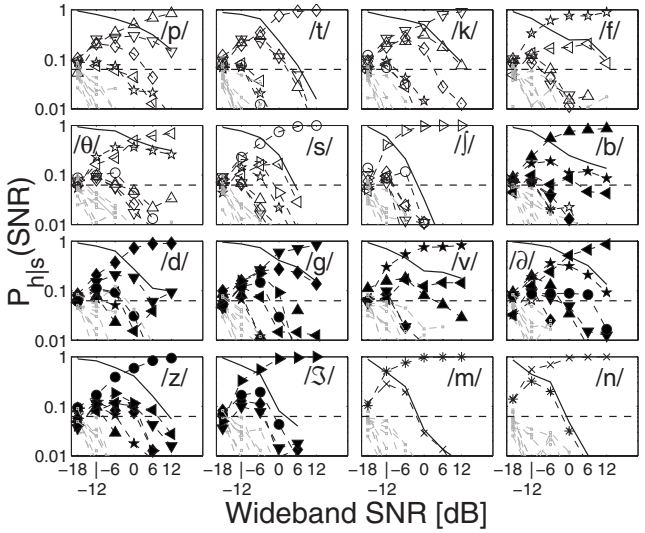


FIG. 7. Confusion patterns (CPs) for $s=/ta/$ from [MN55]. The thick solid line without markers is $1-P_{/b/|/t/}(\mathrm{SNR})$, which is the sum of off-diagonal entries. The horizontal dashed line shows the chance level of 1/16. Weak competitors, which do not exceed the chance performance, are shown by the gray square symbols.

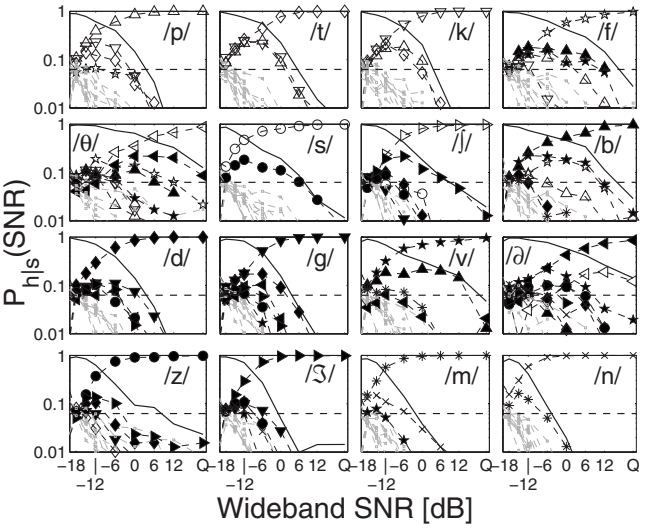
target for SNRs below −8 dB. We say that /t/, /p/, and /k/ form a *confusion group* (or *perceptual group*) at (or near) the *confusion threshold*, indicated by $(\mathrm{SNR}_g)_1 \approx -8$ dB, where $(\mathrm{SNR}_g)_1$ is the point of local maximum in $P_{/p/|/t/}(\mathrm{SNR})$ and $P_{/k/|/t/}(\mathrm{SNR})$ curves. When the SNR is decreased below $(\mathrm{SNR}_g)_2 \approx -15$ dB, consonant group [/f/, /θ/, /s/, and /ʃ/] merges with the [/t/, /p/, /k/] group, forming a supergroup. Since $(\mathrm{SNR}_g)_2 < (\mathrm{SNR}_g)_1$, consonants [/p/, /k/] are perceptually closer to /t/, and thereby form a stronger perceptual group with /t/ than the consonants [/f/, /θ/, /s/, and /ʃ/]. Thus we use the confusion threshold SNR_g as a quantitative measure to characterize the hierarchy in the perceptual confusions.

Figure 8(a) shows all 16 CPs for noise-masking data from MN55. Many confusion groups are not symmetrical. For example, the confusion of /θ/ with /f/ (☆ in second row, left panel) is significantly greater than confusion of /f/ with /θ/ (◁ in top right panel). Thus, the /f/-/θ/ confusion group is biased toward /f/. Allen (2005b) symmetrized the CMs, assuming these asymmetries to be insignificant. While the asymmetries for the /p/-/t/-/k/ and /m/-/n/ groups [the examples considered in Allen (2005b)] are negligible, for other confusion groups in MN55, such as the /f/-/θ/ group, these asymmetries are significant. These asymmetries are important for understanding the perceptual grouping of consonant under noisy conditions. Therefore, the CM should not be symmetrized.

Figure 8(b) shows the same 16 consonant CPs for MN05. These CPs are generated from the CM tables listed in the Appendix. The present experiment consonant confusions for plosives and nasals are very similar to those in MN55. The strong /p/-/t/-/k/, /d/-/g/-/z/ and /m/-/n/ confusion groups are common between the two experiments. However, the confusion thresholds in MN05 are at lower SNRs than those in MN55, indicating that the white noise has greater masking in MN55 than MN05. Part of this difference may be due to differences in the definition of SNR in the two experiments. To set the SNR in MN05, both speech and noise levels were



(a)



(b)

FIG. 8. The 16 consonant confusion patterns (CPs) for (a) the noise-masking data from [MN55] and (b) MN05. The horizontal dashed line shows the chance performance probability of 1/16. The weak competitors [i.e., $P_{h|s}(\mathrm{SNR}_g) < 1/16$] are grayed out for better visualization of the confusion groups. The quiet condition in MN05 (Q) is plotted at +21 dB SNR.

digitally measured by using a software VUMETER, whereas in MN55, the noise level (rms) was electrically measured and the speech level (peak) was measured by using a VUMETER instrument. Other possible factors, which cannot be tested with current data, could be the differences in speech stimuli (live talkers versus recorded) and familiarity of listeners with talkers in MN55.

A striking difference between the two experiments is observed in the fricative CPs. In MN55, consonants have negligible voicing errors, i.e., the unvoiced consonants have unvoiced competitors (hollow symbols) and the voiced consonants have voiced competitors (filled symbols). The unvoiced fricatives form /f/-/θ/ and /s/-/ʃ/ groups and their voiced counterparts have the corresponding /v/-/ð/ and /z/-/ʒ/ groups. These MN55 fricative groups are across place, but not across the voicing. In contrast, the fricatives in MN05 show significant voicing errors, but these voicing confusions

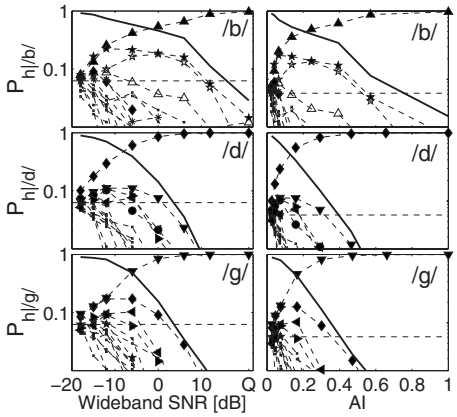


FIG. 9. The present experiment CPs for consonants /b/ (top), /d/ (center), and /g/ (bottom), as a function of SNR (left) and AI (right).

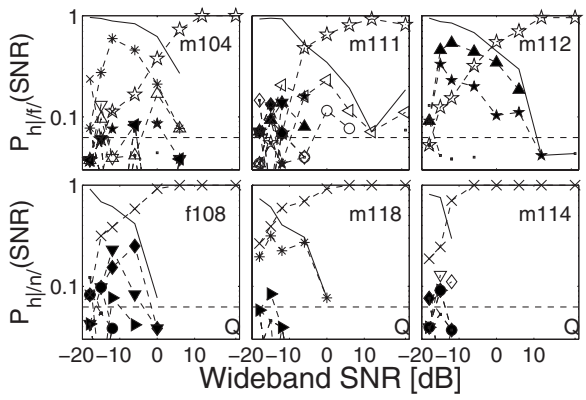


FIG. 10. Examples of confusion heterogeneity. The top row shows CPs for three utterances of /fa/: m104 (left), m111 (center), and m112 (right). The bottom row shows CPs for three utterances of /na/: f108 (left), m118 (center), and m114 (right). (m: male; f: female).

are biased in favor of voicing. That is, the strongest competitors for unvoiced consonants (hollow symbols) are voiced consonants (filled symbols), but not vice versa. For example, /s/ and /ʃ/ are only confused with /z/ and /ʒ/, respectively, but /z/ and /ʒ/ are hardly confused with any unvoiced consonant. The voiced fricative /v/ forms no confusion group with the unvoiced counterpart /f/ (in $P_{h|i/}$), but it is one of the strongest competitors of /f/ (in $P_{h|/f/}$). An interesting behavior is observed for consonant /ð/ in MN05. It is confused with /θ/, but only at SNR above 0 dB. At lower SNRs, it forms confusion groups with /v/ and /z/.

In MN55, consonants /p/ and /k/ form a stronger confusion group with each other than with /t/. Comparatively, the /p/-/t/-/k/ group is more symmetrical in MN05 and the three consonants equally compete with each other. In MN05, /p/ forms a weak group with /f/, but not with /θ/. Thus, MN05 data show /p/-/f/ and /f/-/θ/ groups, but do not show the /p/-/f/-/θ/ group from MN55. Similarly, /b/-/v/ and /v/-/ð/ groups are observed in MN05, but not the /b/-/v/-/ð/ group from MN55. On the other hand, some place confusions from MN05, such as /f/-/b/ and /v/-/m/, are not observed in MN55.

1. AI

As shown in Sec. I, the log-error curves P_e(SNR) for individual consonants become linear on an AI scale [i.e., $\log[P_e(\mathrm{AI})]=\mathrm{AI}\log(e_{\min})+\log(e_{\mathrm{chance}})$]. In this section, we investigate how the abscissa transformation from SNR to AI impacts the confusion patterns. Figure 9 shows the CPs for consonants /b/, /d/, and /g/, as a function of SNR (left) and AI (right). On the SNR scale, the consonant log errors (thick solid lines) have significant curvature. The slope of log-error curve changes as the number of competitors decreases with increasing SNR. The nonlinear SNR to AI transformation compresses the higher confusion regions into a small AI range. On AI scale, all confusion thresholds in the consonant CPs are compressed to $\mathrm{AI}\leqslant 0.2$. The remaining range of AI from 0.2 to 1 has only a small number of competitors, with linearly decreasing log confusions (i.e., log of the off-diagonal elements). The resultant is a linearly decreasing log error for the consonant.

D. Utterance variability

The availability of the confusion data for individual utterance allows us to analyze the utterance variability. In PA07, the utterance CPs were not analyzed because the row sums were too small to reliably analyze a 64×64 CM or even the 16×16 CM. In MN05, the number of listeners is more than twice the number of listeners in PA07, which gives relatively smoother CPs. The row sums for individual utterances in MN05 are slightly greater than the number of listeners (i.e., 23) because some utterances were presented more than once in a block to minimize listener guessing.

There is a significant variation in the recognition scores of different utterances of the same CV. Equally interesting and more complex are the variations observed in the distribution of confusions errors. We cannot directly attribute these variations to either the talker variability or to the within-talker utterance variability because only one utterance of a CV was available from each talker in the LDC database.

The variations in the utterance CPs can be broadly classified into two categories. First is the *confusion heterogeneity*, where the competitors in the confusion group vary from utterance to utterance. Second is *threshold variability*, where the confusion group remains the same, but the SNR and confusion probability at the confusion threshold are utterance dependent.

1. Confusion heterogeneity

The top row of Fig. 10 shows CPs for three different utterances of /fɑ/. In each case, /f/ is confused with different consonants. Talker m104's /f/ is confused mostly with /n/ and somewhat with /p/, while m111 /f/ is confused with /θ/ and /s/. Utterance m112 /f/ forms only one but strong confusion group with /b/ and /v/. The bottom row of Fig. 10 shows CPs for three utterances of /nɑ/. While utterance f108 /n/ forms confusion groups with /d/ and /g/, m118's /n/ is almost solely confused with /m/. Talker m114's /n/ is very robust with no errors at $\mathrm{SNR}\geqslant-6$ dB, and below that SNR it has no strong competitor, but many weak competitors that never exceed a 15% confusion probability.

Morphing. When a confusion significantly exceeds the recognition of the presented sound, such as that top left

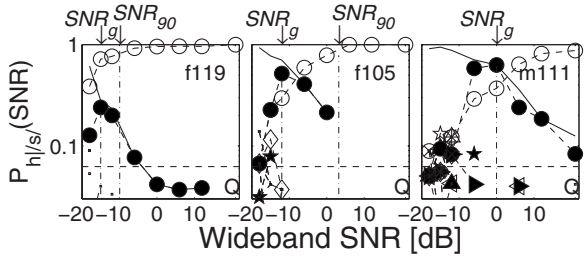


FIG. 11. Examples of threshold variability. CPs for three utterances of /sa/ by talkers f119 (left), f105 (center), and m111 (right). SNR_g denotes the /s/-/z/ group confusion threshold. SNR_{90} denotes the saturation point for /s/ recognition, where the diagonal score is 90%.

(m104 /fɑ/) and top right panels (m112 /fɑ/) of Fig. 10, we denote it as *morphing*. The target sound may be morphed into one or more other sounds by the noise masker. For example, while m104 /f/ is morphed to /m/, talker m112's /f/ has a double morphing toward /b/ and /v/. Not all utterances of a given consonant show morphing. Therefore, morphing is not observed in the average consonant CPs, shown in Fig. 8, which are obtained by pooling the data over utterances. Informal experiments show that at the crossover SNR, where the target and the competitor have equal scores, a listener can *prime* between two sounds.[2] If the target sound is presented in a meaningful word or sentence at the same SNR, then the priming would be resolved by context.

2. Threshold variability

On average, consonant /s/ is exclusively confused with /z/ in MN05 [see Fig. 8(b), second row, second column], with an average confusion threshold $SNR_g=-12$ dB and a confusion probability $P_{/z/|/s/}(SNR_g)\approx 20\%$. However, individual utterance CPs of /sɑ/ show a significant variation in the location of confusion threshold (Fig. 11). For talker f119, the threshold is at −12 dB with $P_{/z/|/s/}(SNR_g)\approx 25\%$, while for m111, it is at 0 dB with $P_{/z/|/s/}(SNR_g)\approx 65\%$. The confusion threshold for talker f105's utterance is between the two. Thus for the same confusion group, the threshold can vary from a minor confusion to a morph, depending on the utterance.

Noise robustness. Talker m111's /sɑ/ has more confusions, compared to f119 (Fig. 11). This means that m111 /sɑ/ is less robust to noise than f119 /sɑ/. To quantify the "robustness" of an utterance, we define a *saturation point*, denoted by SNR_{90}. This point forms a "knee" in the recognition score of the utterance, i.e., below SNR_{90}, the score rapidly decreases, while above SNR_{90}, the recognition score saturates. We quantify the saturation point SNR_{90} as the SNR where $P_c(SNR_{90})=90\%$. If the score for an utterance is always less than 90%, then $SNR_{90}=\infty$ is assigned to it. Thus, a lower SNR_{90} indicates a greater robustness to noise. In Fig. 11, f119 /sɑ/ ($SNR_{90}=-3.55$ dB) is more noise robust than f105's utterance ($SNR_{90}=4.25$ dB), while m111 /sɑ/ ($SNR_{90}=\infty$) is the weakest of the three.

The noise robustness of a sound depends on the masking noise spectrum. A quantitative analysis of SNR_{90} further supports the observations drawn from Fig. 2(b), that the consonants are more robust to SWN (PA07 dataset) than WN (MN05 dataset). Out of 192 common utterances, 174 utterances have $SNR_{90}(WN)>SNR_{90}(SWN)$, 14 have $SNR_{90}(WN)<SNR_{90}(SWN)$ and four have the same SNR_{90} in both experiments. In WN data of MN05, 17 utterances have $SNR_{90}=\infty$, compared to only eight in the SWN data of PA07. Four of these utterances (three /θ/ and one /ð/) have $SNR_{90}=\infty$ in both experiments.

IV. DISCUSSION

We have repeated the MN55 experiment using computerized techniques and a digitally recorded database. With few notable exceptions, the average consonant scores [Fig. 2(a)] and the consonant confusion patterns (Fig. 8) of MN05 closely match with those from the original [MN55]. This verifies that these "modern" computer based testing procedures can reliably reproduce the classic CM experiments. It also implies that the differences observed between PA07 and MN55 are due to differences in speech materials and noise spectra and not due to the procedural factors such as use of computers, digital filters, and a prerecorded database.

The consonant confusions of plosives and nasals in MN05 are virtually identical to those from MN55. However, the significant voicing confusions for fricatives, observed earlier in PA07 (SWN) were not present in MN55 (WN), but are present in MN05 (WN). Therefore, these confusions cannot be attributed to the differences in noise spectra between PA07 and MN55. These confusions were highly asymmetric, biased in favor of the voiced fricatives. Similar confusions are also observed in the Grant and Walden (1996) [GW96] acoustic-only data in SWN, and therefore cannot be attributed to our testing procedures and stimuli. These high voicing errors are responsible for the HE consonant sets, which contain fricatives /f/, /θ/, /v/, and /ð/, in the three experiments (PA07; MN05; and GW96), but not in MN55. One reason for low voicing errors by MN55 could be the familiarity of the listeners with the talker's voice. In MN55, the five listeners also served as the talkers, i.e., when one spoke the syllables, the other four listened and scored. There were no noticeable systematic differences in consonant scores, voicing scores, and consonant confusions for male and female talker utterances in MN05. Therefore, it is unlikely that the differences in MN05 and MN55 are due to the use of male talkers in MN05.

The isoperformance SNR contours (Fig. 2) are particularly useful when comparing performance across two different noise types. A comparison of WN (present experiment) and SWN [PA07] data shows that the difference in the noise spectra induces a constant SNR-loss of about 10–12 dB in WN, with respect to SWN [Fig. 2(b)]. The noise spectrum also impacts the distribution of consonant errors, resulting in different consonant sets in the two experiments. This further supports the conclusion of Phatak and Allen (2007) that consonants /s/, /ʃ/, /z/, /ʒ/, and /t/ have the greatest advantage in SWN. These consonants have lower scores in WN [present experiment; MN55]. The consonants /m/, /n/, and /v/ are almost equally masked by both types of noises. This is in agreement with the SNR-spectra analysis from PA07. The

three consonant sets observed in MN05 are the LE set {/m/, /n/}, the AE set {/p/, /t/, /k/, /s/, /ʃ/, /d/, /g/, /z/, /ʒ/}, and the HE set {/f/, /θ/, /b/, /v/, /ð/}.

The average recognition error in MN05 obeys the exponential AI model of speech recognition, given by Eq. (2). This model was introduced by Fletcher, and was shown to fit the average scores of isolate syllables [CV, VC, and CVC] [Fletcher and Galt (1950)]. Allen (1994) first expressed this relationship in terms of the minimum error $e_{\min}$ (i.e., the error at AI=1) and showed that the model can be extended to the individual consonant errors of the data of MN55 [Allen (2005b)]. The exponential AI model fits the error for individual consonants as well as for the three consonant sets in MN05 (Fig. 3). The log-error curves are linear in AI, relative to SNR, which can be partially explained by the CPs (Fig. 9). When plotted as a function of AI, the confusions in the consonant CPs are restricted to AI$<$0.2. As a result, the log of consonant error becomes more linear on AI scale than on SNR scale. As previously observed by Allen (2005b), the SNR-to-AI transformation also makes the log confusions [i.e., $\log(P_{h|s})$, the off-diagonal entries] more linear. Thus, it is possible to model the entire CM, not just the AE, in terms of AI.

Unlike the popular sigmoidal or ogive approximations of the performance-intensity curve P_C(SNR), the AI-model parametrization of the recognition performance has a solid theoretical psychoacoustic basis. Several standards for measuring speech quality, such the speech intelligibility index (SII) (ANSI-S3.5-1997, 1997) and the speech transmission index (Steeneken and Houtgast, 1980), are based on French and Steinberg's method of estimating AI. Allen (2005b) refined the original expression of French and Steinberg (1947) for estimating AI, to formulate a threshold correction to the AI, and showed that the AI is similar to the Shannon (1948) formula for channel capacity of a communication channel (Allen, 2004). However, this refined expression had a free parameter, which was later demonstrated by PA07 to be equal to the frequency-dependent peak-to-rms ratio of speech. The resulting AI expression [Eq. (5)] is explicitly computable from the speech and noise stimuli, and thus is completely independent of free parameters. This AI is equivalent to the *loudness, audibility, speech recognition* model of Studebaker *et al.* (1994), which is estimated from the peak spectrum of speech and rms spectrum of noise. As a result, the consonant recognition scores across experiments match better on the AI scale than on the SNR scale (Fig. 4). This is because the AI accounts for relative spectral shapes of speech and noise spectra, which are ignored in the wideband SNR calculation.

The CPs for individual utterances (Figs. 10 and 11) are not as smooth as the consonant CPs [Fig. 8(b)] due to lower row sums (N). After the utterance and listener selection, row sums for the consonant CPs range from 162 to 485 responses. In comparison, the typical N for utterance CPs is equal to number of listeners (i.e., 23) because each utterance was presented only once at each SNR to each listener. At low SNRs, when there are multiple competitors, the probability distribution in a row is multimodal. With a low N, the estimation error is relatively high, and thus multimodal distributions cannot be accurately estimated. However, at high SNRs, when there are a small number of competitors, the curves become relatively smooth. Thus, in spite of the low N, the confusion groups and the utterance variability analysis reveal useful results. All published CM data are pooled over listeners and talkers to reduce variance and to obtain an "average response." However, the variations across speech utterances and listener responses provide rich information, such as the morphing phenomenon, which is obscured by such averaging.

The utterance variability could not be analyzed either with MN05 data due to lack of stimuli or with PA07 data due to very low row sums. Thus, one of the aims for conducting MN05 was to investigate whether the utterance variations are random or systematic. The analysis of utterance confusion patterns shows that these variations are not only systematic but also can be quantitatively characterized into two types—confusion heterogeneity and threshold variability.

Morphed utterances provide a unique opportunity to better understand speech perception. The morphing phenomenon is also observed in a time-truncation experiment, where the CV syllables are gated from the consonant side (Régnier and Allen, 2008). For example, when a /sɑ/ utterance is truncated from consonant side, it first morphs into a /zɑ/. When truncated further, it first morphs to /dɑ/ and then to /ðɑ/, until only vowel is perceived. The truncation time at which an /s/ morphs to /z/ is consistent with the voice-onset time of a natural /z/. When a natural /zɑ/ utterance is truncated, it also morphs first to /dɑ/ and then to /ðɑ/, but it never morphs to /sɑ/. This is consistent with the asymmetric /s/-/z/ confusions observed in WN (present experiment) and SWN (PA07) masking data. This asymmetry suggests that the set of perceptual features or events that define consonant /z/ is a subset of those which define /s/. Thus, when the additional features in /s/ are masked or truncated, it is confused with, and in many cases morphs to, the consonant /z/, but /z/ is never confused with /s/. Comparing the /s/ utterance morphed into /z/ with a natural /z/ utterance can reveal these additional features in /s/ which distinguish it from /z/.

Consonants /p/ and /t/ form another pair that show this asymmetric morphing. Ten out of the 12 /tɑ/ utterances tested in the time-truncation experiment morphed to /pɑ/, but none of the /pɑ/ utterances morphed to /tɑ/ (Régnier, 2007). The individual utterance CPs for WN (present experiment) and SWN (PA07) masking data also significantly show more /t/ to /p/ morphing than /p/ to /t/ morphing, in terms of both the number of morphed utterances and the probability of morphing [i.e., $P_{h|s}(\mathrm{SNR}_g)$]. This results in the average $P_{/p/|/t/}(\mathrm{SNR}_g)$ [$\diamond$ in top left panel of Fig. 8(b)] to be lower than the average $P_{/t/|/p/}(\mathrm{SNR}_g)$ [$\triangle$ top row, second panel from left]. Thus, the event set for /p/ is a subset of the event set for /t/. Régnier (2007) show, by using time-frequency modification experiments, that the high-frequency release burst for /t/ is the event which separated /t/ from /p/. This result is consistent with the prediction by Heil (2003) that the envelope onset cues are critically important for speech intelligibility. This prediction is based on the neural data, which provides a physiological basis to the peak-to-rms ratio-based AI. The peak-to-rms ratios account for these perceptually

important temporal variations in speech, thus giving a temporal perspective to the otherwise spectral AI metric. An AI that accounts for the speech peaks can predict the recognition scores better than the ANSI-S3.5-1969 (1969) standard AI (Rankovic, 1998), and has led to the recent SII standard ANSI-S3.5-1997 (1997). The speech peaks are also critically important in extending the AI to predict speech intelligibility in fluctuating noise (Rhebergen and Versfeld, 2005).

Some utterances of a given sound are more robust to noise than others (Fig. 11). A quantitative analysis of the noise robustness, using the saturation point SNR_{90}, revealed that consonants are more robust to speech-weighted noise than WN. Noise-robustness analysis, combined with a spectrotemporal analysis of the stimuli, can lead us to the perceptual coding of speech. For example, Régnier and Allen (2008) found that SNR_{90} of /t/ utterances are highly correlated with the intensity of the transient in the release burst. Such a quantitative correlation would not be possible without the quantitative measures of CPs (i.e., SNR_g and SNR_{90}). Régnier and Allen (2008) also found that the event (i.e., the across-frequency coincidence of energy onset) is invariant, and it is the acoustic correlate (i.e., the intensity of the onset transient) which is responsible for the threshold variability. Similarly, it is likely that the confusion heterogeneity (Fig. 10) is due to differences in the relative intensities of the acoustic correlates of invariant events.

An alternative explanation for the heterogeneity is listener bias. If a listener narrows down the heard sound to a subset of possible choices, but is not confidant about the answer, then the response may be determined by the listener's bias for a specific answer. Such listener biases would dominate the responses in noisy conditions, where weak utterances are not clearly perceptible. This hypothesis can be easily tested by analyzing the consistency of listener responses to these utterances at low SNRs. However, such an analysis is not possible with the current data, as each utterance was presented only once or twice to each listener, at a given SNR. We have collected such data on listener consistency and this analysis is in progress.

V. CONCLUSIONS

The most important conclusions of this study can be briefly summarized as follows.

(1) The results of the classic Miller and Nicely (1955) can be reliably reproduced by using a recorded speech database and modern computerized testing procedures. The differences in the consonant data of Phatak and Allen (2007) (speech-weighted noise) and MN55 (white noise) are primarily due to different noise spectra.
(2) A normal-hearing listener's perception of consonants is more robust to speech-weighted noise than white noise. The noise robustness of an utterance can be quantified by using a saturation point (Fig. 11). Consonants /s/, /ʃ/, /z/, /ʒ/, and /t/ have the most disadvantage in white noise compared to speech-weighted noise, while consonants /v/, /m/, and /n/ are least affected by this difference in noise spectrum [Fig. 2(b)].
(3) An AI calculated from the specific speech and noise stimuli, by using PA07 AI formula [Eq. (5)], was verified to satisfy the exponential AI model [Eq. (1)] for consonant errors in white noise. The model can be extended to individual consonants as well as the consonant groups (Fig. 3).
(4) Masking noise cannot only reduce the recognition of a consonant, but also perceptually morph it into another consonant.
(5) In the presence of masking noise, fricatives show highly asymmetric voicing confusions, biased in favor of voiced consonants. MN55 data are an exception.
(6) There is a significant and systematic variability in the scores and confusion patterns of different utterances of the same consonant, which can be characterized as (a) *confusion heterogeneity*, where the competitors in the confusion groups of a consonant vary (Fig. 10), and (b) *threshold variability*, where confusion threshold (i.e., SNR and score at which the confusion group is formed) varies (Fig. 11).

ACKNOWLEDGMENTS

We thank all members of the HSR group at Beckman Institute, UIUC, for their inputs. We thank the three anonymous reviewers and the Associate Editor for their constructive comments and encouragement. This research was partially supported by a University of Illinois grant. The data-collection expenses were covered through research funding provided by Etymotic Research and by Starkey Labs.

APPENDIX

Tables II–IX show the consonant CMs, pooled over utterances and listeners, after the utterance and listener selection. The "only noise" responses, listed in the last column labeled ϕ, were considered to be chance performance responses and were distributed uniformly over the remaining 16 columns. These CMs were row normalized to have unity row sums for plotting the confusion patterns in Fig. 8(b). The SNR of −21 dB was rarely presented to the listeners due to their low scores at −18 dB SNR, resulting in very low row sums ($8 \leqslant N \leqslant 22$) and high variability at −21 dB SNR. Therefore, data at −21 dB SNR were not used in plotting the CPs and the corresponding CM is not listed here.

Since there were 18 talkers of each CV and 23 listeners, the row sums should be 414. However, as described in Sec. II, two randomly chosen CVs were repeated in each block (i.e., same SNR, same talker) to limit guessing by the listener. These extra presentations make the row sums greater than 414. Consonants /f/, /θ/, /v/, /ð/, and /z/ have row sums lower 414. This is because these consonants occurred most frequently in the ambiguous utterances and their row sums decreased after removing responses to ambiguous utterances.

[1]The experiment was called UIUCs04 by Phatak and Allen (2007), as it was conducted at the University of Illinois at Urbana-Champaign (UIUC) in the summer of 2004.
[2]For a priming condition between choices A and B, the listener will answer "Yes" with 100% probability to both questions-"Do you hear A?" and "Do you hear B?"

TABLE II. Consonant CM table. Quiet condition.

Quiet	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	457	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
/t/	0	469	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/k/	0	3	476	0	0	0	0	0	0	2	0	0	0	0	0	0	0
/f/	0	1	0	386	5	0	0	0	0	0	1	4	0	0	0	0	0
/θ/	0	0	0	5	202	4	0	0	0	0	0	20	0	0	0	0	0
/s/	0	0	0	0	1	429	0	0	0	0	0	1	3	0	0	0	0
/ʃ/	0	0	0	0	0	3	459	0	0	0	0	0	0	6	0	0	0
/b/	5	0	0	1	0	0	0	405	0	0	6	0	0	0	0	0	0
/d/	0	0	0	0	0	0	0	0	463	4	0	0	0	0	0	0	0
/g/	0	0	2	0	0	0	0	0	0	462	0	0	0	0	0	0	0
/v/	0	0	0	2	2	0	0	5	0	0	367	8	1	0	0	0	0
/ð/	0	0	0	0	19	0	0	0	0	0	3	133	0	0	0	0	0
/z/	0	0	0	0	0	0	0	0	0	0	0	3	378	6	0	0	0
/ʒ/	0	0	0	0	0	0	4	0	0	0	0	0	2	412	0	0	0
/m/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	468	0	0
/n/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	463	0

TABLE III. Consonant CM table. SNR=12 dB.

12 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	463	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
/t/	4	460	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
/k/	0	0	474	0	0	0	0	0	0	2	0	0	0	0	0	0	0
/f/	5	0	0	341	3	0	0	27	0	0	22	0	0	1	0	0	0
/θ/	0	0	0	15	159	3	0	9	0	0	3	48	0	0	0	0	0
/s/	0	0	0	0	0	429	0	0	0	0	0	0	12	0	0	0	0
/ʃ/	0	0	0	0	0	0	445	0	0	0	0	0	0	22	0	0	0
/b/	4	0	0	20	0	0	0	369	0	0	23	0	0	0	0	0	0
/d/	0	0	0	0	0	0	0	0	472	2	0	0	0	0	0	0	0
/g/	0	0	2	0	0	0	0	0	2	467	0	0	0	0	0	0	0
/v/	0	0	0	2	0	0	0	57	0	20	332	1	0	0	0	0	0
/ð/	0	0	0	0	29	0	0	1	2	0	5	112	2	0	0	0	0
/z/	0	0	0	0	1	1	0	0	0	0	1	7	377	5	0	0	0
/ʒ/	0	0	0	0	0	0	2	0	0	1	0	0	3	413	0	0	0
/m/	0	0	0	0	0	0	0	0	0	0	1	0	0	3	465	2	0
/n/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	458	0

TABLE IV. Consonant CM table.SNR=6 dB.

6 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	454	6	3	4	2	0	0	1	0	0	0	0	0	0	0	0	0
/t/	9	455	11	1	3	0	0	1	0	0	0	2	0	0	0	0	0
/k/	3	6	466	1	0	0	1	0	1	2	0	0	0	0	0	0	0
/f/	15	1	0	282	5	2	0	50	0	1	37	4	0	0	2	0	1
/θ/	2	0	0	22	133	4	0	15	4	0	4	51	0	0	0	0	0
/s/	0	1	0	1	4	398	2	0	0	0	0	1	28	1	0	0	0
/ʃ/	0	0	0	0	1	1	430	0	0	0	0	0	0	37	0	0	0
/b/	13	0	1	54	2	0	0	272	0	0	66	5	0	0	0	0	0
/d/	0	0	0	0	1	0	0	0	459	11	0	3	0	2	0	0	0
/g/	0	0	2	0	0	0	0	0	13	447	0	1	0	0	0	0	0
/v/	0	0	0	1	0	0	0	76	0	1	299	3	0	0	3	0	0
/ð/	0	0	0	0	26	0	0	0	6	2	14	88	9	0	0	6	0
/z/	0	0	0	0	1	9	0	0	6	0	1	7	351	8	0	0	0
/ʒ/	0	0	0	0	0	0	1	0	0	2	0	0	1	401	0	0	0
/m/	0	0	0	0	0	0	0	0	0	0	1	0	0	3	465	6	0
/n/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	473	0

TABLE V. Consonant CM table. SNR=0 dB.

0 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	390	14	21	22	2	0	0	4	0	0	1	3	0	0	0	0	0
/t/	41	368	54	4	2	2	0	1	0	0	0	3	0	0	0	0	0
/k/	44	25	376	2	10	1	3	1	1	2	1	6	0	0	0	0	0
/f/	35	1	3	226	14	3	0	59	1	1	33	3	0	0	5	2	1
/θ/	3	5	4	33	84	11	0	27	8	4	7	52	0	0	0	0	0
/s/	0	0	0	4	14	364	2	1	0	0	7	8	50	1	0	0	0
/ʃ/	0	0	0	0	3	17	382	0	6	0	0	0	4	55	0	0	0
/b/	15	0	1	74	5	0	0	220	0	0	76	8	0	2	6	0	0
/d/	1	0	0	0	2	0	1	0	402	35	0	10	1	7	0	0	0
/g/	0	0	1	0	8	1	2	2	44	409	0	10	1	7	0	0	0
/v/	1	1	0	8	0	0	0	87	0	1	276	9	0	1	14	4	0
/ð/	0	2	0	1	16	0	0	2	17	4	25	69	15	7	0	4	0
/z/	0	0	0	0	1	12	0	0	6	3	0	5	350	9	0	0	0
/ʒ/	0	0	0	0	0	1	2	0	0	11	1	0	3	383	0	2	0
/m/	1	0	0	0	0	0	0	1	0	0	2	0	0	0	453	14	0
/n/	0	0	0	0	0	1	0	0	1	1	0	0	0	0	6	454	0

TABLE VI. Consonant CM table. SNR=−6 dB.

−6 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	287	46	83	24	7	0	1	10	0	3	6	1	0	1	3	5	0
/t/	114	208	101	19	7	0	0	1	2	2	2	3	0	1	5	1	0
/k/	86	59	252	18	9	3	1	5	2	10	5	7	0	1	3	6	0
/f/	43	10	6	146	20	4	1	66	2	5	62	16	2	1	11	2	0
/θ/	5	15	14	30	63	5	2	21	10	16	18	34	3	1	1	0	0
/s/	0	2	1	10	18	313	3	8	5	1	7	10	56	6	0	0	0
/ʃ/	0	1	0	0	12	21	268	1	26	5	0	7	16	99	0	2	0
/b/	24	4	4	67	16	0	0	170	8	1	88	14	3	2	5	1	1
/d/	0	0	0	0	11	7	6	3	281	51	4	35	21	38	1	5	0
/g/	0	3	7	4	14	6	5	6	82	246	8	49	17	28	0	2	0
/v/	13	2	5	14	8	2	0	68	3	3	229	18	1	1	29	2	0
/ð/	1	7	0	0	8	3	1	4	8	6	34	34	15	15	5	9	0
/z/	0	4	0	1	5	13	1	0	14	7	2	8	287	39	0	3	0
/ʒ/	1	1	1	2	2	0	14	0	16	30	2	15	16	301	1	11	0
/m/	3	3	3	2	3	0	0	6	0	0	8	5	0	0	394	32	0
/n/	0	0	1	0	1	0	1	1	8	3	0	0	0	2	24	436	1

TABLE VII. Consonant CM table. SNR=−12 dB.

−12 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	183	57	107	31	8	3	0	16	2	3	12	5	3	1	26	7	0
/t/	103	122	116	25	10	6	1	12	8	15	17	8	3	4	12	9	0
/k/	81	100	126	19	18	7	1	19	12	10	16	14	3	3	18	25	4
/f/	51	23	30	65	15	11	3	71	8	11	50	14	4	6	21	8	1
/θ/	23	19	26	45	18	7	3	21	11	13	22	14	5	2	3	4	1
/s/	5	11	14	23	15	212	15	5	10	5	11	12	82	16	5	5	3
/ʃ/	3	22	18	13	21	26	110	7	47	20	7	23	41	91	3	10	9
/b/	36	8	14	41	10	9	1	106	25	12	92	12	10	5	27	3	0
/d/	5	15	12	3	14	11	13	25	135	51	17	29	47	49	14	18	0
/g/	5	21	16	11	24	20	10	24	81	85	30	45	32	43	5	22	0
/v/	21	9	6	17	7	10	3	72	15	12	124	25	12	7	32	15	3
/ð/	8	8	6	10	5	2	0	11	13	9	22	20	15	11	10	13	1
/z/	4	12	5	5	8	21	11	10	41	24	16	24	145	53	4	6	1
/ʒ/	6	7	6	6	7	9	22	12	44	41	25	16	44	120	12	31	2
/m/	23	10	13	13	2	1	2	25	7	10	24	11	4	4	263	51	0
/n/	6	7	12	3	6	4	3	4	22	22	12	16	4	14	57	274	1

TABLE VIII. Consonant CM table. SNR=−15 dB.

−15 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	101	63	88	28	10	7	4	31	5	15	22	10	6	5	23	12	44
/t/	67	77	84	22	18	15	15	27	8	14	20	10	5	7	28	23	37
/k/	53	71	73	26	8	7	8	22	12	23	20	7	7	5	32	34	62
/f/	41	26	25	26	11	12	8	55	13	19	44	12	12	2	26	15	41
/θ/	19	19	26	24	12	3	4	26	12	16	24	9	8	6	7	8	23
/s/	10	20	16	19	10	132	19	16	10	7	16	13	58	9	8	8	74
/ʃ/	9	24	19	18	10	31	35	16	38	35	25	20	29	44	9	17	91
/b/	34	16	25	31	8	14	6	53	35	20	66	17	13	5	26	12	30
/d/	15	16	16	14	16	12	13	28	81	42	20	36	37	30	25	22	50
/g/	16	18	29	16	18	14	18	23	57	58	32	26	34	37	6	31	41
/v/	16	12	19	30	14	14	3	49	20	26	73	16	28	12	27	22	26
/ð/	8	12	8	7	0	3	0	12	7	8	17	6	12	10	14	12	19
/z/	3	16	7	12	7	26	12	15	35	23	28	11	57	48	12	19	61
/ʒ/	10	19	10	10	8	17	8	7	36	29	31	15	38	59	20	34	58
/m/	26	28	20	28	8	7	3	25	6	12	34	7	10	7	136	67	49
/n/	8	24	14	1	8	11	9	9	25	17	20	11	19	19	73	138	66

TABLE IX. Consonant CM table. SNR=−18 dB.

−18 dB	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	ϕ
/p/	37	25	35	12	3	6	7	22	11	13	11	6	13	5	17	28	216
/t/	28	26	39	13	4	15	11	14	9	15	18	9	14	11	15	28	198
/k/	27	31	28	16	6	18	12	20	14	16	12	6	4	2	24	30	207
/f/	18	21	12	13	6	5	14	18	17	15	20	9	12	6	21	24	157
/θ/	9	11	7	9	3	4	5	9	12	16	14	3	4	2	13	6	102
/s/	13	20	12	6	4	43	16	5	8	7	8	10	19	6	10	10	238
/ʃ/	12	16	17	9	8	13	8	16	17	22	9	10	6	17	13	13	273
/b/	22	21	18	15	11	10	3	22	18	18	22	6	7	13	17	16	175
/d/	13	24	12	11	9	15	9	25	35	19	17	12	22	17	17	16	198
/g/	8	29	14	7	9	10	16	12	26	31	16	10	13	12	14	27	210
/v/	13	21	15	13	4	11	6	29	12	21	27	3	8	8	21	8	171
/ð/	11	7	7	0	4	3	4	3	6	6	10	3	2	3	5	12	68
/z/	8	17	12	8	8	7	10	8	12	11	10	11	13	6	8	17	220
/ʒ/	12	11	10	2	6	12	12	13	18	18	17	6	15	16	13	16	202
/m/	18	16	10	12	3	10	5	16	11	10	18	7	11	5	58	47	215
/n/	8	19	14	3	6	11	8	10	14	16	12	8	10	11	29	60	241

Allen, J. B. (**1994**), "How Do Humans Process and Recognize Speech?" IEEE Trans. Speech Audio Process. **2**, 567–577.

Allen, J. B. (**2004**), "The Articulation Index is a Shannon channel capacity," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer Verlag, New York), Chap. Speech, pp. 314–320.

Allen, J. B. (**2005a**), in *Articulation and Intelligibility*, Synthesis Lectures in Speech and Audio Processing, edited by B. H. Juang (Morgan and Claypool, USA).

Allen, J. B. (**2005b**), "Consonant recognition and the articulation index," J. Acoust. Soc. Am. **117**, 2212–2223.

ANSI-S3.5-1969 (**1969**), "American National Standard methods for the calculation of the articulation index" (American National Standards Institute, New York).

ANSI-S3.5-1997 (**1997**), "American National Standard methods for calculation of the speech intelligibility index" (American National Standards Institute, Inc., New York).

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., 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 International Conference on Spoken Language Processing (ICSLP), pp. 2749–2752.

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.

Heil, P. (**2003**), "Coding of temporal onset envelope in the auditory system," Speech Commun. **41**, 123–134.

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.

Pavlovic, C. V. (**1984**), "Use of articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.

Phatak, S. A., and Allen, J. B. (**2007**), "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2316.

Rankovic, C. M. (**1998**), "Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **103**, 1043–1057.

Rankovic, C. M. (**2002**), "Articulation index predictions for hearing-

impaired listeners with and without cochlear dead regions," J. Acoust. Soc. Am. **111**, 2545–2548.
Régnier, M. (**2007**), "Perceptual features of some consonants studied in noise," Master's thesis, University of Illinois at Urbana-Champaign, Urbana, IL.
Régnier, M., and Allen, J. B. (**2008**), "A method to identify noise-robust perceptual features: application for consonant /t/," J. Acoust. Soc. Am. **123**(5), 2801-2814.
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.
Shannon, C. E. (**1948**), "The mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.
Sroka, J., and Braida, L. D. (**2005**), "Human and machine consonant recognition," Speech Commun. **45**, 401–423.
Steeneken, H. J. M., and Houtgast, T. (**1980**), "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
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.
Wang, M. D., and Bilger, R. C. (**1973**), "Consonant confusions in noise: a study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English

Alexander L. Francis[a)]
Department of Speech, Language and Hearing Sciences and Program in Linguistics, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, Indiana 47906

Natalya Kaganovich[b)]
Program in Linguistics, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, Indiana 47906

Courtney Driscoll-Huber
Department of Speech, Language and Hearing Sciences, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, Indiana 47906

(Received 16 May 2007; revised 17 March 2008; accepted 21 May 2008)

In English, voiced and voiceless syllable-initial stop consonants differ in both fundamental frequency at the onset of voicing (onset F0) and voice onset time (VOT). Although both correlates, alone, can cue the voicing contrast, listeners weight VOT more heavily when both are available. Such differential weighting may arise from differences in the perceptual distance between voicing categories along the VOT versus onset F0 dimensions, or it may arise from a bias to pay more attention to VOT than to onset F0. The present experiment examines listeners' use of these two cues when classifying stimuli in which perceptual distance was artificially equated along the two dimensions. Listeners were also trained to categorize stimuli based on one cue at the expense of another. Equating perceptual distance eliminated the expected bias toward VOT before training, but successfully learning to base decisions more on VOT and less on onset F0 was easier than vice versa. Perceptual distance along both dimensions increased for both groups after training, but only VOT-trained listeners showed a decrease in Garner interference. Results lend qualified support to an attentional model of phonetic learning in which learning involves strategic redeployment of selective attention across integral acoustic cues. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945161]

PACS number(s): 43.71.An, 43.71.Es, 43.71.Rt [PEI] Pages: 1234–1251

I. INTRODUCTION

The acoustic patterns of speech sounds are highly multidimensional, in the sense that multiple acoustic properties typically correlate with the production of a particular phonetic category. Most, if not all, of these correlates have the potential to function as perceptual cues to categorization under appropriate circumstances, but not all cues are weighted equally in a given contrast. There are at least two major reasons that listeners might prefer to make a particular phonetic judgment on the basis of one cue over another. On the one hand, the perceived difference between two phonetic categories might be greater along one contrastive dimension than the other. Alternatively, some cues may be privileged (for particular phonetic decisions) because of learned or innate biases in the way they are processed.

The multiplicity of cues to phonetic contrasts is well documented. For example, Lisker (1986) describes a wide variety of acoustic correlates that differ systematically between productions of intervocalic /p/ and /b/ in English. Most or all of these correlates have been shown to be sufficient to cue the perception of this contrast in syllable-initial position, even in the absence of other cues (Lisker, 1978), but we will focus on four that have been more intensively studied: Voice onset time (VOT; Abramson and Lisker, 1970), the fundamental frequency at the onset of voicing (onset F0; Haggard *et al.*, 1970; Haggard *et al.* 1981), the degree of delay in the onset of the first formant (F1 cutback or voiced transition duration; Stevens and Klatt, 1974) and the relative amplitude of any aspiration noise in the period between the burst release and the onset of voicing (Repp, 1979). Despite the multiplicity of sufficient cues to the English stop-consonant voicing contrast, when more than one of these cues are presented to listeners, a pattern of dominance appears that suggests that some correlates are better able to serve as cues (often called *primary cues*) than others (*secondary cues*), at least in specific phonetic contexts. For the purposes of this study, the most relevant observation is that VOT appears to dominate other cues to voicing of syllable-initial stop consonants in English (Raphael, 2005). In particular, a variety of studies have shown that, in this context, VOT is preferred over onset F0 (Abramson and Lisker, 1985; Gordon *et al.*, 1993; Lisker, 1978; Whalen *et al.* 1993; see Francis and Nusbaum, 2002 for discussion). However, although such patterns of relative dominance are generally agreed upon, there is little consensus regarding the

[a)]Author to whom correspondence should be addressed. Tel.: (765) 494-3815. Electronic mail: francisa@purdue.edu
[b)]Also at: Department of Speech, Language and Hearing Sciences, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, Indiana 47906.

psychological basis for such apparent prioritization of one acoustic cue over another.

One factor of note in this regard is that the results of group studies on this topic (including the present one) may obscure the presence of real individual differences in the relative weighting of these two cues. For example, Haggard *et al.* (1970) found that onset F0 "can be of some importance, but the wide differences in performance between subjects show that it is unimportant for some listeners" (p. 616). Similarly, Massaro and Cohen (1976, 1977) found a range of individual differences in reliance on onset F0 as compared to VOT and fricative duration in a series of studies on the perception of voicing in syllable-initial fricatives. Such differences in individual listeners' weighting of normally covarying acoustic cues are consistent with other studies showing similar differences even in the perception of nonspeech cues (e.g., Lutfi and Liu, 2007), and clearly invite further study. However, the observation of individual differences in weighting still does not address the question of what might motivate the prioritization of one cue over another and to what degree such weighting might be changed by experience.

A. Perceptual weighting

One possible reason for the relative dominance of one cue over another is that the perceptual distance between two categories may be different along two different dimensions of contrast. For example, the perceptual distance between two prototypical exemplars of English /b/ and /p/ is quite large according to VOT and may be somewhat smaller according to onset F0.[1] In this case, listeners would be expected to give more weight to VOT than to onset F0, if only because the VOT differences are more easily distinguished. On the other hand, it is also possible that one dimension might be intrinsically better at attracting listeners' attention to it than another, such that, when given a choice between the two dimensions, listeners prefer to make decisions on the basis of one rather than another, even when the two contrasts are equated in terms of perceptual distance in isolation. That is, some acoustic properties may be privileged, at least with respect to their use in distinguishing a given phonetic contrast.

There seem to be at least two or three possible explanations of how such an intrinsic bias might arise. On the one hand, biases might arise as a function of (possibly innate) biological mechanisms, for example, as a consequence of differences in the efficiency of neural systems for processing different kinds of features, e.g., differences in neural systems specialized for processing temporally versus spectrally defined properties, see Zatorre and Belin (2001). Alternatively, such biases might derive from auditory/acoustic interactions between features that result in one feature enhancing the perception of another (Diehl and Kluender, 1989; Kingston and Diehl, 1994) or the two features together contributing to a higher-order, combinatoric perceptual feature (Kingston *et al.*, 2008). Finally, such biases might be explicitly learned, developing through years of experience listening to a language in which linguistically salient differences are more frequently made on the basis of one feature rather than another (a pattern whose origins might itself ultimately have a socio-historical as well as or instead of a psychophysiological basis) (see Holt *et al.* 2001 for discussion). That these kinds of explanations need not be mutually exclusive is supported by recent evidence suggesting that listeners' native language experience affects the efficiency of neural encoding of pitch properties at the brainstem level (Xu *et al.*, 2006).

One of the most recent and thorough discussions of the idea that listeners may be predisposed to use certain acoustic properties rather than others in a categorization task was presented by Holt and Lotto (2006). They trained adult listeners to categorize unfamiliar nonspeech sounds that differed according to two orthogonal dimensions, the center frequency (CF) of the carrier sine wave and the frequency of a modulating sine wave. They found that listeners showed a consistent preference for the CF cue, even when the perceptual distances between the two categories were equal along the two dimensions. This suggests that there may be intrinsic biases favoring the ability to learn (and therefore use) certain acoustic dimensions rather than others (see also Lutfi and Liu, 2007), but it is not known whether this is the case for dimensions that are relevant to perceiving speech sounds.

If English speakers' preference for using VOT over onset F0 in determining a syllable-initial stop-consonant voicing contrast results from a privileged status for VOT, then we would expect VOT to be given more weight than onset F0 when perceiving a voicing contrast even when the perceptual distance between tokens is equalized along the onset F0 and VOT dimensions. Thus, the first goal of the present study is to determine whether VOT and onset F0 exhibit different weighting in a voicing decision when perceptual distance is not a factor. These two commonly studied acoustic correlates of the phonetic voicing contrast were chosen because of the extensive literature on the perception of these two features and because previous research strongly suggests that VOT is more heavily weighted than onset F0 for perceiving the English voicing contrast in syllable-initial stops, yet it is not known whether this pattern still obtains after equating the two distances perceptually.

B. Dimensional integrality

Another consequence of the multidimensionality of speech sounds is that many acoustically independent correlates covary consistently with one another in the speech signal. The covariance of onset F0 and VOT has been argued to arise from a variety of sources. Abramson (1977) and Lisker (1978) suggest that the two features share a common origin in the unfolding of the same laryngeal timing gesture, while Hombert (1978) links the two via aerodynamic demands (higher airflow following the release of voiceless stops leading to a greater onset F0 and longer VOT),[2] In contrast, others ascribe the covariance to perceptual factors. For example, Kingston and Diehl (1994) and Kingston *et al.* (2008) argue that the two cues contribute to the perception of an overarching property of low frequency energy continuing into the stop closure (near short VOT/low onset F0 consonants) or its absence (in long VOT/high onset F0 conso-

nants), while Holt *et al.* (2001) claim that the covariance is learned simply because the two cues are reliably associated in the ambient language (without specifying a basis for this association).

In all cases, however, we might expect covarying cues to be highly integral in the sense of Garner (1974). Listeners who are accustomed to hearing that two cues covary in a consistent manner might be expected to have difficulty ignoring irrelevant variability in one of the cues when making a decision based on the properties of the other, especially if the two cues are integrated into a distinct "intermediate perceptual property" (Kingston *et al.*, 2008). When perceptual distances along the two covarying dimensions are not equal, variability along the more distinctive dimension tends to interfere more with classification along the less distinctive one in a pattern of performance known as asymmetric integrality (see Garner, 1974, 1983; Melara and Mounts, 1994). Thus, in the case of the covarying cues of onset F0 and VOT, if the perceptual distance between long- and short-lag VOT categories is naturally greater than that between falling and rising onset F0 categories, then this would be sufficient to explain the primacy of VOT as a cue to voicing, but artificially equating the perceptual distances along both dimensions should result in a symmetrical pattern of interference.

On the other hand, if VOT is intrinsically more attention demanding than onset F0, then variability in VOT should interfere more with classification according to onset F0 than vice versa. Moreover, this dominance should be maintained even when the perceptual distances between stimuli are equated (that is, even when stimuli are selected such that their perceptual distance is equivalent along each of two dimensions tested in isolation), because trial-to-trial changes along a more attention-demanding dimension should attract attention more than those along a less demanding one (see Tong *et al.*, 2008, for a review of some such cases).

In support of the possibility that VOT may simply be a more attention-demanding dimension of contrast, Gordon *et al.* (1993) argue that VOT is a "stronger" phonetic feature than onset F0, in the sense that VOT is more closely linked to the phenomenal quality of voicing than is onset F0. They suggest that under ideal listening conditions onset F0 is more likely to be ignored as a cue to voicing if VOT is unambiguous than vice versa (cf. Abramson and Lisker, 1985). Moreover, Gordon *et al.* (1993) showed that the primacy of VOT over onset F0 as a cue to stop-consonant voicing was mitigated by attentional demands. Under conditions of high cognitive load, listeners showed a decreased reliance on VOT and a corresponding increase in the relative weight given to onset F0, suggesting that, all else being equal, the use of VOT as a cue to voicing attracts or demands greater attentional commitment than using onset F0. However, in the study of Gordon *et al.* (1993) no attempt was made to equate the perceptual distance along the two dimensions. Thus, the second goal of this study was to investigate the symmetry of dimensional interference between onset F0 and VOT when making a voicing decision after equating perceptual distances along both dimensions. In this case, any observation of asymmetric integrality, such that variability in VOT interferes more with classification according to onset F0 than vice versa, would support the hypothesis that VOT is an intrinsically more attention-demanding dimension of phonetic contrast.

C. Perceptual learning

If, in fact, VOT is a privileged dimension for voicing (as compared to onset F0), then listeners might be expected to be better at learning new categories distinguished in terms of VOT than ones distinguished according to onset F0. A variety of studies (e.g., Holt *et al.*, 2004; Pisoni *et al.* 1982) have shown that listeners are able to learn new VOT-based categories with relatively little training, while Francis and Nusbaum (2002) showed that a few hours of laboratory training with Korean speech stimuli were sufficient to induce English listeners to make use of onset F0. However, due to methodological differences it is difficult to compare results across studies. Thus, the third goal of the present study was to determine whether training to identify categories differing only along one of these two dimensions (VOT or onset F0) would have comparable effects, or whether there would be differences in the effects of training based on the dimension being learned.

D. Enhancement and inhibition

A final question concerned the mechanism or mechanisms by which training affected perception of the two dimensions. A few theories of general perceptual learning (Gibson, 1969; Goldstone, 1994; Nosofsky, 1986) have been applied to perceptual learning of speech, primarily to explain the results of first- and second-language learning (Francis and Nusbaum, 2002; Iverson *et al.*, 2003). According to such theories, category learning requires increasing the similarity of tokens within the same category (acquired similarity), while increasing the perceived differences between tokens in different categories (acquired distinctiveness) (see Liberman, 1957, for what is probably the first application of these terms in speech research, and Jusczyk, 1993, for a comprehensive model of first language acquisition that explicitly incorporates these concepts). Such changes are argued to result from changing the relative weighting of different dimensions: Dimensions that are good at distinguishing categories are given more weight (enhanced), while those that do not differentiate categories well are given less weight (inhibited). Existing research provides tentative support for the hypothesis that both enhancement and inhibition of specific dimensions of contrast may operate in perceptual learning of speech. For example, Francis *et al.* (2000) trained two groups of listeners to use one of two competing cues to syllable-initial stop-consonant place of articulation: The slope of the formant transitions or the spectrum of the burst release. While listeners in the formant-trained condition learned to give increased weight to the formant cue, results from those in the burst-trained group were more suggestive of their having learned to give less weight to the formant cue rather than more weight to the burst cue. However, because the perceptual distance between tokens was not equated across the two cues, we cannot tell whether training caused listeners to adjust the weight given to formant transitions because the

stimuli differed more along this dimension of contrast (formant transitions) or because formant transitions are a privileged cue compared to the spectrum of the burst release. Thus, the final goal of the present study was to provide additional data relevant to determining whether training-related changes in the relative weight given to a specific dimension result from inhibition of the uninformative dimension or enhancement of the more informative one.

E. Summary

In the present investigation listeners were trained to hear a familiar consonantal contrast (voiceless aspirated versus voiceless unaspirated stops, e.g., [p] and [b]) according to either onset F0 or VOT while ignoring variability in the other cue. We used acoustic differences that were within a single category (voiceless aspirated) with the goal of ensuring that our stimuli were located within a region of perceptual space that did not contain any already-known discontinuities in auditory sensitivity such as the well-known discontinuity around 20–30 ms along the VOT dimension (cf. Holt *et al.* 2004) or the probable discontinuity between falling and rising frequency transitions (Schouten, 1985).

We used a variety of training stimuli, incorporating aspects of "high variability" training which has been argued by some researchers to be more effective than other common types of laboratory training (see discussion by Iverson *et al.*, 2005), in an attempt to improve learning over what is often observed in short-term laboratory training studies. We included stimuli produced at a variety of places of articulation of the initial consonant, with a variety of vowels, and produced by two different talkers. However, because the pretest and post-test results we report here derive from stimuli that were identical to (some of) those used in training, we cannot make any strong assumptions about what listeners were actually learning because there is no possibility to measure generalization, e.g., to a novel talker, place of articulation, or vowel context.

We measured the perceptual distance between tokens differing according to these two dimensions both before and after training and compared it to the distribution of selective attention between the two dimensions at the same times. All measurements were made from listeners who exhibited a high degree of success in learning. Our focus is on the performance of these successful learners because we were interested in the effect of *successful learning* on the distribution of weight to acoustic cues. By focusing on learners who showed clear improvement in performance, we also increase the validity of any comparison between the effects of learning observed here and those observed in more natural learning tasks (Francis and Nusbaum, 2002) and in actual cases of native language acquisition (e.g., Iverson *et al.*, 2003). We expected that training would increase perceptual distances along the trained dimension while possibly also decreasing distance along the (task-irrelevant) untrained dimension. Corresponding to these changes, following the results of Melara and Mounts (1994), we expected to see an increase in Garner interference when classifying according to the untrained dimension, and a similar decrease in interference when classifying according to the trained dimension.

II. METHOD

A. Subjects

A total of 42 young adults between the ages of 18 and 36 were initially enrolled in this experiment. All of them were undergraduate or graduate students or staff of Purdue University, or residents of the surrounding community. All participants underwent a standard hearing screening [pure tone audiometry at octave intervals between 500 and 4000 Hz at 20 (500 Hz) or 25 dB HL] and a linguistic background questionnaire designed to identify individuals with strongly monolingual perceptual experience. No applicant was enrolled if they failed the hearing screening, had lived for more than two weeks in a non-English speaking environment, grew up speaking any language other than English, or had lived in a household where the predominant language was anything other than English.

Participants were initially randomly assigned to one of two training conditions, VOT training or onset F0 training. However, as the experiment progressed and it became apparent that the VOT training condition was easier than the F0 condition, more participants were assigned to the onset F0 training group to increase the probability of ending up with relatively balanced numbers of successful learners in both conditions. Of the 42 initial participants, 34 completed all phases of the experiment (producing analyzable data), and 24 of these showed evidence of some learning (improvement of at least five percentage points). In all, 16 of these learners (11 women, 5 men) showed evidence of progressing toward expert perception of the contrast on which they were trained, defined as improvement of at least five percentage points above pretest level as well as a final proportion correct of at least 0.70. There were nine such expert learners in the VOT-trained condition (six women, three men) and seven in the F0-trained condition (five women, two men) (see Sec. III B, below).

B. Design

The goal of this study was to investigate the relationship between changes in perceptual distance and the distribution of selective attention before and after successful training to make phonetic decisions based on one acoustic cue as opposed to another. Thus, in addition to the usual pretest-training-post-test structure commonly used in phonetic training studies (e.g., Francis *et al.*, 2000; Francis and Nusbaum, 2002; Guenther *et al.*, 1999; Guion and Pederson, 2007), three kinds of measures were needed, one to assess degree of learning (in order to identify successful learners), one to determine the distribution of selective attention, and one to evaluate perceptual distance. It was also important that this last measure be obtainable even on the pretest, when listeners were expected to be close to chance when using cues on which they had not been trained. To assess learning, the measure of proportion correct responses was used, calculated over the first and last sessions of training. For measuring the distribution of selective attention, a set of related tasks often referred to as a *Garner paradigm* (Garner, 1974) was used.

Finally, to measure perceptual distance, two quantities were obtained: Sensitivity (d') computed from a speeded target monitoring (STM) task and response time (RT) computed from the baseline component of the Garner paradigm (see below). Sensitivity in a STM task was used in addition to the Garner base line task (which was collected in the course of evaluating selective attention, see below) for two reasons. First, the validity of response-time measures may be less reliable when participants are close to chance, as there will be fewer correct responses on which to base average scores, but the stimuli used in this experiment necessarily sounded quite similar to listeners (prior to training) to increase the likelihood of observing training-related improvement, meaning that performance on the initial Garner task would likely be close to chance. Second, since the primary goal of this study was to compare changes in perceptual distance with changes in selective attention, it was thought desirable to obtain a measure of perceptual distance through methods independent of, though similar in task structure to, the methods used to measure selective attention.

A final aspect of the experimental design that may play a role in interpreting the results is the choice of response categories in the Garner paradigm. In a typical Garner paradigm, stimuli differ along dimensions that are consciously identifiable to listeners, e.g., pitch and loudness, or hue and brightness. In such cases, participants can be instructed to identify stimuli according to a value along either dimension (e.g., is the sound "loud or soft" or "high or low pitched"?). However, in the present case the dimensions are expressly not accessible to conscious processing (Allen *et al.*, 2000). In such cases, researchers frequently first train listeners on novel, arbitrarily labeled categories (e.g., "type 1" versus "type 2"), but this was not an option in the present experiment because one of our research questions involved the effects of training and therefore we did not want to train listeners on the stimuli before we could establish a baseline measure of their performance. Instead, listeners were asked to identify stimuli as belonging to one of two categories (e.g., "B" or "P") when the decision was made along the dimension they were (to be) trained on, or according to one of two alternative categories when the decision was made along the untrained (to be ignored) dimension. The identity of the alternative categories, stressed and unstressed, was chosen based on the correspondence between both VOT and onset F0 with stress in English: Stressed syllables typically exhibit both a higher overall F0 and longer VOT than unstressed syllables, and a sharply falling F0 contour is associated with emphatic stress (as in the final syllable of the response "You don't believe that story, do you?" "Yes, I *do*"). However, listeners were not necessarily expected to be as facile with this classification as with the voicing classification so it was used only for the untrained dimension.

C. Stimuli

Six sets of 100 stimuli varying in two dimensions (onset F0 and VOT) were generated from naturally recorded tokens using PSOLA resynthesis (PRAAT 4.2, Boersma and Weenink, 2006).

1. Recording

Initially, multiple productions of each of the nine syllables [p^hi], [p^ha], [p^hu], [t^hi], [t^ha], [t^hu], [k^hi], [k^ha], and [k^hu] were recorded by one adult male and one adult female native speaker of a Midwestern dialect of American English. Recordings were made to digital audio tape using a hypercardioid microphone (Audio-Technica D1000HE) and digital audio tape-recorder (Sony TCD-D8) in a sound-isolated booth (IAC, model No. 403A), and redigitized to disk for analysis and resynthesis at 22.05 kHz sampling rate and 16 bit quantization using PRAAT 4.2 via a SoundBlaster Live! Sound card on a Dell Optiplex running Windows XP. Speakers recorded multiple instances of three repetitions of each syllable. For example, two or three utterances of [p^ha p^ha p^ha] were recorded by each speaker. Only the second token of each group was digitized to maintain similar intonational properties across tokens. The resulting set of 54 tokens (three repetitions of each of nine syllables by two speakers) was carefully analyzed to identify the acoustically cleanest recording of each syllable. Tokens with a comparatively high degree of line noise or breathiness, irregularities in voicing during vowel production, or other acoustic artifacts that could be compounded by the resynthesis process were discarded. In the end, six tokens were selected for each speaker, creating two mostly overlapping sets (with the lack of complete overlap due to acoustic artifacts in specific recordings). For the female speaker, [p^hi], [p^hu], [t^hi], [t^hu], [k^hi], and [k^ha] were selected, and for the male talker [p^hi], [p^ha], [t^hi], [t^h], [k^hi], and [k^ha]. Stimuli derived from the male [p^ha] tokens were used for testing, and stimuli derived from all tokens (including the male [p^ha]) were used in training.

2. Resynthesis

Starting with each of the 12 base syllables, a set of 100 tokens were resynthesized using the PSOLA methods implemented in PRAAT 4.2, creating a grid varying in ten steps along each of two phonetically relevant acoustic dimensions, onset F0 and VOT, for a total of 1200 tokens (100 tokens for each of 12 starting syllables). Along the VOT dimension, stimuli ranged from 35 to 65 ms VOT in approximately 3 ms steps.[3] Variation in onset F0 ranged from a starting frequency of 1.21 times the starting frequency of the unmodified (base) syllable to 0.91 times (125 Hz for the male [pa]), in steps of about 4 Hz (i.e., for the male [pa] stimulus, the starting frequency ranged from 165 to 125 Hz). All onset F0 contours were linear interpolations starting at the defined initial value and decreasing to the original F0 contour over the first 100 ms of the token (ending at 118 Hz). Thus, all onset F0 contours ranged from sharply falling to nearly flat. There were no rising contours in any stimuli. Slopes ranged from −0.07 Hz/ms (in the shortest VOT, lowest slope stimulus) to −0.47 Hz/ms for the most sharply falling contour.

3. Nomenclature

The goal was to identify four stimuli that differed orthogonally according to two dimensions to perceptually equivalent degrees [forming a square in perceptual space, as

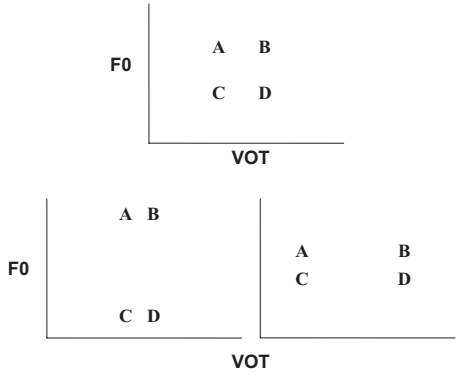


FIG. 1. Hypothetical illustration of changes in perceptual space from equally balanced performance on pretest (1a) to increased attention to VOT/decreased attention to F0 (1b) or decreased attention to VOT/increased attention to F0 (1c). Axes are measured in arbitrary units of perceptual distance.

shown in Fig. 1(a)]. For the purposes of testing and training, stimuli were identified differently to each group, based on the dimension on which each group was trained. For participants in the VOT-trained group, tokens A and C were both treated as exemplars of B while B and D were categorized as P. Conversely, A and B were both considered stressed while C and D were unstressed. In contrast, for participants in the F0-trained group, A and C were both considered unstressed and B and D were stressed, while A and B were labeled as P and C and D were labeled as B.

D. Procedure

Participants completed a total of 11 to 12 sessions, each about an hour in duration, over the course of three to four weeks (one session per day, usually with no more than three days between any two sessions).

The first three sessions and last three sessions constituted the pretest and post-test, respectively, with six sessions of training between them. In the first pretest session participants completed the hearing test, language background questionnaire, and initial assessment of perceptual distance to identify subject-specific, perceptually equal distances along the two dimensions. In the second and third pretest sessions, participants completed the tasks associated with the Garner selective attention paradigm using both male and female stimuli (one talker in each session). The post-test was accomplished in the reverse order of the pretest, but consisted of the same tests (Garner paradigm followed by perceptual distance measurement). When time permitted, the last two sessions of the post-test were conducted on the same day. Training was carried out in the intervening sessions.

1. Perceptual distance measurement (STM)

The goal of this stage of the pretest was to identify four tokens whose pairwise perceptual distances were approximately equal in each of the two dimensions, roughly forming a square in the VOT-by-onset-F0 space, as shown in Fig. 1(a). Sensitivity, d', was used as a measure of perceptual distance because, with listeners expected to be close to chance on the pretest, such a measure would be more informative than response time for correct responses, which might be highly variable due to a high incidence of guessing. Testing always proceeded in the same order. Starting with the B token (step 7 along both the VOT and onset-F0 dimensions, indicating a token close to but not quite prototypical for [p^h]), a corresponding A token was selected having the same onset-F0 value (step 7), but a more [b]-like (shorter) VOT (generally step 3 or 4). Participants then completed a series of eight repetitions of a speeded target monitoring task (STM, see below) using these two stimuli, and sensitivity (d') was calculated as the difference between the z-score transformed proportion of hits and false alarms [$z(H) - z(\mathrm{FA})$ (Macmillan and Creelman, 2004), where hits were counted as correct responses to presented targets, while false alarms were incorrect responses to distractors (nontargets). If the listener's sensitivity to the initial A-B pair was less than 1, a more distant candidate for the A token was selected (e.g., step 2 or 1) and the STM task was repeated. Conversely, if the listener's sensitivity to the initial A-B pair was greater than 1, a closer candidate for the A token (e.g., step 4 or 5) was selected and the STM task was repeated. This process was repeated until either (1) a VOT step value was identified that was approximately 1 d' distant from the B token along the VOT dimension or (2) the perceptual distance between the B token and the most distant possible A token (VOT step 0) was determined. At this point the A token was *fixed* and the selection of a D token began. If the most distant A token was selected (i.e., if the maximum distance between the B and A tokens was still less than 1 d'), then the d' value calculated between this A and the B token was used as the critical value (instead of 1) for the next leg of the square. A similar quasi-iterative process was used to select a D token located approximately the same distance away from the B token along the onset-F0 dimension (typically close to 1 d', but sometimes less if the step-0 A token was used). This process took between one and five repetitions for the AB distance (mean=2.2, SD=0.81) and between one and four repetitions for the BD distance (mean=2.2, SD=0.76). After A and D tokens had been identified through these iterative procedures, a C token was automatically selected having the onset-F0 step value of the D token and the VOT step value of the A token. Once all tokens were selected, the perceptual distances between the remaining adjacent pairs (DC and AC) as well as the diagonals (AD and BC) were computed using the same STM task (see Sec. III). In this way, a set of four tokens were selected that were approximately equidistant in perceptual space for each individual listener. Step values identified in this session were then used for all stimuli, both in testing and training. Note that, since the order of presentation of each pair was the same for all listeners, some effect of order of presentation may have occurred.

The task used to determine d' for a given pair of stimuli was STM. For every pair of tokens, listeners completed one set of eight trials with each trial consisting of a total of 20 stimulus presentations. In each trial, participants were shown a type of sound to monitor for (e.g., B or P for tokens differing only along the trained dimension or stressed or unstressed for tokens differing only along the untrained dimension). The stimulus corresponding to this identifier was considered a target for this trial, while the other stimulus was

considered the distractor. For example, if a member of the VOT-training group was being tested on the distance between the C and D tokens, in a trial specified as monitoring for B, the C token (more [b] like) would be the target while the D token (more [p] like) would be the distractor. If the trial involved monitoring for P then the D token would be the target and the C token would be the distractor. The category identifier (e.g., B) remained on the screen for the duration of the trial. Beginning 1 s after the target identifier appeared on the screen, listeners heard a series of 20 tokens, presented with 1250 ms stimulus onset asynchrony. There were an equal number of target and distractor tokens, and these could appear in any order within the trial with the constraint that a target token could not appear first or last in the trial. Participants were instructed to press a response key every time they heard a syllable starting with the sound shown on the screen and not to respond if the syllable began with a sound different from the symbol shown. They were asked to response as quickly as possible, but also to be as accurate as possible. Responses were scored as hits (responses to targets) or false alarms (responses to distractors) and combined over all eight trials (total of 80 target presentations and 80 Distractors) and used to calculate d'.

Before each trial, listeners were familiarized with the two tokens to be used, and their respective labels for the particular contrast being tested (e.g., for a participant in the VOT-trained group, the A versus B stimulus contrast would be presented as exemplars of B (paired with the A token) and P (paired with the B token). Familiarization consisted of presentation of a stimulus label (e.g., B) with instructions to click on the mouse button in order to hear an example (the A token). After one presentation, listeners were instructed to click the mouse again to hear the sound again. Then the task proceeded to the next stimulus/label pair. Thus, each stimulus was presented a total of 16 times with its associated label in a given block (twice per each of eight trials).

2. Garner paradigm

A complete Garner selective attention paradigm consists of three kinds of tasks, each using stimuli drawn from a set of four stimuli, arranged in a square in perceptual space. The tasks are typically referred to as *baseline, correlation*, and *orthogonal* or *filtering* (Garner, 1974; Pomerantz *et al.* 1989). Each task involves classifying two or four stimuli as exemplars of two categories, e.g., B or P. In this experiment participants completed two base line tasks, two correlation tasks, and one filtering task for each dimension of classification. Because our focus is on Garner interference, only results from the baseline and filtering tasks will be discussed in detail, although responses to some of the stimuli in the correlated condition (specifically, the A and D tokens) are informative with respect to the question of the relative weighting of the two cues in a directly conflicting condition analogous to that used by Francis *et al.* (2000). Moreover, although both male and female voices were used, only results for the male stimuli will be discussed because performance was noticeably better for this talker, especially among the F0-trained listeners. Tasks were blocked by talker (in different sessions) and by dimension: All tasks involving classification by the trained dimension were grouped together, as were all involving classification according to the untrained dimension. Furthermore, the order of labels on the screen (e.g., B and P) and their associated response keys was counterbalanced within blocks for each listener, such that the first half of each block of trials used one order (e.g., B on the left, P on the right) while the second half used the other order. Other than this, tasks were randomized.

TABLE I. Structure of Garner paradigm experiment showing stimuli and tasks for both groups in all conditions.

VOT-trained group			
Trained dimension "Is it B or P?"		Untrained dimension "Is it stressed or unstressed?"	
Task	Stimuli	Task	Stimuli
Base line 1	A, B	Base line 1	A, C
Base line 2	C, D	Base line 2	B, D
Filtering	A, B, C, D	Filtering	A, B, C, D
Correlation 1	A, D	Correlation 1	A, D
Correlation 2	B, C	Correlation 2	B, C
F0-trained group			
Trained dimension "Is it B or P?"		Untrained dimension "Is it stressed or unstressed?"	
Task	Stimuli	Task	Stimuli
Base line 1	A, C	Base line 1	A, B
Base line 2	B, D	Base line 2	C, D
Filtering	A, B, C, D	Filtering	A, B, C, D
Correlation 1	A, D	Correlation 1	A, D
Correlation 2	B, C	Correlation 2	B, C

In each of the baseline and correlation tasks, listeners heard repetitions of only two different stimuli, e.g., the A and B tokens or the A and the C tokens, and classified them according to the appropriate categories by pressing a button on a button box corresponding to the category label shown on that side of the screen. For example, A and B would be classified as B and P, respectively, by participants in the VOT-trained group classifying stimuli along the trained dimension, but as unstressed and stressed by participants in the F0-trained group classifying stimuli along the untrained dimension. In the correlation condition stimuli were classified according to both dimensions. For example, the contrast between A and D would be classified as "B and stressed" versus "P and unstressed" by listeners in the VOT-trained condition, and as "P and unstressed" versus "B and stressed" by listeners in the F0-trained condition. In the filtering condition listeners still made a binary decision, e.g., B or P, but all four stimuli were presented in random order (see Table I for a complete description of the distribution of stimuli in each task).

In the base line and correlated conditions there were a total of 64 trials with each pair of sounds (32 trials per stimulus, in random order within blocks). Response choice location and corresponding button was counterbalanced within each block (e.g., half of the trials showed the order "B" "P" and the other half showed "P" "B" from left to right), for a

total of 128 stimulus presentations for both dimensions of contrast (trained and untrained). In the filtering condition there were also 128 total trials (32 per stimulus) and response choice location was similarly counterbalanced. Before the Garner paradigm began, listeners completed a minisession consisting of two trials of each of the two baseline tasks (in random order). Before every block (practice, each baseline condition, each correlated condition, and the filtering task) listeners were also familiarized with the stimuli and their respective labels to be used in the current block, in the same manner as for the STM task. However, unlike the STM task, familiarization was carried out before each block of the Garner task, not before each trial.

Response times for each correct response were averaged according to Dimension of classification (either trained or untrained) and task (base line, filtering) for each subject, and Garner interference was calculated as (filtering RT—baseline RT) for each dimension. Response times were measured from the beginning of the stimulus and no response times less than 350 ms (the maximum duration of the longest male stimulus) were recorded.

3. Training

The six sessions between the pre-test and post-test consisted of training. In each session, listeners heard six blocks of trials, three with the male voice and three with the female one. Each block of trials consisted of stimuli with a different place of articulation (bilabial, alveolar, and velar). Possible responses were always appropriate to the place of articulation (e.g., P or B for the bilabial blocks, "T" or "D" for the alveolar blocks, and "K" or "G" for the velar blocks). In each block, listeners heard eight different stimuli, presented in random order, ten times each. As in the Garner tasks, the trials in the first and second halves of each block used a different response order left to right. The stimuli consisted of the tokens corresponding to those identified in the initial perceptual distance measurement, but with the appropriate consonant place of articulation and vowel quality for the given block. For example, once a given participant demonstrated roughly equal perceptual distances between four /pa/ stimuli, then in the velar blocks of trials that participant would have heard /ka/ and /ki/ syllables with onset F0 and VOT values corresponding to the same steps along their respective continua.

III. RESULTS

A. Training

Overall, training was successful. Looking at performance on the first and last (sixth) days of training, across all training stimuli (male and female, at all places of articulation and in all vowel contexts included in the experiment), listeners in the VOT group improved from 68% to 81% correct, while those in the F0 group improved from 60% to 67% correct. Results of a repeated measures ANOVA with the two factors of group (VOT trained and F0 trained) and training session (days 1 and 6) showed a significant effect of session, $F(1,32)=4.40$, $p=0.001$, and of group, $F(1,32)=9.31$, $p=0.005$, but no interaction, $F(1,32)=3.18$, $p=0.08$. Planned comparisons of means (significance reported for tests at $p<0.05$ or better) on the pretest showed a significant difference between participants assigned to VOT training and those assigned to F0 training, and this difference remained significant on the post-test. However, both groups improved significantly from day 1 of training to day 6. A t-test of difference scores showed no significant difference between the improvement from day 1 to day 6 for the VOT group (13%) and that shown by the F0-trained group (8%). However, this may be a result of the large amount of variance in changes in performance, since 13 out of the 14 participants (93%) in the VOT-trained group showed an improvement from pretest to post-test, as compared to only 15 out of 20 (75%) in the F0-trained group, despite the equalization of perceptual distance along each dimension on a participant-specific basis. This suggests that listeners who were able to learn the F0 contrast were comparatively few, but showed relatively large improvements, while those who learned the VOT contrast were more common, but did not generally show such extreme improvements.

Because we were interested in understanding the effects of learning (successful training), we restricted subsequent analyses to results only from those participants who both achieved at least 70% correct on the final day of training and showed at least 5% improvement in token identification from the first to the last day of training. Repeating the same analysis on only these 16 participants (7 in the F0 group, 9 in the VOT group) showed the expected significant effect of test, $F(1,14)=69.75$, $p<0.001$, but no effect of group, $F(1,14)=3.23$, $p=0.09$, and no interaction, $F(1,14)=0.10$, $p=0.76$ (Fig. 2). Planned comparisons of means showed again that both groups improved significantly (VOT, from 72% to 88% correct; F0 from 64% to 82% correct), but there was no significant difference between the groups on either the pretest or post-test. This suggests that successful learners from both groups showed comparable improvements in performance along the dimension on which they were trained.

B. Perceptual distance (STM)

Responses to targets in the go/no-go STM task were scored as hits while responses to distractors were scored as false alarms. Perceptual distances between each pair of tokens are shown in Table II. Results of a mixed factorial ANOVA of the pretest distances with the between-groups factor of group (VOT-trained, F0-trained) and within-groups factor of pair (AB, CD, AC, BD, and the diagonals AD and BC) showed a significant effect of pair, $F(5,70)=8.05$, $p<0.001$, but no effect of group, $F(1,14)=3.79$, $p=0.07$, and no interaction, $F(5,70)=0.62$, $p=0.68$. *Post hoc* (Tukey HSD, $p=0.05$) tests showed a significant difference *only* between the pairs that make up the sides of the square (AB, CD, AC, BD) and those making up the diagonals (AD and BC), as Euclidean geometry would predict for a square. There were no significant differences between any two sides of the square, and none between the two diagonals, suggesting that the stimuli selected were perceptually "square" (all sides equal, and both diagonals equal). A similar analysis of the post-test data showed comparable results: A significant

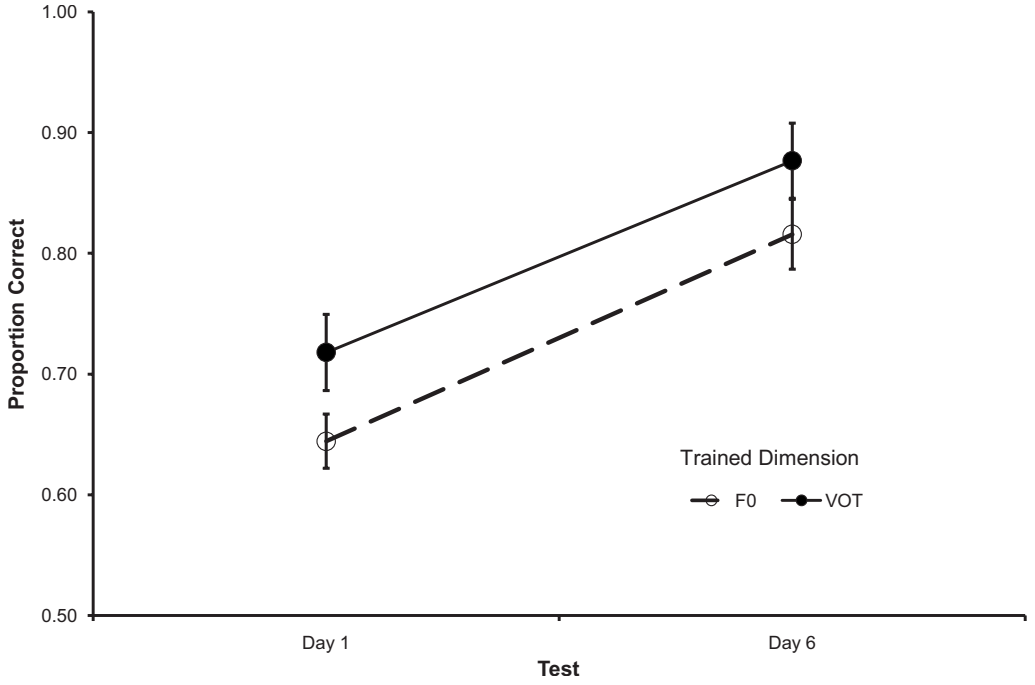


FIG. 2. Proportion of correct consonant identification responses on the first and last days of training for both training groups (successful learners only, see text). Error bars indicate standard error of the mean.

effect of pair, $F(5,70)=9.88$, $p<0.001$, but no effect of group, $F(1,14)=0.18$, $p=0.68$, and no interaction, $F(5,70)=1.78$, $p=0.13$. Again, *post hoc* analyses showed no significant differences between any two sides of the square, and no difference between the two diagonals, although the diagonals were again significantly longer than the sides.

In order to compare performance from pretest to post-test, parallel legs of each square were averaged (e.g., AB and CD were averaged, as were AC and BD) to derive a measure of sensitivity to each dimension for each subject. Results of a mixed factorial ANOVA with between-groups factor of group (VOT-trained, F0-trained) and repeated measures of test (pretest, post-test) and dimension (VOT, onset F0) showed a significant effect of test, $F(1,14)=36.39$, $p<0.001$, but no main effects of group, $F(1,14)=1.02$, $p=0.33$, or of dimension, $F(1,14)=0.30$, $p=0.59$. There was a significant interaction between dimension and group, $F(1,14)=5.35$, $p=0.04$, but no significant interactions between dimension and test, $F(1,14)=0.45$, $p=0.51$, group and test, $F(1,14)=0.29$, $p=0.60$, or between group, dimension and test, $F(1,14)=1.30$, $p=0.27$. Planned comparisons of means (all values reported as significant at $p<0.05$ or better) showed that, for the VOT group, there was a significant increase in sensitivity to the VOT dimension (from a d' of 1.93–3.30) and the F0 dimension (from a d' of 1.51–2.72). Similarly, for the F0-trained group, d' for the VOT dimension increased significantly from 1.28 to 2.51, while for the F0 dimension it increased significantly from 1.29 to 3.12. This suggests that the effect of training on perceptual distance was robust and not constrained to the dimension on which listeners were trained. Overall, these results suggest that the perceptual distances between tokens along each dimension were successfully equated on the pretest, and remained equal after training. Thus, with respect to measures of perceptual distance based on accuracy of speeded target monitoring, training primarily served to increase perceptual distances, and did so to an equivalent degree along both the trained and untrained dimensions.

TABLE II. Perceptual distance, in d' units, fro all pairs of stimuli for both groups on pretest and post-test.

	VOT-trained				F0-trained			
	Pretest		Post-test		Pretest		Post-test	
Pair	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AB	1.59	0.45	2.83	1.65	1.06	0.55	2.20	0.75
CD	2.28	0.84	3.77	1.59	1.49	0.79	2.87	1.81
AC	1.54	0.90	2.88	1.20	1.32	0.70	3.68	1.91
BD	1.48	0.42	2.55	1.15	1.25	0.43	2.56	0.75
AD	2.86	1.59	4.32	1.65	1.74	0.89	4.37	0.82
BC	3.32	1.73	4.52	1.00	2.49	1.26	3.98	1.29

C. Perceptual distance (Garner baseline RT)

Although perceptual sensitivity can be measured in terms of response sensitivity (hit rate and false alarm rate), measures based on response time may be better at differentiating subtle training-related differences between groups. Thus, response times for correct responses in the base line Garner task were averaged for each learner and dimension of classification to provide another measure of perceptual distance between tokens before and after training. Responses made when classifying according to the trained dimension reflect correct responses to the question "is this B or P" while those made when classifying according to the untrained di-

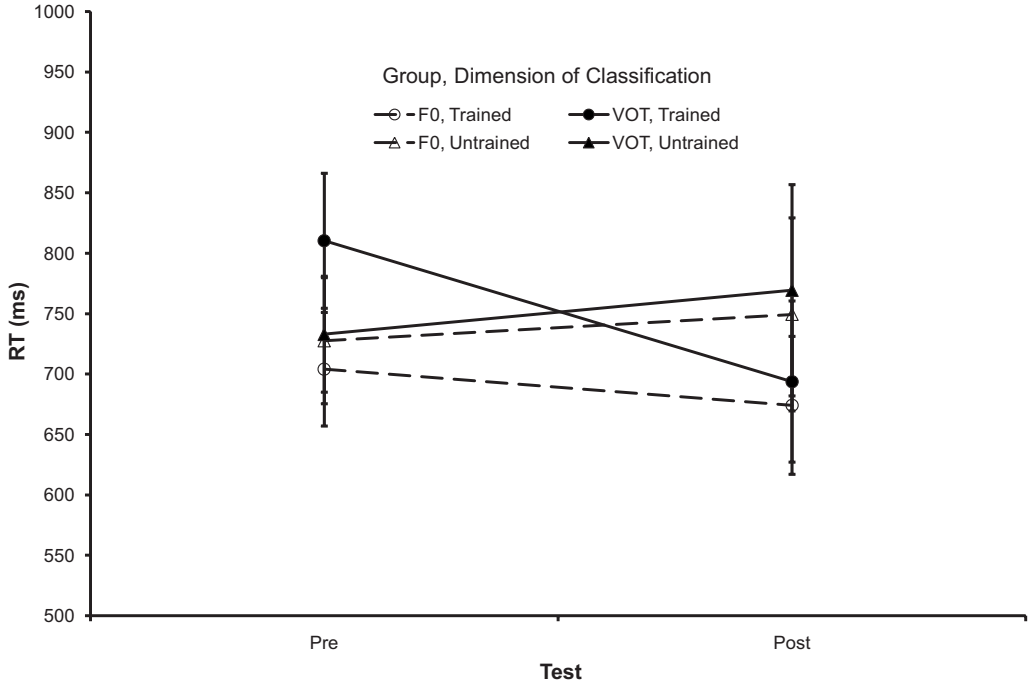


FIG. 3. Pretest and post-test response times on the Garner base line task, classifying stimuli as either [b] or [p] (trained dimension) or "stressed" or "unstressed" (untrained dimension) for both training groups, separated by dimension of classification. Error bars indicate standard error of the mean.

mension reflect response times for classifying according to the other dimension, in response to the question "Is this sound stressed or unstressed?"

A repeated measures ANOVA with one factor between groups (training group, either VOT or onset F0) and two factors within group (test and dimension) showed no significant effects of group, $F(1,14)=0.23$, $p=0.64$, test, $F(1,14)=0.50$, $p=0.49$, or dimension, $F(1,14)=0.53$, $p=0.48$, and no significant interactions between test and group, $F(1,14)=0.33$, $p=0.57$, or between dimension and group, $F(1,14)=0.53$, $p=0.48$. However, the interaction between group, test, and dimension was significant, $F(1,14)=13.35$, $p=0.003$, as shown in Fig. 3. *Post hoc* (Tukey HSD) analysis with a significance threshold of $p=0.05$ showed that the only significant pairwise comparison in the three-way interaction was the 116 ms decrease in baseline response time from pretest (810 ms) to post-test (694 ms) for the VOT-trained group classifying tokens according to the trained (VOT) dimension. The observation that none of the pairwise comparisons for pretest response times showed a significant difference corroborates the findings from the STM task, supporting the claim that stimuli were indeed a perceptual square prior to training. However, the pattern of change in RT, unlike the pattern of change in sensitivity, suggests that only the VOT-trained group showed any appreciable change in perceptual distance between tokens as a result of training, specifically an increase in the distance between tokens along the VOT dimension.

D. Cue weighting

On the pretest, in the correlated task, learners showed no strong evidence in favor of one dimension over another. In the correlated condition involving the A and D tokens stimuli exhibited conflicting values of VOT and onset F0. The A token had a short VOT (similar to a [b]) but a falling F0 contour (like a [p]), while the feature values were reversed for the D token (long VOT like [p] but level F0 onset, more like [b]). Thus, a response of B to the A token or P to the D token would indicate a decision made according to VOT, while a P response to A or a B response to D would indicate a decision made according to onset F0. Overall, learners showed no preference for either cue: 49% of responses to the A token and 51% of those to the D token were consistent with the F0 cue, and this pattern remained even on the post-test (51% and 48%, respectively). This lack of a preference for one cue over another suggests that the bias toward using VOT under normal circumstances (when other cues do not conflict) is not due to something about the VOT dimension *per se*, but rather has to do with the relative size of the interstimulus differences in VOT as compared to those in onset F0.

There was also a very large difference in response patterns between the two training conditions, even on the pretest. The F0-trained group made 88% of pretest and 96% of post-test responses to both the A and D tokens based on onset F0 (responding P and B, respectively), while the VOT group made only 10% and 7% of their responses to the A and D tokens based on F0, respectively (again, responding P to A and B to D). This suggests that the small amount of familiarization that listeners received prior to beginning the pretest was already sufficient to induce them to make phonetic decisions on the basis of the trained rather than the untrained cue. These results suggest, in turn, that listeners' use of a particular cue may be strongly influenced by even short-term experience with a talker or context.

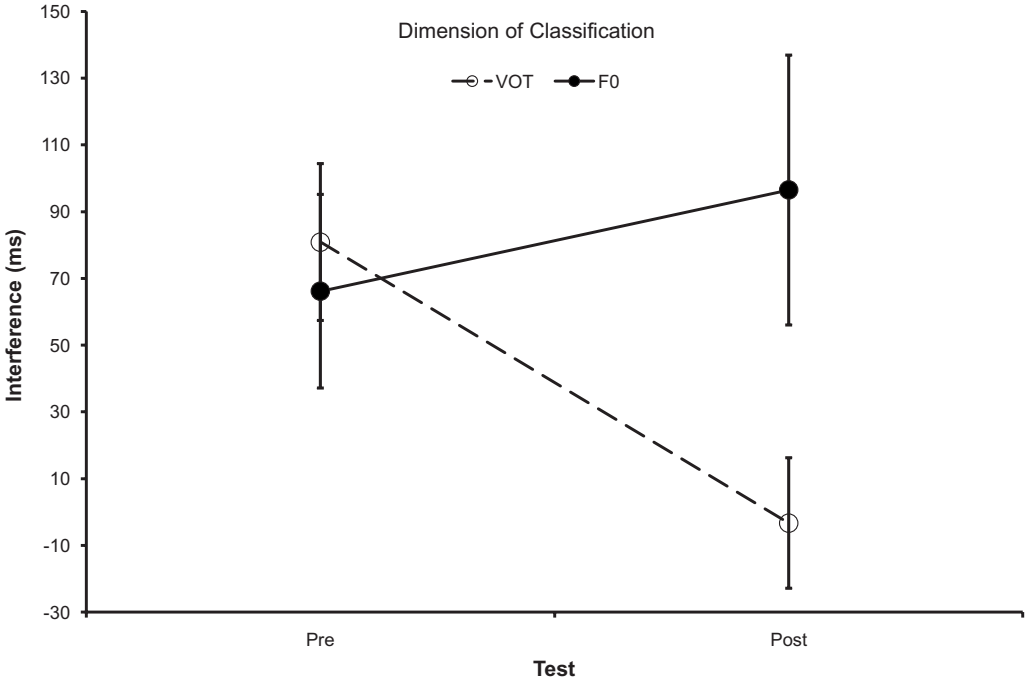


FIG. 4. Garner interference (difference between RT on the Garner filtering task and RT on the Garner base line task, see text for description of tasks) showing significant interaction between test and dimension of classification. Error bars indicate standard error of the mean.

E. Garner interference

Comparison of learners' pretest base line RT with their corresponding filtering RT using a three-way mixed factorial ANOVA with repeated measures of task (baseline, filtering) and dimension of classification (VOT and onset F0), and between-groups factor of training group (VOT and onset F0) showed a significant effect of task, $F(1,14)=9.51$, $p=0.008$, but no effects of group, $F(1,14)=0.54$, $p=0.48$, or dimension, $F(1,14)=2.47$, $p=0.14$, and no interactions. Filtering performance was overall slower (817 ms) than baseline (743 ms) by 74 ms, suggesting that the two dimensions are indeed integral.

Garner interference was computed as the difference in response time between classification according to a given dimension in the filtering task and the average response time for classifying stimuli according to the same dimension in the two baseline tasks using that dimension. These values were submitted to a repeated measures ANOVA with one factor between groups (training group) and two factors within group (test and dimension). Results showed no significant effect of group, $F(1,14)=0.08$, $p=0.78$, test, $F(1,14)=0.62$, $p=0.45$, or dimension, $F(1,14)=1.69$, $p=0.21$, and no interactions between group and test, $F(1,14)=0.08$, $p=0.78$, or group and dimension, $F(1,14)=1.86$, $p=0.19$, and the three-way interaction between test, group, and dimension was not significant, $F(1,14)=1.27$, $p=0.28$. However, there was a significant interaction between test and dimension, $F(1,14)=8.26$, $p=0.01$, suggesting that training had a different effect on the degree of interference of each dimension (Fig. 4). After training, irrelevant variation in F0 no longer interfered with classification according to VOT, but irrelevant variation in VOT continued to interfere with classification according to onset F0.

Although the overall three-way interaction (group by test by dimension) was not significant (Fig. 5), the theoretical basis for the study, namely, the question of whether different kinds of training induce different changes in the processing of the two different dimensions, justified closer examination of some of the contrasts within this interaction. Therefore, a series of planned comparisons were carried out to compare, for each group, the amount of interference for each of the two dimensions on the pretest and on the post-test, as well as the amount of interference for each dimension on the pretest versus the post-test. Significance was set at $p<0.05$. Results showed that, for the F0-trained listeners, there was no significant difference between the degree to which F0 interfered with VOT classification and vice versa on either the pretest or the post-test, and there was no significant difference from pretest to post-test in either the interference of F0 on VOT or vice versa. For the VOT-trained listeners there was no significant difference between VOT or F0 interference on the pretest, but a significant increase from pretest to post-test in interference of VOT on classification according to onset F0 resulted in there being a significant difference on the post-test between the interference of VOT on F0 as compared to vice versa. There were no significant differences in F0 interference from pretest to post-test for this group either.

IV. DISCUSSION

A. Training

Although training can be considered successful for both groups, the degree of learning was unexpectedly low as measured in terms of change in proportion of correct responses from first to last day of training and in terms of the number of trained listeners who exhibited the requisite improvement

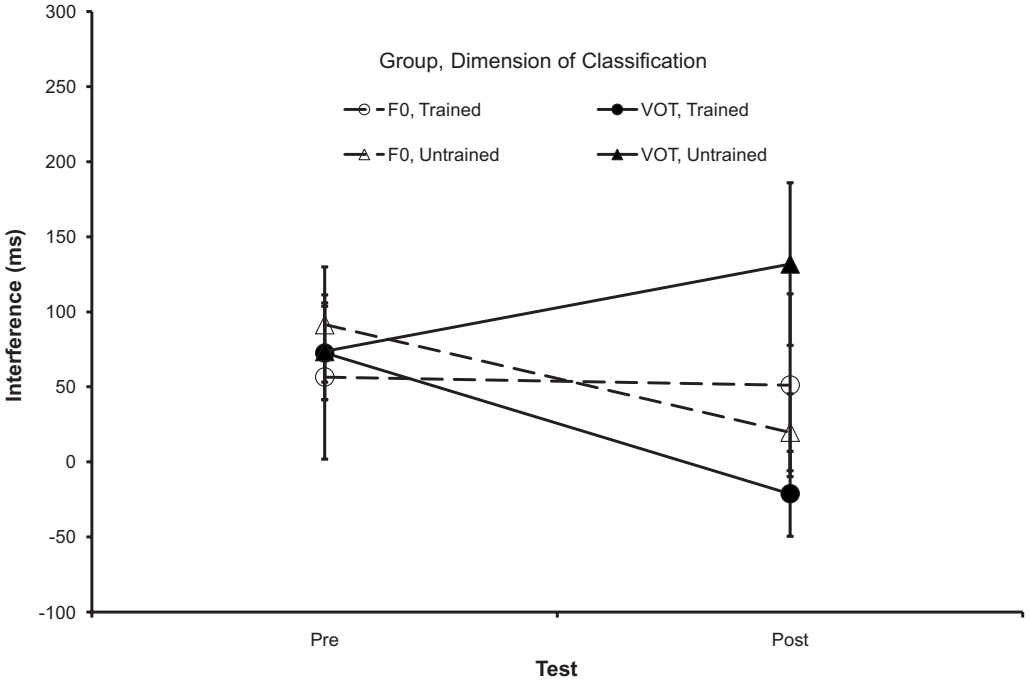


FIG. 5. Differential effects of training on Garner interference (difference between RT on the Garner filtering task and RT on the Garner base line task) for VOT- and F0-trained groups, separated by dimension of classification. Error bars indicate standard error of the mean.

in performance. Previous studies training listeners to develop new categories based on non-native VOT differences (e.g., Holt *et al.*, 2004; Pisoni *et al.* 1982) gave less training and yet showed noticeably better improvement than was found in the present experiment, even for the VOT-trained listeners. Although the training results of Pisoni *et al.* (1982) may have been better than those observed here because of their use of a different location in the VOT continuum (they trained listeners to distinguish between a prevoiced category with negative VOT and a short-lag category), the intended category boundary of experiment 1 of Holt *et al.* (2004), "inconsistent" group, was quite similar to the VOT difference in the present experiment, yet listeners of Holt *et al.* (2004) achieved an identification rate of 90% correct or better within about eight blocks of training (about 380 stimulus presentations).

One possible explanation for the poor rate of learning in the present experiment is that, by using very similar VOT and onset F0 values for all of the training stimuli, regardless of place of articulation (POA), we provided less variability than would be found in natural speech. More significantly, this lack of variability is contrary to the typical correlation between VOT and POA, in which VOT increases as POA moves back in the oral cavity (from bilabial to alveolar to velar) (Lisker and Abramson, 1964). The lack of an expected correspondence of this sort between POA and VOT may have made the additional (non-[pa]) tokens less effective for training, and might conceivably have interfered with learning in some way.

Another major factor that probably contributed significantly to the comparatively low learning rate for listeners in the present experiment is the inconsistent mapping between response category and response button in both testing and training. Although this was done intentionally in an attempt to encourage listeners to develop more abstract categories less closely associated with a specific response key, it almost certainly made the task considerably more difficult. Shiffrin and Schneider (1977) have shown that it is much harder to learn an inconsistent mapping between stimulus and response in which the assignment of stimulus to response changes than a consistent mapping in which stimuli have the same response across trials. Although in the present case the mapping was, at one level, consistent (i.e., shorter VOT values always mapped onto the response B for listeners in the VOT-trained condition), the mapping between the category label B and the response key (left or right) was inconsistent, and this presumably contributed to poorer performance on this task.[4]

B. Perceptual weighting

In the present study, perceptual distance was successfully equated along the two dimensions of VOT and onset F0, as indicated by the results of the pretest STM (d') and Garner base line (RT) tasks. This suggests that the typically observed pattern of using VOT in preference to onset F0 as a cue to voicing in syllable-initial stops (e.g., Abramson and Lisker, 1985; Francis and Nusbaum, 2002; Gordon *et al.*, 1993; Lisker, 1978) can apparently be eliminated at least at the level measurable by discrimination and classification (and at least for tokens that lie within the onset F0 and VOT range of voiceless aspirated stops). In addition, overall performance on the conflicting-cue tokens in the correlated task suggested that listeners showed no *a priori* preference for using VOT over F0, and just a few instances of familiarization were sufficient to induce listeners from both groups to rely heavily on one cue instead of the other. This further supports the hypothesis that preference for VOT is based

strongly on unequal perceptual distance, and does not derive from any special intrinsic property of VOT as a dimension of perceptual contrast.

C. Dimensional integrality

With respect to the question of integrality, results from the Garner interference task on the pretest suggest that the two dimensions of VOT and onset F0 are integral in the sense of Garner (1974). This is consistent with other research on the integrality of speech dimensions (Kingston and Macmillan, 1995; Kingston *et al.*, 1997; Macmillan *et al.*, 1999). Interference was symmetrical on the pretest, such that there was no significant difference in magnitude between the interference of irrelevant variability in onset F0 on classification according to VOT and vice versa, for either of the two groups of learners. This pattern of results is consistent with the hypothesis that any preference for using VOT over onset F0 in classifying voicing in syllable-initial stop consonants derives from unequal perceptual distances along the two dimensions, and not from any preferred quality of VOT. When the perceptual distances were equated along both dimensions in the present experiment, integrality was symmetrical. However, after training, asymmetry increased, at least for the learners in the VOT group, such that there was significantly less interference from irrelevant variability in the untrained dimension (onset F0) on classification according to the trained dimension (VOT) than vice versa. These results (for the VOT-trained listeners), in turn, are consistent with the hypothesis that training served primarily to increase perceptual distance along the trained dimension (VOT). As demonstrated by Melara and Mounts (1994), unequal perceptual distances between tokens along two different dimensions result in increased interference from the larger dimension. Results of the present experiment suggest that, after successfully learning to rely more heavily on VOT and to better ignore onset F0, the perceptual distance between tokens along the VOT dimension was increased with respect to that along the onset F0 dimension for successful VOT-trained listeners, resulting in the observed pattern of increased interference. As discussed below in Sec. IV E other results together suggest that this change resulted primarily from increased distance along the VOT dimension, and not decreased distance along onset F0.

D. Perceptual learning

In this experiment, perceptual distance was calculated in two ways, using d' (sensitivity) in a STM task, and using response time on a Garner speeded classification task. Results were somewhat contradictory, in that the STM task indicated that both groups of learners showed significantly increased perceptual distance along both their untrained and trained dimensions as a result of training, while the classification task indicated that only the VOT-trained group showed an increase in perceptual distance as a result of training, and that occurred only along the trained dimension (see below for a discussion of possible reasons for these differences between monitoring sensitivity and classification response time). At the least, the results from the Garner baseline task lend tentative support to the hypothesis that there may be something special about VOT, as a phonetic cue, that makes it easier to learn than onset F0 (though not easier to use as a cue when perceptual distances are equated): Both groups of listeners were given the same number of trials with the same stimuli, but the VOT-trained group showed, overall, more evidence of stronger learning, including (1) a greater improvement as a result of training (for the entire training group), (2) a greater proportion of listeners showing evidence of learning (greater than five percentage-point increase, with a final score above 70% correct), and (3) the significant changes in Garner interference discussed in the previous section.

While the present results suggest that it may be easier to direct (even) more attention to VOT than to either divert attention from VOT or distribute more attention to onset F0, it is only possible to speculate in a broad manner about possible reasons for such asymmetry in learnability. The most obvious explanation is that American English listeners are simply more used to directing attention to VOT than to onset F0 (cf. Francis and Nusbaum, 2002; Gordon *et al.*, 1993), and thus increasing attention to an already dominant dimension of contrast comes relatively easily. In contrast, inhibiting such a cue may be considerably more difficult, especially since listeners in these studies spend relatively little time in training compared to the amount of time they spend speaking their native language outside the laboratory (where giving greater weight to VOT is clearly a beneficial strategy).

This possibility may be further compounded by the fact that, in testing, listeners were not directed to make judgments about the specific dimensions in question, as would occur in a typical Garner paradigm (e.g., "classify the sounds according to the *pitch* dimension, as either high or low"). Rather, because the dimensions of VOT and onset F0 are not usually thought of as being consciously accessible to untrained listeners, linguistically plausible contrasts were chosen ([b]/[p] for voiced/voiceless, and stressed/unstressed) with the intent that each of these two dimensions should map sufficiently well onto either of the two acoustic cue contrasts (VOT or onset F0). That is, the goal was to use two dimensions such that the mapping between a short VOT stimulus and the response B would be equally acceptable to naïve listeners as that between a short VOT stimulus and the response "unstressed" (and similarly for mappings between shallow onset F0 declines and B and unstressed responses, as well as for long VOT/sharp onset F0 declines and P or "stressed" responses). However, although all expected mappings are plausible *a priori* (stressed syllables do have longer VOT and higher F0 than unstressed ones, and voiced sounds do have shorter VOT and a less negative slope of onset F0 than do voiceless ones), these linguistic dimensions do not, in fact, map equally well onto each respective response for native speakers of English. Not only are English speakers more accustomed to making voicing distinctions based on VOT, not onset F0 (as discussed in the previous paragraph), but they are also more accustomed to making stress distinctions on the basis of F0 than on the basis of VOT. Thus, testing conditions, in terms of the mappings between response items and acoustic dimensions, were much more natural for the

VOT-trained listeners, who were tested with the P/B contrast mapping onto the VOT difference and stressed/unstressed mapping onto onset F0 difference, than for the onset F0-trained listeners, who were tested with the P/B contrast mapping onto the onset F0 difference and stressed/unstressed mapping onto VOT. In other words, our indices of perceptual distance and the distribution of selective attention may be confounded, for the onset F0 group, with experiment design-specific factors, and this might explain why the onset F0 group showed a comparable degree of improvement to the VOT-trained group on the training task (measured in terms of proportion correct identification), but failed to show any evidence of a differential change in the processing of onset F0 as opposed to VOT that might explain this improvement.

On the other hand, it is also possible that there is something intrinsically more learnable about the acoustic properties that comprise VOT as opposed to onset F0 (i.e., an advantage for learning temporal as opposed to spectral contrasts), but to test this hypothesis would require eliminating the bias induced by native language experience, for example, by identifying and training listeners whose native language weighted onset F0 equally with VOT (one such possible example might be Korean, cf. Francis and Nusbaum, 2002). Finally, it may also be noted that training of this sort served primarily to improve the speed with which listeners were able to make a decision, and such an improvement was disproportionately advantageous for decisions based on VOT which is fundamentally temporal in nature and occurs earlier in the syllable, as opposed to onset F0 which involves both spectral and temporal properties and occurs later in the syllable.

E. Enhancement versus inhibition

Although the two methods used to measure perceptual distance (sensitivity in speeded target monitoring versus response time in speeded classification) provided somewhat discrepant results (see below), it is important to note that both methods provided strong evidence that training served only to increase the perceptual distance between tokens (acquired distinctiveness), not to decrease it (acquired similarity). Only the VOT group showed a change in interference, and this was only in terms of the decrease in interference of the untrained on the trained. The (expected) corresponding increase in interference of the trained on the untrained was not significant, although the trend was definitely in the expected direction. Given that the dimensions of VOT and onset F0 are highly integral, these results are entirely consistent with results from previous research. In particular, Goldstone (1994) also found evidence for increased perceptual distance along a variety of trained dimensions in a visual category learning experiment, but only found evidence of decreased perceptual distance along a to-be-ignored dimension when the two dimensions were perceptually separable in the sense of Garner (1974). Indeed, cases of true acquired similarity seem to be relatively rare in the perceptual learning literature [cf. Guenther *et al.* (1999) for discussion, and Francis and Nusbaum, (2002), for an example of acquired similarity with more natural stimuli].

There are at least two ways to characterize the difference between acquired similarity and acquired distinctiveness. Iverson and co-workers (Iverson and Kuhl, 2000; Iverson *et al.*, 2003) have argued that acquired similarity arises from properties of the statistical distribution of input stimuli in perceptual space in a manner independent of attention, while acquired distinctiveness results from the operation of an attentionally demanding process. Although there is now evidence that even passive statistical learning depends on the availability of attentional resources (Toro *et al.* 2005), there is also evidence that the development of acquired similarity can be facilitated by certain distributional properties of the training stimuli. Thus, the Iverson argument may still be valid, despite the almost certain involvement of attention in the process of phonetic cue learning. In support of a role for distributional factors, Guenther *et al.* (1999) found that in order to induce increased similarity, it was necessary to provide not only categorization training (as in the present experiment) but also multiple exemplars of each category. They argued that experience with multiple exemplars encouraged listeners to ignore small (noncategorical) differences between stimuli within a single category, an effect impossible to achieve when training with only a single exemplar [see also Iverson *et al.* (2005) for similar arguments related to a test of the efficacy of high variability training].

On the other hand, Goldstone (1994) and Francis and Nusbaum (2002) argued that the processes of acquired distinctiveness and acquired similarity may be employed at different stages in the learning process, and/or under different conditions of stimulus properties. In cases such as the present experiment and those of Goldstone (1994) in which stimuli are perceptually highly similar (located within a single native category in the present case, or within one (just noticeable difference) JND of one another in the Goldstone case), acquired distinctiveness is the most effective strategy for significantly improving categorization quickly. In contrast, under conditions in which stimuli are already relatively easy to categorize (e.g., certain contrasts in the Korean stimuli used by Francis and Nusbaum, 2002), acquired similarity, especially along an irrelevant dimension of contrast, leads to a more significant improvement in categorization that would simply further increase the already salient difference between the two categories along an already contrastive dimension.

Of course, the two accounts are not necessarily mutually exclusive, in the sense that the presence of multiple exemplars within each category increases the probability that variance within the category is relatively high, which in turn increases the likely benefit of applying a process of acquired similarity to reduce within-category variability. In the present case, however, listeners were trained with multiple exemplars, but these exemplars were acoustically extremely similar to the test stimuli along the critical dimensions of onset F0 and VOT, and yet the two categories represented by these exemplars (and by the test stimuli) were extremely close to one another in perceptual space. Thus, in this case, although listeners received multiple training exemplars, one might argue that they were not distributed in a manner that would be expected to promote acquired similarity on the basis of either

of these two hypotheses. The distribution of training exemplars was not sufficiently broad to engage a Guenther/Iverson type of mechanism, and the overall similarity of the two categories was sufficiently great to engage a mechanism of acquired distinctiveness over one of acquired similarity in a Francis/Goldstone type of model. Further research is clearly necessary to explore the basis for these two kinds of processes.

F. Differences between monitoring sensitivity and classification response time

One curious finding in the present results is the apparent disagreement between the two measures of perceptual distance employed, sensitivity in speeded target monitoring and response time in speeded classification. While the sensitivity results indicated that listeners in both groups showed equivalently increased perceptual distances along both their trained and untrained dimensions of contrast, the response-time data suggested that only the VOT-trained listeners showed a change in perceptual distance, and this increase occurred only along VOT, the dimension on which they were trained.

This finding is particularly curious given the commonly accepted assumption that response time and accuracy tasks are assumed to measure more or less the same thing (perceptual distance between tokens). Ashby and Maddox (1994) discuss the widespread nature of this assumption as they develop an explicit model relating RT performance to perceptual distance between tokens and decision (category) boundaries, based on general recognition theory (GRT) (Ashby and Townsend, 1986). Specifically, they propose that RT should decrease monotonically as a function of the perceptual distance between the stimulus and the decision bound. Furthermore, the GRT as well as other theories of similar phenomena (e.g., Luce, 1986) clearly demonstrate that difficult discriminations are associated with longer response times. Thus, we have every reason to expect a correspondence between RT and accuracy measures: As stimuli become more distant from one another in perceptual space, they should become both easier to identify (in the STM task) and correct identifications should be faster (in the Garner base line task). However, it is possible that, in the present case, specific details of the experiment design unintentionally predisposed listeners to treat the two tasks differently with respect to the type of memory or attentional mechanisms they employed, resulting in a divergence between the results of the two tasks.

One potentially important difference between the two tasks in the present experiment is that, in the STM task, listeners received much more frequent familiarization with exemplars of the two categories they using than they did on the classification task. In the STM task, listeners heard two presentations of each of the two stimuli in a given trial (e.g., the A and B tokens), accompanied by visual presentation of their associated category label, before every trial. On the other hand, in the classification task, listeners were familiarized with the stimulus-symbol pairing only three times, once before each *block* of trials (baseline, correlated, and filtering). Thus, performance in the STM task may better reflect listeners' ability to compare each test stimulus with short-term memory traces of the familiarization stimuli, while performance on the Garner base line task better reflects listeners' ability to compare test stimuli with long(er)-term category representations [see Xu *et al.* (2006) for a model of memory for phonetic categorization].

Macmillan (1987) distinguishes between sensory or trace and context modes of processing. In the trace mode, processing is dominated by comparison of (temporary) sensory traces of stimuli, while in context coding processing involves comparison between sensory traces of stimuli and (longer-term) perceptual anchors, including category representations. In this sense, the different familiarization protocols for the two types of tasks may have encouraged a greater degree of reliance on sensory coding in the STM task and on context coding in the classification (Garner base line) task. That is, performance measured in terms of accuracy on the STM task may serve mainly to indicate listeners' ability to retain and make use of short-term memory traces of the familiarization stimuli. As listeners learned which properties of the signal (VOT and onset F0) varied across the training stimuli, they may have become better able to encode and retrieve these properties as short-term memory traces (i.e., when exposed to the tokens during familiarization). Since both properties varied equally across the training set, listeners showed an equal degree of improvement in encoding and retrieving memory traces of these properties.

On the other hand, RT performance on the Garner base line task may better reflect listeners' ability to access stored long-term representations of phonetic categories (context coding). It has been argued that perceptual learning based on categorization training (as used here) primarily affects categorization at the level of context coding (Guenther *et al.*, 1999). According to this hypothesis, training was successful in changing the long-term representations of the categories that listeners were learning (e.g., B versus P), but this only became obvious in the Garner baseline (RT) task because there was sufficient time between the presentation of the familiarization stimuli and the actual test trials that listeners were not able to rely solely on trace memories of the familiarization stimuli and instead had to depend on their long-term memories of the different (learned) category representations. Thus, it may be argued that the results of the Garner base line task are more indicative of the overall phonetic consequences of this kind of training than are those of the STM task, because they better reflect changes in listeners' attention to features encoded in long-term memory representations of the learned categories, while the results of the STM task reflect instead an increase in overall sensitivity to those acoustic properties that varied during training as a result of increased attention to the speech signal under conditions of higher uncertainty (Nusbaum and Magnuson, 1997; Nusbaum and Schwab, 1986; Wong *et al.*, 2004).

G. The role of attention in phonetic learning

Gordon *et al.* (1993) showed that, under conditions of (comparatively) unlimited attentional load, American English listeners gave more weight to VOT than to onset F0 in a voicing decision. In contrast, under conditions of more limited attentional availability, listeners showed a greater re-

duction in the weight given to VOT than in that given to onset F0. They argued that weak acoustic cues (e.g., onset F0) require comparatively little attention to make their full contribution to a phonetic decision (thus benefiting little from an increased availability of attention), while stronger cues (e.g., VOT) benefit more from increased availability of attentional resources. We elaborate on this hypothesis by proposing that using *any* cue requires some commitment of attention, but that attention is allocated dynamically depending on the current diagnosticity of specific cues. Under normal circumstances those cues that have proven to be most diagnostic (e.g., over the course of prior experience) receive the lion's share. Under conditions of limited attentional availability, the proportion of capacity devoted to each cue is reduced proportionally, with strong cues continuing to receive proportionally more of the smaller pool of available resources. In new contexts or under conditions of uncertainty (i.e., multiple talkers, high noise, etc.), the distribution of attention to individual cues may vary as the speech perception mechanism begins to seek out cues that are potentially more diagnostic under those conditions (Nusbaum and Magnuson, 1997; Nusbaum and Schwab, 1986; Wong *et al.*, 2004). Such reallocation may result in a more even distribution of resources across cues as attention is withdrawn from cues that are typically stronger but fail to be sufficiently diagnostic in the present context, and reallocated toward cues that, though typically weaker, might potentially be more diagnostic in the present case.

In this dynamic redistribution of attention we see a reconciliation between the effects of training and the effects of experimental task observed in the present experiment. On the one hand, perceptual training may alter the base line distribution of attention to specific cues, increasing the weight given to cues that are sufficient for identifying the newly learned categories, and reducing that given to less diagnostic cues. That it does so preferentially for VOT and less so for onset F0 suggests that there is something special about VOT, at least as a cue to the perception of syllable-initial stop-consonant voicing by native speakers of English. On the other hand, frequent presentations of representative stimuli differing along two dimensions (as in the STM task) may encourage listeners to maintain a high level of attention to both cues to facilitate the use of trace coding. Thus, the ability of training to accomplish the redistribution of attention among acoustic cues may only become obvious under conditions in which listeners are not constantly reminded of the multiple dimensions (diagnostic and nondiagnostic) along which stimuli differ, and instead are forced to focus on stimulus differences that have been encoded in the long-term mental representations of the learned categories.

Ultimately, this perspective is compatible with Kuhl's *neural commitment* theory (Kuhl *et al.*, 2006), in the sense that English listeners appear to have committed to VOT to a greater degree than to onset F0 (at least as a cue to the phonetic property of voicing in syllable-initial stops), and reducing that commitment, or increasing their commitment to onset F0, seems to require more training, or different kinds of training, than we have employed here. Whether this commitment derives from innate differences in the neural systems that process VOT as compared to onset F0, or from experience-dependent development of such systems is a question beyond the scope of the present paper. However, by considering such neural commitment in terms of the distribution of attentional resources we are able to link the role of attention in perceptual learning (Guion and Pederson, 2007; Strange, 2006) to processes of online speech perception (Gordon *et al.*, 1993), making a connection that is obviously necessary, but thus far only occasionally discussed (Nusbaum and Goodman, 1994; Stevens *et al.*, 2006; Toro *et al.*, 2005).

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute on Deafness and other Communication Disorders (NIH-NIDCD R03DC006811) to A.L.F. We would like to thank Bob Melara, Howard Nusbaum, John Kingston, and an anonymous reviewer for suggestions on earlier drafts of this article. Some of these results were presented at the fourth Joint Meeting of the Acoustical Society of American and the Acoustical Society of Japan, Honolulu, HI, November 28–December 2, 2006.

[1]Although the dimension of VOT has been explored in considerable depth, the dimension of onset F0 is less well investigated, and to our knowledge there are no studies that provide quantitative data on listeners' sensitivity to onset F0 differences comparable to the wealth of information available regarding VOT (see Holt *et al.*, 2004 for discussion).

[2]Note that subsequent research (e.g., Löfqvist *et al.*, 1989) supports a physiological origin of the onset F0 property of stop consonants in the degree of tension of the cricothyroid muscle, suggesting that there is no direct physiological link between onset F0 and VOT cues. This physiological dissociation is further supported by the patterning of these two cues in three-way stop consonant systems such as that of Korean and Thai, in which stop categories are distinguished by independent onset F0 and VOT properties (Thai: Gandour, 1974; Korean: Francis and Nusbaum, 2002; see Francis *et al.*, 2006 for discussion).

[3]While step size was maintained as closely as possible across tokens and talkers, when specific values are given here they refer to the test stimuli based on [p^ha]. Other stimuli varied slightly from these specific values to preserve some degree of interstimulus variability, but never by more than 5 ms or two percentage points (for frequency modifications) from the values given here.

[4]In an ongoing study using nonspeech sounds in a similar testing/training paradigm, we have found that simply eliminating this inconsistent mapping between response label and response key improves learning considerably both in terms of the number of listeners who are able to reach criterion, and in terms of the magnitude of the overall change in proportion correct identification from the first to the last day of training.

Abramson, A. S. (**1977**). "Laryngeal timing in consonant distinctions," Phonetica **34**, 295–303.

Abramson, A. S., and Lisker, L. (**1970**). "Discrimination along the voicing continuum: Cross-language tests," Proceedings of the 6th International Congress on Phonetic Science, Prague, **1967**, Academia, Prague, pp. 569–573.

Abramson, A. S., and Lisker, L. (**1985**). "Relative power of cues: F0 shift versus voice timing," in *Linguistic Phonetics*, edited by V. Fromkin (Academic New York), pp. 25–33.

Allen, J., Kraus, N., and Bradlow, A. (**2000**).. "Neural representation of consciously imperceptible speech sound differences," Percept. Psychophys. **62**, 1383–1393.

Ashby, F. G., and Maddox, W. T. (**1994**). "A response time theory of separability and integrality in speeded classification," J. Math. Psychol. **38**, 423–466.

Ashby, F. G., and Townsend, J. T. (**1986**). "Varieties of perceptual independence," Psychol. Rev. **93**, 154–179.

Boersma, P., and Weenink, D. (**2006**). Praat: doing phonetics by computer (Version 4.2) (computer program). http://www.praat.org/ (last accessed March 17, **2008**).
Diehl, R. L., and Kluender, K. R. (**1989**). "On the objects of speech perception," Ecological Psychol. **1**, 121–144.
Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.
Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**, 1668–1680.
Francis, A. L., Ciocca, V., Wong, V. K. M., and Chan, J. K. L. (**2006**). "Is fundamental frequency a cue to aspiration in initial stops?," J. Acoust. Soc. Am. **120**, 2884–2895.
Gandour, J. (**1974**). "Consonant types and tone in Siamese," J. Phonetics **2**, 337–350.
Garner, W. R. (**1974**). *The Processing of Information and Structure* (Erlbaum, Hillsdale, NJ).
Garner, W. R. (**1983**). "Asymmetric interactions of stimulus dimensions in perceptual information processing," in *Perception, Cognition, and Development: Interactional Analyses*, edited by T. J. Tighe and B. E. Shepp (Erlbaum, Hillsdale, NJ), pp. 1–37.
Gibson, E. J. (**1969**). *Principles of Perceptual Learning and Development* (Appleton-Century-Crofts, New York).
Goldstone, R. (**1994**). "Influences of categorization on perceptual discrimination," J. Exp. Psychol. Gen. **123**, 178–200.
Gordon, P. C., Eberhardt, J. L., and Rueckl, J. G. (**1993**). "Attentional modulation of the phonetic significance of acoustic cues," Cogn. Psychol. **25**, 1–42.
Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (**1999**). "Effects of categorization and discrimination training on auditory perceptual space," J. Acoust. Soc. Am. **106**, 2900–2912.
Guion, S. G., and Pederson, E. (**2007**). "Investigating the role of attention in phonetic learning," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. Munro (Benjamins, Amsterdam), pp. 57–77.
Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.
Haggard, M. P., Summerfield, Q., and Roberts, M. (**1981**). "Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F0 cues in the voiced-voiceless distinction," J. Phonetics **9**, 49–62.
Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.
Holt, L. L., Lotto, A. J., and Diehl, R. L. (**2004**). "Auditory discontinuities interact with categorization: Implications for speech perception," J. Acoust. Soc. Am. **116**, 1763–1773.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?," J. Acoust. Soc. Am. **109**, 764–774.
Hombert, J. M. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 77–111.
Iverson, P., and Kuhl, P. K. (**2000**). "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?," Percept. Psychophys. **62**, 874–886.
Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.
Jusczyk, P. W. (**1993**). "From general to language-specific capacities: The WRAPSA model of how speech perception develops," J. Phonetics **21**, 3–28.
Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–494.
Kingston, J., Diehl, R. L., Kirk, C. J., and Castleman, W. A. (**2008**). "On the internal perceptual structure of distinctive features: The [voice] contrast," J. Phonetics **36**, 28–54.
Kingston, J., and Macmillan, N. A. (**1995**). "Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm," J. Acoust. Soc. Am. **97**, 1261–1285.
Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., and Bartels, C. (**1997**). "Integrality in the perception of tongue root position and voice quality in vowels," J. Acoust. Soc. Am. **101**, 1696–1709.
Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (**2006**). "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months," Dev. Sci. **9**, F13–F21.
Liberman, A. M. (**1957**). "Some results of research on speech perception," J. Acoust. Soc. Am. **29**, 117–123.
Lisker, L. (**1978**). "In qualified defense of VOT," Lang Speech **21**, 375–383.
Lisker, L. (**1986**). ""Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.
Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.
Löfqvist, A., Baer, T., McGarr, N. S., and Seider Story, R. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.
Luce, R. D. (**1986**). *Response Times: Their Role in Inferring Elementary Mental Organization* (Oxford University Press, Oxford).
Lutfi, R. A., and Liu, C.-J. (**2007**). "Individual differences in source identification from synthesized impact sounds," J. Acoust. Soc. Am. **122**, 1017–1028.
Macmillan, N. A. (**1987**). "Beyond the categorical/continuous distinction: A psychophysical approach to processing modes," in *Categorical Perception*, edited by S. Harnad (Cambridge University Press, New York), pp. 53–85.
Macmillan, N. A., and Creelman, C. D. (**2004**). *Detection Theory: A User's Guide*, 2nd ed. (Lawrence Erlbaum Associates, Hillsdale, NJ).
Macmillan, N. A., Kingston, J., Thorburn, R., Dickey, L. W., and Bartels, C. (**1999**). "Integrality of nasalization and F1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions," J. Acoust. Soc. Am. **106**, 2913–2932.
Massaro, D. W., and Cohen, M. M. (**1976**). "The contribution of fundamental frequency and voice onset times to the /zi/-/si/ distinction," J. Acoust. Soc. Am. **60**, 704–717.
Massaro, D. W., and Cohen, M. M. (**1977**). "Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction," Percept. Psychophys. **22**, 373–382.
Melara, R. D., and Mounts, J. R. W. (**1994**). "Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty," Percept. Psychophys. **56**, 73–90.
Nosofsky, R. M. (**1986**). "Attention, similarity, and the identification-categorization relationship." J. Exp. Psychol. Gen. **115**, 39–57.
Nusbaum, H. C., and Goodman, J. C. (**1994**). "Learning to hear speech as spoken language," in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum, (MIT Press, Cambridge, MA), pp. 299–338.
Nusbaum, H. C., and Magnuson, J. (**1997**). "Talker normalization: Phonetic constancy as a cognitive process," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix, (Academic, San Diego, CA), pp. 109–132.
Nusbaum, H. C., and Schwab, E. C. (**1986**). "The role of attention and active processing in speech perception," in *Pattern Recognition by Humans and machines*, edited by E. C. Schwab and H. C. Nusbaum (Academic, San Diego), Vol. **1**, pp. 113–157.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," J. Exp. Psychol. Hum. Percept. Perform. **8**, 297–314.
Pomerantz, J. R., Pristach, E. A., and Carson, C. E. (**1989**). "Attention and object perception," in *Object Perception: Structure and Process*, edited by B. Shepp and S. Ballesteros (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 53–89.
Raphael, L. J. (**2005**). "Acoustic cues to the perception of segmental phonemes," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA), pp. 182–206.
Repp, B. H. (**1979**). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," Lang Speech **22**, 173–189.
Schouten, M. E. (**1985**). "Identification and discrimination of sweep tones," Percept. Psychophys. **37**, 369–376.
Shiffrin, R. M., and Schneider, W. (**1977**). "Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory," Psychol. Rev. **84**, 127–190.
Stevens, K. N., and Klatt, D. H. (**1974**). "Role of formant transitions in the voiced-voiceless distinction for stops," J. Acoust. Soc. Am. **55**(3), 653–659.

Stevens, C., Sanders, L., and Neville, H. (**2006**). "Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment," Brain Res. **1111**, 143–152.
Strange, W. (**2006**). "Second-language speech perception: The modification of automatic perceptual routines. Paper presented at the Fourth Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, November-December, 2006, Honolulu, HI. [Abstract]," J. Acoust. Soc. Am. **120**, 3137.
Tong, Y., Francis, A. L., and Gandour, J. T. (**2008**), "Processing dependencies between segmental and suprasegmental features in Mandarin Chinese," Lang. Cognit. Processes **23**, 689–708.
Toro, J. M., Sinnett, S., and Soto-Faraco, S. (**2005**). "Speech segmentation by statistical learning depends on attention," Cognition **97**, B25–B34.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "F0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.
Wong, P. C. M., Nusbaum, H. C., and Small, S. L. (**2004**). "Neural bases of talker normalization," J. Cogn Neurosci. **16**, 1173–1184.
Xu, Y., Gandour, J. T., and Francis, A. L. (**2006**). "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am. **120**, 1063–1074.
Xu, Y., Krishnan, A., and Gandour, J. (**2006**). "Specificity of experience-dependent pitch representation in the brainstem," NeuroReport **17**, 1601–1605.
Zatorre, R., and Belin, P. (**2001**). "Spectral and temporal processing in human auditory cortex," Cereb. Cortex **11**, 946–953.

Coding of intonational meanings beyond F0: Evidence from utterance-final /t/ aspiration in German

Oliver Niebuhr[a)]
Institute of Phonetics and Digital Speech Processing (IPDS), Christian-Albrecht-University, Kiel, Germany

(Received 6 May 2007; revised 3 May 2008; accepted 20 May 2008)

An acoustic analysis of a German read-speech corpus showed that utterance-final /t/ aspirations differ systematically depending on the accompanying nuclear accent contour. Two contours were included: Terminal-falling early and late F0 peaks in terms of the Kiel Intonation Model. They correspond to H+L*L−% and L*+HL−% within the autosegmental metrical (AM) model. Aspirations in early-peak contexts were characterized by (a) "short", (b) "high-intensity" noise with (c) "low" frequency values for the spectral energy maximum above the lower spectral energy boundary. The opposite holds for aspirations accompanying late-peak productions. Starting from the acoustic analysis, a perception experiment was performed using a variant of the semantic differential paradigm. The stimuli were varied in the duration and intensity pattern as well as the spectral energy pattern of the final /t/ aspiration. Results revealed that the different noise patterns found in connection with early and late peak productions were able to change the attitudinal meaning of the stimuli toward the meaning profile of the respective F0 peak category. This suggests that final aspirations can be part of the coding of meanings, so far solely associated with intonation contours. Hence, the traditionally separated segmental and suprasegmental coding levels seem to be more intertwined than previously thought. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2940588]

PACS number(s): 43.71.Es, 43.70.Fq [MSS] Pages: 1252–1263

I. INTRODUCTION

Many different concepts were developed to describe and to structure the intonation patterns of languages. Basically, these concepts may be divided into two classes: level tone analyses and contour analyses. The former include, for example, the model of Bruce (1977) and the influential AM model, starting from Pierrehumbert (1980). The contour class, on the other hand, comprises among others the IPO model ('t Hart *et al.*, 1990) and the Kiel Intonation Model (KIM) (Kohler, 1991a,b). Independent of the respective model, however, research efforts were directly or indirectly motivated by the assumption that the intonation codes and transmits a wide spectrum of meanings in languages. This assumption led to a common ground between all models: The phonologically differentiated units are defined with reference to F0 or to the resulting perceived pitch, respectively.

Moreover, it is generally assumed that the transmission of intonation patterns and their corresponding meanings are determined and limited by a further independent coding layer: the segmental string. Segmental determination is reflected in studies investigating the alignment of turning points in the F0 course relative to segmental landmarks as well as in postulations of a temporally rigid anchoring between the turning points and segmental or articulatory landmarks (cf. Arvaniti *et al.*, 1998; Ladd, 2003; Niebuhr and Ambrazaitis, 2006; Mücke *et al.*, 2006). The assumption that intonation patterns are limited by the segmental string is expressed in time pressure phenomena (Caspers and van Heuven, 1993) and in the notion of utterance-final truncation (Grabe, 1998).

However, none of these assumptions hold in a strict understanding. For instance, Hirschberg and Ward (1992) found for American English that meanings, which were exclusively associated with F0 patterns, are accompanied in production by consistent differences in duration, intensity, and voice quality, and that at least the latter is used by hearers in addition to F0 to distinguish the meanings. Also, duration and intensity patterns contribute to the identification of "intonational" meanings in addition to F0 parameters, like peak synchronization, shape, and height, was more recently demonstrated in perception experiments for German by Niebuhr (2003, 2007b). Hence, the respective meanings are coded by *prosodic* units rather than by mere intonation units in the traditional F0-based understanding of the term.

Further, in regards to the separation of intonation patterns (i.e., glottal prosodies) on the one hand and the segmental string (i.e., quick successions of supraglottal articulatory configurations) on the other, it is well documented across languages that phonological segments can be realized as long-term articulatory or phonatory settings superimposed on events at the segmental level. Additionally, these long phonetic components are relevant for the identification of the corresponding morphs (cf. Kohler, 1990; Pierrehumbert and Talkin, 1992; Wesener, 2001; Local, 2003). Contrariwise, there are also indications that local segmental properties are involved in the signaling of meanings associated with intonation units. This concerns the duration or the degree of opening of (accented) vowels (cf. Gartenberg and Panzlaff-

a)Electronic mail: oliver.niebuhr@lpl-aix.fr

 0001-4966/2008/124(2)/1252/12/$23.00

Reuter, 1991, Erickson *et al.* 2004), for example So, as the segmental string is probably much more involved in the coding of the traditional intonational meanings than has been considered so far (e.g., in Pierrehumbert and Talkin, 1992), the necessary step from an F0-based unit to a prosodic unit may not be sufficient.

In this connection, noise signals are particularly worth investigating. In the realm of speech, they occur when the created air stream is forced through a narrow opening in the vocal tract. In this way, the laminal stream becomes turbulent for adequate combinations of the degree of the stricture and the velocity of the air flow (cf. Ladefoged, 2001). The resulting signals may be called friction noises. Within the field of speech, this term refers to global voice qualities like whisper, as well as to more local events like fricatives and aspirations, that are assigned to the level of segmental units. Such friction noises are basically also capable of conveying pitch impressions. Thus, they are also potential intonational elements in a more general perceptual understanding of the term. Pitch impressions evoked by friction noises in speech are called, for example, whisper pitches (Whalen and Xu, 1992) or sibilant pitches (Traunmüller, 1987), depending on the kind of friction noise and on the context in which it is found. However, it is likely that the possible range of pitch variation that can be covered by friction noises is considerably smaller than the F0-based one (cf. Meyer-Eppler, 1957).

Nevertheless, studies dealing with tone (accent) languages agree that communication within such languages is possible in the absence of F0, i.e., in whispered speech, at least in the case of longer utterances (cf. Abramson, 1972). Under these circumstances, changes in F0 that are necessary to disambiguate lexical items are substituted by changes in whispering. However, across different studies and languages a heterogeneous picture emerges as to the acoustic parameters that are modified in whispered speech. Most of the studies report spectral changes, either based on shifts in the lower formant frequencies and/or energies (Meyer-Eppler, 1957; Thomas, 1969; Higashikawa and Minifie, 1999; Kong and Zeng, 2006; Kiefte, 2005; Konno *et al.*, 2006) or on changes in the spectral tilt of the noise (cf. Krull, 2001; Meyer-Eppler, 1957; van Rossum 2005). These spectral changes seem to be directly related to the pitch impression caused by the noise. Other studies point to a systematic relationship between F0 patterns in voiced utterances and duration and intensity patterns in whispered utterances (cf. Hadding-Koch, 1962; Abramson, 1972; Whalen and Xu, 1992; Nicholson and Teig, 2003). It may thus be assumed that durations and intensities of friction noises also convey pitch information [a relation between pitch and intensity was already assumed by Pike (1948)]. However, compared with the spectral variation, the relation between duration, intensity, and pitch information may be an indirect one. For instance, it is not clear from the literature, whether duration and intensity changes are directly perceived as changes in whisper pitch, or whether the pitch information is derived from other perceptual measures (like the strength or the harshness of the friction noise).

Further, it is reported in many of the studies mentioned previously that the acoustic patterns that are directly or indirectly responsible for the different pitch courses identified in whispered speech are also present in normally voiced utterances. That is, the findings generally point to a redundant (or synergetic) coding of pitch patterns and their corresponding meanings, instead of a compensatory use of non-F0 parameters in whispered speech.

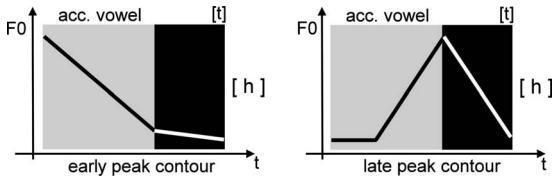


FIG. 1. Schematic representations of characteristic F0 courses for German early and late peak contours, starting from the accented-vowel (acc. vowel. gray box). The black box represents the utterance-final syllable coda ending in aspirated t. As the coda is voiceless in the selected context, the white parts of the F0 contours are expected to be truncated.

In view of the sketched research background, the following general question concerning intonation languages is addressed in the present paper: Do segmental friction noises of utterances with normal voicing patterns contribute to the signaling of meanings that are traditionally associated with F0-based intonation units? It is the aim of the present study to give the first answer to this general question on the basis of the following research subject: /t/ aspiration noises at the end of utterance-final accented syllables, which contain a phonologically short vowel and a voiceless syllable coda.

According to the findings of Grabe (1998), the selected condition should lead to an extensive truncation of the F0 patterns associated with the utterance-final accented syllable. In consequence, the F0 patterns may convey the corresponding meanings less clearly. Therefore, there may be a special need to transfer part of the signaling into the utterance-final aspiration. Thus, in a first approach, those conditions are investigated in which the assumed phenomena are particularly likely to be observed.

For the same reason, two intonation units were contrasted in the selected context that are maximally different with regard to meaning and coding. In KIM they are called early and late peak. As illustrated in Fig. 1, their F0 patterns are, in principle, both marked by an utterance-final terminal F0 fall. In the present context, however, this fall is expected to be truncated to a different degree by the voiceless coda. The structural difference between German early and late peaks is in the preceding F0 movement. The early peak is characterized by a fall into the accented vowel, whereas an F0 rise starting after the accented-vowel onset is an essential feature of the late peak (cf. Niebuhr 2007a,b). In studies of Dombrowski (2003) and Kohler (2005), the two intonation units were found to convey opposed attitudinal meanings, which have been captured by semantic differentials. While terms like "nonquestioning," "certain," and "matter-of-fact" outline the attitudinal meaning profile of early peaks, the profile of late peaks is constituted by meaning shifts in the opposite direction. The early and late peak contours of the KIM are represented by H+L*L−% and L*+HL−% in the German AM framework (GToBI, Grice and Baumann, 2000).

On the basis of empirical findings combined with a theoretical framework for the perceptual construction of speech melody, Niebuhr (2007a) developed coding profiles for the German early and late peaks. Parallel to the meaning relations, he postulates that early and late peaks are marked by opposite coding profiles. In particular, while the early peak is perceptually based on a salient low pitch impression, the late peak emphasizes high pitch more strongly. So, provided that the utterance-final /t/ aspiration is, in fact, involved in the signaling of the attitudinal meanings of German early and late peaks, their diametrically opposed coding should make it easier to detect differences in the acoustic manifestations of the investigated aspirations. Further, it should be easier on this basis to obtain a clearly structured judgement behavior for stimuli in perception experiments.

Thus, if and in which way the final /t/ aspiration noise is involved in the coding of early and late peaks in the selected segmental and prosodic context is investigated by a combination of acoustic analysis and perception experiment. The results of the former will guide the manipulations for the stimuli in the perception experiment, which is the decisive part of this study.

II. ACOUSTIC ANALYSIS

A. Method of the acoustic analysis

For the acoustic analysis a speech corpus of Standard German was used. It is based on 24 short utterances. Most of them consist of two words: subject and verb. The utterances were read by a trained male phonetician (kk) in his 50s, who produced them with the terminal-falling F0 peak categories of the KIM on the accented syllable of the verb. This does not only include early and late, but also medial peaks (which show an alignment in between the ones of the early and late categories, cf. Kohler 1991a,b). The KIM-based research is founded on meaning. It represents the starting point for the postulation of intonational categories, and it serves as the point of reference in experimental investigations (cf. Kohler, 1991a, b; Niebuhr, 2007a, b). Due to this central position of meaning, the aim of the speaker (kk) to produce particular intonation categories corresponds to the aim to produce particular (attitudinal) meanings in the sentences. The 24 sentences were repeated 10 times for the medial peak, and 5 times for the early and late peaks. So, the whole speech corpus comprises 480 utterances; 240 medial peak tokens as well as 120 early and late peak tokens.

The corpus was recorded almost 20 years ago to investigate effects of different syllable structures and numbers on the F0 peak contours differentiated in the KIM. The findings were meant to be applied in speech synthesis (cf. Gartenberg and Panzlaff-Reuter, 1991). It follows from this that the speaker (kk) did not know the research question addressed in the present study. Moreover, even if the speaker (kk) had been aware of the original research question of Gartenberg and Panzlaff–Reuter during the recording, it would have drawn his main attention to the intonation (i.e., to F0) and not to the underlying segmental string. Hence, it is very unlikely that the segments, including the utterance-final aspirations, were *consciously* influenced by the speaker in view of the intonation categories. However, the latter aspect and the meaning-based elicitation of the intonation categories are not the only arguments in favor of the selected speech corpus. It is also the only known high-quality recording in which the German peak categories can be contrastively investigated under otherwise comparable conditions within the relevant (utterance-final) syllable structures.

Regarding the latter, the following two-word utterances were selected from the 120 early and late peak tokens. They start with the personal pronoun "*Sie*" ("she/it," /zi:/), followed by a verb consisting of a single accented syllable with the short vowel /ɪ/ in the nucleus and a voiceless consonant cluster in the coda, with /t/ as the final element. The plosive /t/ was either preceded by another (unreleased velar) plosive /k/ or by a palatal fricative /ç/. By the restriction to the closed front vowel quality /ɪ/, co-articulatory effects on the following final consonant cluster and hence on the acoustic manifestation of the /t/ aspiration were avoided. The syllable onset was also marked by voiceless obstruents. The selected utterances were (1) "*Sie spricht*" ("she talks"), (2) "*Sie strickt*" ("she knits"), (3) "*Sie schickt*" ("she sends"), (4) "*Sie schrickt*" ("it gives her a start"), and (5) "*Sie tickt*" ("it ticks"). As five repetitions of the five selected two-word sequences were produced with both peak contours, the early and the late, 50 utterances were analyzed in total, 25 within each of the two intonation categories.

The acoustic analysis should consider parameters that were identified as substitutes for F0 in whispered speech (cf. Sec. I). Moreover, the amount of truncation of the early and late peak contours has to be determined. Therefore, the following five measurements were taken for each of the 50 utterances: (a) the F0 value at the end of voicing, (b) the duration of the aspiration noise, (c) the intensity (i.e., the short-term energy) maximum in the aspiration noise, (d) the lower spectral energy boundary (SEB), i.e., the frequency in the fricative spectrum at which the first clear increase in energy is observed, as well as (e) the frequency with the highest spectral energy above the SEB.

In addition to the relevant effects caused by the production process itself, the absolute intensity measurements (c) can also be influenced by the distance between the microphone and the mouth of the speaker. However, while it is well known that accented-syllable productions themselves may be accompanied by head movements (cf. Krahmer and Swerts, 2006; Scarborough *et al.*, 2006) and hence affect the speaker's microphone distance, the intonation categories associated with the accented syllable do not seem to cause any further specific head movement patterns. Rather, the production of intonation categories was found to be synchronized with manual gestures or facial gestures, such as eyebrow movements (cf. Cavé *et al.*, 1996; McClave, 1998). In any case, the speaker (kk) is a trained phonetician and experienced in recording speech. This includes the general suppression of body movements when speaking. So, the distance between the microphone and the mouth of the speaker should have been comparable for the utterances produced with early and late peaks. If at all, it might have shown small random variations. Further, it was shown by Niebuhr (2006, 2007a) that the intensity course in the accented syllable is involved

TABLE I. Results of the statistic analyses of the measurements (a)–(e) yielded for the selected utterances; Means and standard deviations (sd) are given at the top. Below, critical values (F,t) of F tests and of t tests for independent samples are given, together with the probability of α errors (p); stars indicate the significance level; sample size $n=25$. The tests compare the means of variables (a)–(e) yielded for early and late peak conditions. For significant F tests, critical values and degrees-of-freedom (dof) in the corresponding t tests were adjusted. Values in parentheses refer to the second measurements of (d) and (e) shortly before the aspiration offset.

		(a) final F0 (Hz)	(b) dur asp (ms)	(c) Imax asp (dB)	(d) SEB (Hz)	(e) Emax spec (Hz)
Early	Mean	74.6	96.2	58.5	1801 (1757)	3611 (3457)
peak	sd	8.1	9.8	2.2	53.8 (47.6)	284.7 (108.6)
Late	Mean	127.5	106.2	54.8	1803 (1773)	4050 (3663)
peak	sd	6.3	19.7	2.5	48.6 (43.9)	413.6 (350.1)
F test	F	1.6767	0.2713	0.7864	1.2242 (1.1788)	0.4737 (0.0962)
	p	0.1064	0.0011	0.2803	0.3121 (0.3451)	0.0366 (<0.0001)
t test	t	−25.875	−2.321	5.5950	−0.1380 (−1.2360)	−4.3713 (−2.8100)
	dof	48	36	48	48 (48)	43 (29)
	p	<0.0001***	0.0261*	<0.0001***	0.8908 (0.2225)	<0.0001*** (0.0088**)

in the signaling of early and late peaks. From this point of view, it would have been problematic to use measurements that relate the intensity maximum of the aspiration noise to immediately preceding events like the intensity maximum of the accented vowel in order to compensate for potential variations in the speaker–microphone distance. These relative measurements are likely to yield systematic differences between the peak categories, which do not stem from the aspiration, but from the preceding vowel. For these reasons, absolute measurements were preferred. The spectral frequency measures (d) and (e) have characteristic values for fricatives at different places of articulation, such as /f/ versus /s/ (cf. Strevens, 1960; Ladefoged, 2001), which are also associated with different perceptual qualities and pitch impressions (cf. Laver, 1994). It was therefore assumed that the two measures (d) and (e) are suitable to represent differences in pitch impression between the aspiration noises. However, it had to be considered that the two measures only provide snapshots of the whole aspiration noise. Therefore, two measurements were taken for (d) and (e): The first one approximately 20 ms after the aspiration onset, and the second one approximately 20 ms before the aspiration offset.

Measurements (a)–(c) were taken in PRAAT with the default analysis settings (www.praat.org). For (d) and (e), the XASSP software package was used (cf. www.ipds.uni-kiel.de/forschung/xassp.en.html), and the measurements were taken by means of a spectral section window (corresponds to "spectral slice" in PRAAT) based on discrete Fourier transformation (DFT) analyses averaged over 50 ms of the signal (i.e., over 25 ms to both sides of the cursor) to compensate for the high variability of noise signals.

B. Results of the acoustic analysis

The measurements taken in the selected utterances were subjected to descriptive and inferential statistic analyses. Results are summarized in Table I. Regarding the inferential statistics, t tests for independent samples were performed. They compared for measurements (a)–(e), whether the 25 values from the early and the late peak conditions differ significantly. Prior to the t tests, F tests were used to check whether the compared samples have significantly heterogeneous variances. In that case, the critical values and degrees-of-freedom in the t tests were adjusted accordingly.

Further, the mean frequency values of the lower SEB (d) and of the highest spectral energy above the SEB (e) for each of the two-word sequences are illustrated in Fig. 2. The measurements taken around 20 ms after aspiration onset and before aspiration offset (the latter are given in the brackets of Table I) are connected by linear interpolation to capture the changes of the two spectral frequency measures across the aspiration.

Table I shows for (a) that the F0 values found at the end of voicing in the selected utterances (i.e., at the accented-vowel offsets) differ highly significantly for early and late peaks. In the case of early peaks, F0 fell to a mean value of 74.6 Hz. This comes close to the terminal F0 level of the speaker (kk), which is usually reached slightly below 70 Hz (e.g., in other utterances of the same speech corpus with voiced segments in final position). In the late peak condition, however, F0 ends on average at almost 130 Hz. That is, the terminal fall is extensively cut off by the voiceless obstruents in the coda of the accented syllable.

In addition to this basic finding, the results further reveal that the /t/ aspirations in the early and the late peak contexts show multidimensional differences. For instance, the aspiration noises produced after late peaks are significantly longer than the ones following early peaks (measurement b in Table I). The mean values diverge by 10 ms, but individual comparisons showed that the /t/ aspirations in the late peak context can be more than twice as long as the ones in the early peak context. This is also reflected in the standard deviations for the duration measurements in Table I. Moreover, the mean intensity maximum found in /t/ aspirations after early peaks is considerably higher than the one after late peaks. This difference, which amounts to almost 4 dB (cf. measurement c in Table I), is highly significant.

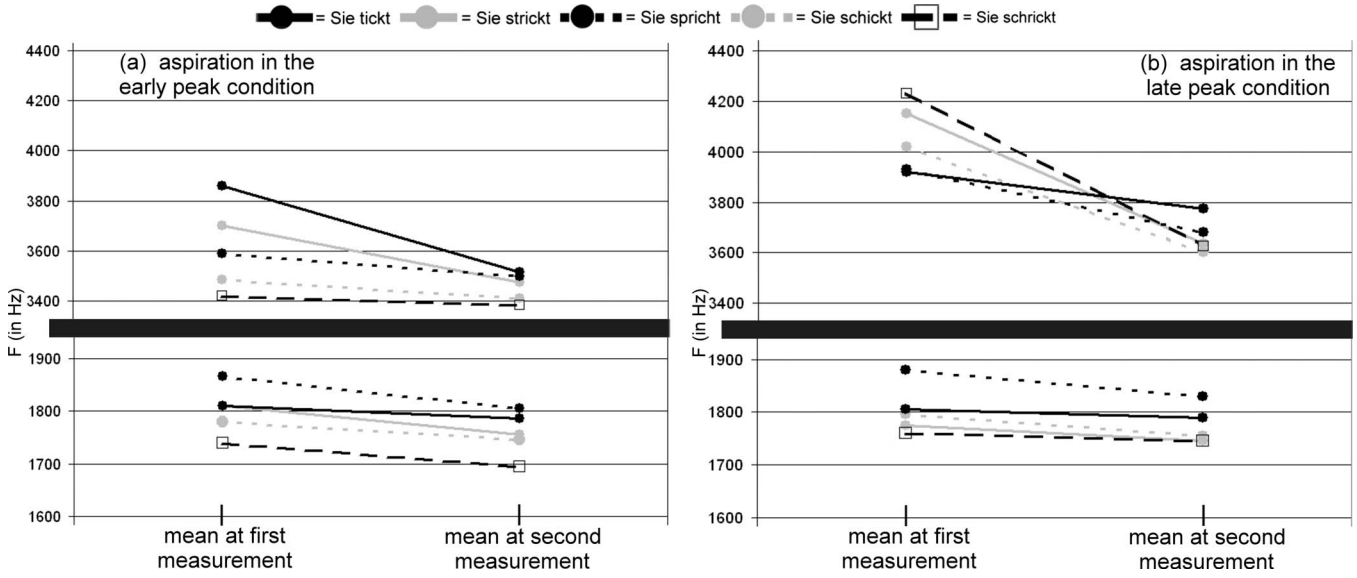


FIG. 2. Mean values of the SEB (measure d, bottom) and of the frequency with the highest spectral energy above the SEB (measure e, top) for the final aspiration noises in the early peak condition (left, a) and in the late peak condition (right, b). Measurements for (d) and (e) were taken at two points in time, about 20 ms after/before aspiration onset/offset.

Regarding the spectral measures (d) and (e), significant differences between the aspiration noises in the early and the late peak conditions only showed up for the frequency with the highest spectral energy above the SEB (i.e., for e) and not for the SEB measurements (d) themselves. As can be seen from Table I and Fig. 2, the highest spectral energy is, on average, located at higher frequency values in the aspiration noises after late peaks, compared with the aspiration noises following early peak productions. This holds for the measurements taken at both points in time, i.e., about 20 ms after aspiration onset and before aspiration offset. Moreover, the mean value for the first measurement after the aspiration onset is generally higher than the one immediately before the aspiration offset. That is, the frequency with the highest energy above the SEB decreases across the aspiration. Figure 2 shows that this down-drift is also present at the level of each of the individual five two-word sequences and holds for the SEB measurements, too.

Figure 2 reveals two further aspects: First, the downdrift of the frequency with the highest spectral energy is clearly stronger in the aspirations of the late peak condition. In particular, it starts from a much higher value than in the early peak condition (cf. also Table I). Second, there are also differences between the individual two-word sequences, within, as well as across, the early and late peak conditions. In a cross-conditional comparison, e.g., Fig. 2, shows very pronounced differences between the aspiration spectra in early and late peak contexts for the utterance *Sie schrickt*. The mean frequency of the highest spectral energy above the SEB is even higher toward the end of the aspiration in late peak contexts than at the beginning of the aspiration in early peak contexts. Moreover, there are also slight differences in the mean SEB values in connection with *Sie schrickt*. They are higher in late than in early peak contexts. Compared with *Sie schrickt*, the spectral differences are less pronounced in *Sie tickt*. Finally, compared with the spectral measures in general, the intensity maxima of the aspirations in early and late peak contexts show less variation between the individual utterances in the two intonational contexts.

C. Discussion of the acoustic analysis

First, it is consistent with previous findings by Grabe (1998) and hence also with the expectation shown in Fig. 1 that the measurements of the utterance-final F0 values revealed an extensive truncation of the terminal fall for late peak productions, whereas a smaller, but still substantial truncation was found in the case of early peaks. Further, the acoustic analysis revealed that the /t/ aspiration noises at the end the short two-word utterances are systematically related to the two intonation conditions, i.e., to the productions of early and late peak contours. Aspiration noises following early peak productions are significantly shorter, more intensive and have significantly lower frequency values for the spectral energy maximum above the SEB, compared with the aspiration noises preceded by late peak productions.

Hence, the two aspiration noises differ in all acoustic parameters that were found in previous studies to convey pitch information in friction noises of speech: duration, intensity, and spectral energy distribution. Shorter durations and higher intensity levels were used by speakers in whispered speech to substitute F0 courses, which emphasize low pitch; and listeners interpreted the changes in noise parameters accordingly (e.g., Hadding-Koch, 1962; Abramson, 1972; Krull, 2001). In the case of spectral changes, speakers transfer noise energy into lower spectral regions in order to replace low-pitched F0 courses (cf. Meyer-Eppler, 1957; Kiefte, 2005; Konno *et al.*, 2006; van Rossum, 2005). Also this strategy was shown to yield matching perceptual effects, i.e., it results in lower pitch impressions (cf. Thomas, 1969; Higashikawa and Minifie, 1999).

According to the empirically founded coding schemes of Niebuhr (2007a), highlighting low pitch is essential for the identification of the attitudinal meaning of the early peak,

whereas the late peak involves emphasis on high pitch. This goes well with the different acoustic configurations of the aspiration noises and their known effects on perceived pitch. Thus, it seems that the utterance-final aspirations are in fact modified to contribute to the coding of the attitudinal meanings associated with the two peak categories.

On the other hand, the acoustic analysis is just based on the productions of a single speaker. This raises doubts whether the results can be generalized. While this concern is justified, it is of minor importance at this stage of the investigation. In order to give the first general answer to the research question raised in the introduction, it is sufficient to check whether there *can*, in principle, be systematic differences between the aspirations in the two peak contexts. Moreover, the decisive evidence for the research question must come from an additional perception experiment. That is, it is essential to examine whether listeners actually project the differences found in the acoustic analysis onto the attitudinal meaning profiles of the early and late peaks. Without these perceptual effects, the acoustic differences cannot be interpreted in view of the research question.

In addition, the question as to what extent the results of the acoustic analysis can be generalized is not only a matter of the number of subjects included. Among others, the following aspects also need to be taken into account: Do the intonational categories influence /p/ and /k/ aspirations in the same way as /t/ aspirations? What happens in the case of phrase-final accented syllables that contain more clearly pronounced F0 cues to due, for example, phonologically long vowels and/or sonorant consonants in the syllable coda? What happens in high-frequency and function words, which are usually produced with less effort? Is the (assumed) involvement of the aspiration noise in the coding of attitudinal meanings a general phenomenon or tied to stylistic, sociophonetic, or gender factors? These aspects illustrate that the question of the generalization of the acoustic findings is closely intertwined with a more detailed investigation of the formulated research question. This goes beyond the aim of the present study. Therefore, the question of generalization should be dealt with in separate follow-up studies, after it has been shown in a perception experiment that a link exists between friction noises in contexts of intonational categories and the attitudinal meanings associated with these categories.

III. PERCEPTION EXPERIMENT

A. Method of the perception experiment

1. Stimulus generation

It was the aim of the perception experiment to examine whether the acoustic differences that were found for /t/ aspiration noises at the end of utterance-final accented syllables in early and late peak contexts lead to attitudinal meaning changes that go with the meaning profiles established for the two peak categories. Correspondingly, the stimuli of the perception experiment were based on the same speech corpus (of the speaker kk) that was subjected to the acoustic analysis. Specifically, one of the repetitions of the utterance *Sie tickt* (it ticks) was selected as the point of departure for the stimulus generation. Compared with the other utterances, *Sie tickt* is most likely to occur in this isolated form. The selected utterance was marked by a medial peak on the only accented syllable *tickt* (ticks). A "neutral" medial peak token was to exclude the possibility that any cues for the early or the late peak category are contained in the segmental string (e.g., in the duration structure, cf. Niebuhr 2007a), which could have biased the results of the perception experiment.

In the first step of the stimulus generation, the final /t/ aspiration was cut off. Then, the F0 contour of the resulting base stimulus was flattened in PRAAT around a value of 100 Hz. However, most of the microprosodic perturbations in the F0 contour were maintained in order to prevent the utterance from sounding metallic. Further, the F0 level of the syllable *tickt* was set about 2 Hz above the level of *Sie* to compensate for pitch differences resulting from the vowel qualities in the two syllables Sie and tickt, which differ slightly in height [/i/ and /ɪ/ (cf. Niebuhr, 2004; Rossing and Houtsma, 1986)]. The F0 changes within the base stimulus fall below the just noticeable difference estimated in various studies ('t Hart, 1981; Mack and Gold, 1986). Accordingly, the two syllables of the base stimulus were perceived with a constant pitch. In consequence, the early and late-peak meanings may be regarded as being removed from the base stimulus, as they are not conveyed by certain pitch registers, but by pitch changes [irrespective of the question whether they are analyzed as contours or not (cf. Kohler 1991a, b; Pierrehumbert 1980)]. Also the semantics of the two-word utterance *Sie tickt* is neutral with regard to the attitudinal meanings of early and late peaks.

As a second step, two aspiration noises had to be selected from the investigated speech corpus, one from the context of each of the two peak categories. The two aspiration noises were to replace the original aspiration of the base stimulus, which was cut off in the preceding step. In this way, two (different) base stimuli were created. The aspirations were to diverge as much as possible with regard to the different acoustic configurations found in connection of early and late peaks. However, differences in the frequency of the highest spectral energy above the SEB had to be given priority over differences in duration and intensity, as, unlike the latter two, this parameter could not be manipulated with existing speech processing software. Therefore, two aspirations were selected that were marked by the greatest difference in the frequency with the highest spectral energy above the SEB (and that also showed substantial parallel differences in the SEB values), even though they only differed marginally in duration and intensity. The values of the latter two parameters were manipulated separately. Both aspiration noises attached to the base stimulus come from *Sie schickt*.

In the following step, the two base stimuli were further manipulated in COOL EDIT (cf. www.cooledit.com). For each stimulus, the complete aspiration was marked in the signal window and its overall energy was increased and decreased by 6 dB (i.e., it was doubled or halved, respectively). In this way, the number of base stimuli was raised to 4. As the manipulation started from different naturally produced aspirations, the increase as well as the decrease of overall energy yielded slightly different values for the two aspiration noises. In terms of the intensity maxima measured in PRAAT, these

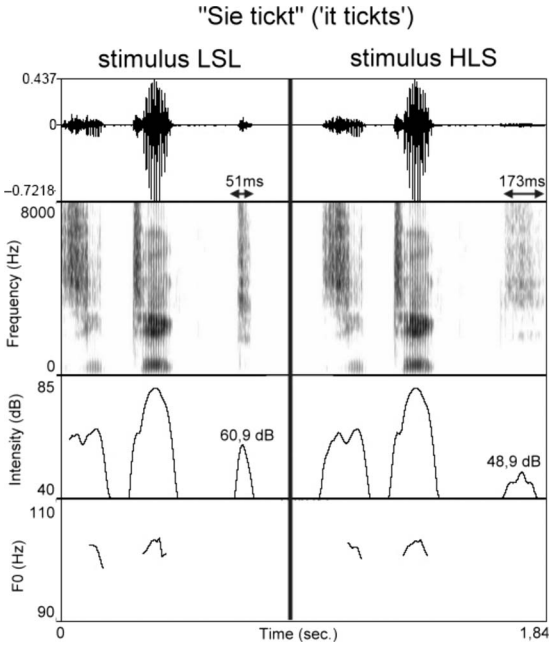


FIG. 3. Acoustic analyses of the two stimuli LSL and HLS generated on the basis of the utterance "Sie tickt," including oscillogram (top), spectrogram, intensity course, and F0 course (bottom). For the final aspiration noises, measurements of durations and intensity maxima are given.

differences amounted to about 1 dB, with values of 61.9 and 60.9 dB for the increased and of 49.9 and 48.9 dB for the decreased condition.

In the final step, each of the four base stimuli was subjected to a duration manipulation in PRAAT. The duration of each aspiration was doubled and halved. As in the preceding energy manipulation, this yielded different results for each pair of base stimuli: The shortened durations either amounted to 51 or 58 ms. The extended aspirations had durations of 153 or 173 ms, respectively. The latter is not a slight difference; however, it is considerably smaller than the decisive differences between the long and short duration conditions. The generated energy and duration values are around the limits found in the acoustic analysis.

In total, the stimulus generation yielded eight stimuli. These eight stimuli were completely identical except for the final aspiration noises. The latter differ in three two-level variables: (1) The frequency with the highest spectral energy above the SEB (high versus low), (2) duration (long versus short), and (3) overall energy or intensity maximum (loud versus soft). For variable (3), perceptual terms will be used to avoid confusions with the difference "high versus low" in (1). However, only four stimuli were used in the perception experiment reported in this paper. In view of the research question, they are either marked by a long and soft or by a short and loud aspiration noise, each combined with a low or a high frequency of the highest spectral energy above the SEB. So, as (2) and (3) were linked, the present experiment is based on two experimental variables. The stimuli are referred to as LLS (low, long, soft), LSL (low, short, loud), HLS (high, long, soft), and HSL (high, short, loud). Figure 3 presents acoustic analyses of the two stimuli LSL and HLS, which resemble the aspiration properties found in the natural productions of the aspiration noises in the early and the late peak conditions.

2. Experimental setup

The four stimuli were used in a perception experiment based on the semantic differential paradigm. It comprised the scales that diverged for the attitudinal meanings of German early and late peaks in the perception experiments of Kohler (2005) and Dombrowski (2003). That is, the scales yielded (statistically significant) average values for early and late peaks that differed by at least one scale point. Based on this criterion, the following five scales were selected: (1) *fragend/nicht fragend* ("questioning/nonquestioning"), (2) *sicher/unsicher* ("certain/uncertain"), (3) *abschließend/weiterweisend* ("concluding/continuing"), (4) *sachlich/erstaunt* ("matter-of-fact/surprised"), and (5) *akzeptierend/nicht akzeptierend* ("accepting/not accepting"). Table II summarizes how these word pairs are related to the early and late peak category.

Scales (1) and (2) were adopted from Dombrowski (2003); scales (3)–(5) come from Kohler (2005). Dombrowski (2003) also used a scale that is similar to (4), viz. *sachlich/ironisch* ("matter-of-fact/ironic"), and the judgements on this scale also differed by more than one point for early and late peaks. However, as *erstaunt* ("surprised") was regarded to be a more suitable counterpart of *sachlich* ("*matter-of-fact*") than *ironisch* ("ironic"), the scale of Kohler (2005) was preferred. For the same reason, scale (5) was modified from *akzeptiert/kontrastiv* ("accepted/contrastive") in Kohler (2005) to *akzeptierend/nicht akzeptierend*. Moreover, considering that attitudinal meanings like the ones conveyed by early and late peaks refer to current processes rather than to fixed states, present participle verb forms were used.

As it was expected that judging the stimuli on the five scales would be quite an unusual and hard task for the subjects, the experimental paradigm was marked by three major modifications, compared with the ones used in Dombrowski (2003), Ambrazaitis (2005), and Kohler (2005). The modifications aimed at facilitating the task of the subjects and at increasing the sensitivity of the test procedure to detect meaning changes in the stimulus utterances. First, *pairs* of stimuli were presented, and the subjects judged the meaning change of the second stimulus relative to the first one on the corresponding scale. The meaning of the first stimulus was represented at the center of the scale. Second, *continuous* instead of ordinal scales were used. That is, the strength of the perceived meaning change could be indicated by a cross

TABLE II. Profiles of the attitudinal meanings of German early and late peaks based on differences on semantic scales found by Dombrowski (2003) and Kohler (2005).

Early peak	Concluding	Accepting	Matter-of-fact	Nonquestioning	Certain
Late peak	Continuing	Nonaccepting	Surprised	Questioning	Uncertain

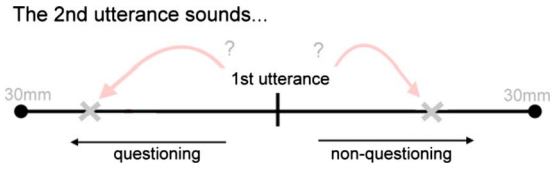


FIG. 4. (Color online) Modified example of a response display presented to the subjects in the perception experiment. Gray symbols and texts were added and annotations are translated into English. In total, a scale was 60 mm long, 30 mm in each meaning direction.

at any point on the scale or at any distance from the center of the scale, respectively. According to Chen (2004) continuous scales are superior to ordinal ones in revealing (meaning) differences between stimuli. The method is illustrated in Fig. 4. Finally, all stimulus pairs were presented *three times* to the subjects. However, if it was necessary, each pair was repeated until all subjects responded.

The experiment contained 6 stimulus pairings in both orders (A–B and B–A), i.e., 12 stimulus pairs, resulting from a complete cross combination of the 4 stimuli. Identical pairs (e.g., A–A or B–B) were not included. Regarding the latter, it is important to see that, although pairs of stimuli were presented to the subjects, the task itself is not about perceptual discrimination, but about relative meaning association. To this end, it is a prerequisite that the subjects were able to discriminate the stimuli of a pair. Accordingly, the subjects were explicitly informed in the instructions that all pairs will consist of different stimuli. If the subjects were not able to discriminate the stimulus pairs, the latter would not be judged systematically (in the two orders as well as across the subjects). So, the results of the perception experiment will show indirectly whether the subjects were able to discriminate the pairs. Moreover, in view of the aim of the experiment, it is of minor importance to check if the subjects actually marked stronger meaning differences as stronger on the continuous scale. This could be done by comparing the judgements for identical pairs with the remaining ones. At the present stage, it is more important to see whether they do judge the stimuli systematically, and if so, in which direction on the scales. For these reasons, it was decided to exclude identical pairs. This also reduced the duration of the experiment and hence eased the task of the subjects.

3. Subjects and procedure

As each of the 12 stimulus pairs had to be presented for five scales, 60 pairs were played in a randomized order. A complete session took about 50 min. In total, 25 subjects participated in the experiment (10 male, 15 female, between the ages of 21 and 40). They were all native speakers of German with normal hearing. At the beginning of the experimental session, they listened to examples of *Sie tickt* produced naturally by the author (on) with early and late peak contours. However, all final aspiration noises were substituted for a constant one. In this context, the attitudinal meanings of the peak categories were explained to them. Then, they were told that they would hear further productions of *Sie tickt*, arranged in heterogeneous pairs and produced by a male speaker. However, unlike the preceding examples, the

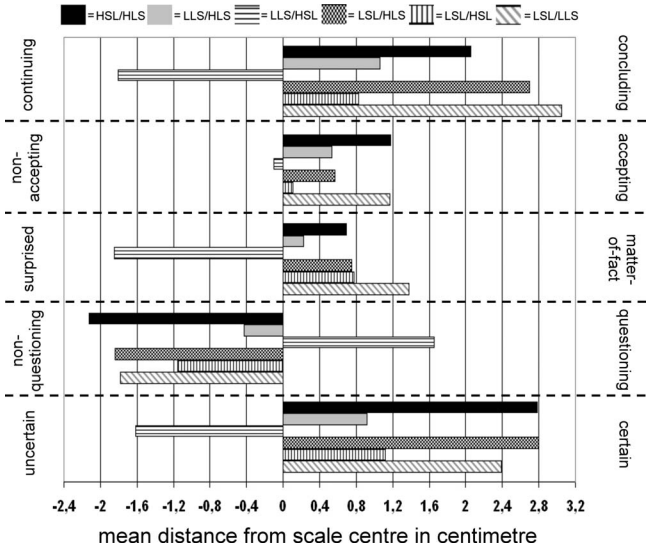


FIG. 5. Mean distances of the crosses from the centers of the five scales, based on the sums of measurements formed for the two orders of a stimulus pairing across the 25 subjects.

utterances in each pair had identical flat pitch courses and differed only by the pronunciation of the final "*t*." With this in mind, the subjects were instructed to judge for each pair direction and strength of the meaning change of the second utterance in relation to the first one by making crosses at an appropriate point of a scale. The subjects were familiarized with the scales, and they received answer sheets with 65 displays like the one illustrated in Fig. 4 (except for the gray symbols and texts). The subjects listened to the pairs via loudspeaker in a sound-treated room. Prior to the 60 experimental pairs, 5 pairs, one for each scale, were judged by the subjects to familiarize them with the task.

B. Results of the perception experiment

For the 60 scales judged by each of the 25 subjects, the distance between cross and scale center was measured manually in millimeters. For pairs judged in the order A–B, crosses placed on the left-hand side of the scale received positive values and crosses on the right-hand side received negative values. The signs were inverted in the order B–A. The results of this procedure are illustrated in Fig. 5. It shows for the 6 stimulus pairings of the perception experiment the mean distances of the crosses from the centers on the five scales. A stimulus pairing represents the combination of two stimuli in both orders (A–B and B–A). That is, the bars in Fig. 5 result from two steps: First, for each of the subjects and scales, the two distance values measured for a stimulus pairing were added up (yielding sums of, at most, ± 6 cm, $2 \times \pm 3$ cm, cf. Fig. 4). Then, the mean values were calculated from the sums of the 25 subjects.

In addition to the descriptive analysis, the measurements were subjected to an inferential statistic analysis. To this end, 30 paired samples were formed, one of each combination of stimulus pairing and scale (i.e., $6 \times 5 = 30$). The distance values measured for each of the two orders of a stimulus pairing across the 25 subjects represent one sample. They were compared by t tests for paired samples. Results are given in Table

TABLE III. Results of the 30 t tests for paired samples, comparing for each combination of stimulus pairing (columns) and scale (rows, $6 \times 5 = 30$) the distance values measured for each of the two orders of a stimulus pairing across the 25 subjects (i.e., $n = 25$). For each test, t values (in parentheses) and p levels are given. Asterisks indicate the significant outcomes, and italics point to significant trends. Both are based on Bonferroni corrections to account for multiple testing (i.e., α significance level p[0.017).

	LSL/LLS	HSL/HLS	LSL/HLS	LLS/HSL	LSL/HSL	LLS/HLS
Concluding/continuing	(−6.836)	(−4.981)	(−5.868)	(1.834)	(−1.808)	(−2.737)
	<0.0001*	<0.0001*	<0.0001*	0.0791	0.0832	0.0114*
Accepting/non accepting	(−2.418)	(−3.143)	(−2.508)	(0.406)	(−0.975)	(−2.444)
	0.0235	0.0044*	*0.0193*	0.6885	0.3395	*0.0222*
Matter-of-fact/surprised	(−2.542)	(−0.697)	(−1.772)	(3.633)	(−3.532)	(0.643)
	0.0178	0.4494	0.0890	0.0013*	0.0017*	0.5261
Questioning/nonquestioning	(3.277)	(3.372)	(3.819)	(−3.528)	(2.439)	(1.458)
	0.0032*	0.0025*	0.0008*	0.0017*	*0.0225*	0.1578
Certain/uncertain	(−6.274)	(−6.274)	(−5.969)	(3.373)	(−1.691)	(−1.573)
	<0.0001*	<0.0001*	<0.0001*	0.0025*	0.1038	0.0924

III. With regard to Fig. 5, t tests with a statistically significant outcome indicate that the corresponding bar represents a relevant deflection in a particular meaning direction, which juts out from the statistic background noise (e.g., due to order effects and guessing subjects). With regard to the multiple testing, statistical significances are based on Bonferroni corrections. As this procedure is quite conservative, the following explanations will also consider significant trends.

The descriptive and inferential analyses of the data clearly reveal that the 6 stimulus pairings were judged differently on the five scales. These differences are mainly found for stimulus pairs, whose utterance-final aspiration noises are distinguished by duration and overall energy patterns and which were judged on scales other than *akzeptierend/nicht akzeptierend* (accepting/nonaccepting).

In the case of the two stimulus pairings differing only in the duration and the overall energy of the final aspiration, i.e., LSL/LLS and HSL/HLS, the stimuli with the short and loud aspirations (LSL and HSL) significantly changed the meaning of the stimulus utterance toward concluding and certain, compared with the stimuli ending in long and soft aspirations (LLS and HLS). Additionally, the LSL and HSL stimuli sounded more nonquestioning and accepting (the latter is a significant trend for LSL/LLS). Finally, there is another significant trend indicating that the LSL stimuli were more strongly associated with matter-of-fact compared with the LLS ones.

These findings hold in a comparable way for the two stimulus pairings, whose aspirations differ not only in duration and overall energy, but also in the high–low dimension of the frequency with the highest spectral energy above the SEB, i.e., LSL/HLS and LLS/HSL. Regarding the pairing LSL/HLS (which resembles the aspiration differences found in connection with early and late peaks in the acoustic analysis), the low, short, and loud configuration changed the meaning of the stimulus utterance significantly toward concluding, nonquestioning, certain, and accepting (the latter is a significant trend), compared with the high, long, and soft one. In the case of LLS versus HSL (which combine the duration and energy patterns found for the aspirations in the early and late peak conditions with the inverted spectral energy difference), only two of the scales reached significance. In these, the stimuli with the (high) short and loud aspiration sounded more nonquestioning and certain than the stimuli with the (low) long and soft aspiration. The latter, however, sounded significantly more surprised to the subjects.

The two stimulus pairings that contrast aspiration noises with high and low spectral energy maxima above the SEB, LSL/HSL, and LLS/HLS, yielded only two significant results. Compared with the (high) HSL stimuli, the meaning of the (low) LSL stimuli was changed toward matter-of-fact. Moreover, the (low) LLS stimuli were judged to be more concluding than the (high) HLS ones. However, there are two further significant trends, revealing that stimuli with (low) aspirations (LSL and LLS) are more strongly associated with nonquestioning and accepting than the stimuli that end in high aspirations (HSL and HLS).

C. Discussion of the perception experiment

Comparing Table II with the findings from the perception experiment represented by Fig. 5 and Table III clearly reveals a very close correspondence. That is, the attitudinal meaning profiles for stimuli containing the (fully realized) intonation patterns of early and late peaks were largely replicated by stimuli which differed solely in the acoustic configuration of the final /t/ aspiration noise by combinations of acoustic patterns found in natural contexts of early and late peak intonations. Moreover, apart from the final /t/ variation, the stimuli contained no other potential phonetic or semantic cues that could have biased or changed the attitudinal meanings of the stimuli (toward the profiles of early or late peaks).

The match between the findings of the present perception experiment and the ones of Dombrowski (2003) and Kohler (2005) is not perfect, as not all scales yielded statistically significant effects for all pairs of stimuli (which cross combined aspiration properties found in the two intonation peak contexts). However, all significant effects and trends found for meaning differences within the stimulus pairs are in the direction expected from the profiles for early and late peaks given in Table II. There is no conflicting evidence. For instance, the short and loud aspirations, which were found to accompany early peak intonations in the preceding acoustic analysis, yielded only significant effects (or trends) toward

the early-peak meanings concluding, accepting, matter-of-fact, nonquestioning, and certain, compared with the long and soft aspirations, which characterized late peak productions. Likewise, in comparisons of aspiration noises with high and low spectral energy maxima, the latter, marking the naturally produced aspiration after early peak intonations, were also associated with the meanings concluding, accepting, matter-of-fact, nonquestioning, and certain.

In summary, the perception experiment showed that the systematic acoustic and perceptual differences found for utterance-final /t/ aspirations in naturally produced early and late peak contexts can be related to systematic meaning differences. Further, the latter refer to the attitudinal meanings of the early and late peak contours.

Two further aspects of the results need to be addressed. First, with regard to Fig. 5 and to Table III, the scale accepting/nonaccepting did not yield as many significant effects as the remaining four scales. This is probably due to the semantic concepts represented by the word pair. In view of the remarks of some subjects after the experimental session, these concepts were not clear and intuitive enough and hence difficult to judge. Thus, subsequent studies should search for alternative word pairs. Apart from that, however, the method of the present experiment, which applied the semantic scales to stimulus pairs and which used continuous scales to judge them, was successful. This also means that each of the 6 pairs of stimuli yielded at least one statistically relevant effect. This shows that the subjects were in principle able to judge all pairs systematically.

Second, the perception experiment was based on the duration and intensity pattern as well as on the spectral energy distribution of the aspiration noise. The results consistently show that the duration and intensity pattern was more influential than the spectral energy distribution. On the one hand, this is mirrored in the number of significant effects. Aspiration noises differing in the duration and intensity pattern at constant spectral envelope yielded seven significant effects, but there are only two significant effects for aspiration noises that differed solely in their spectral envelopes. Further, the fact that judgements were primarily guided by the duration and intensity pattern can also be seen in the case of conflicting cues. For example, the short and loud aspiration changed the meaning of the stimulus utterance toward the early peak profile, even in connection with a high-pitched aspiration noise (cf. LLS/HSL). The superiority of the duration and intensity pattern may be interpreted in terms of a cue hierarchy. This would also go well with the observation from the acoustic analysis that the differences in intensity were less variable than the spectral differences. However, it is also possible that the (artificial) changes in the duration and intensity pattern were simply more extreme than the (natural) difference in the spectral energy distribution and therefore masked the relevance of the latter cue. Future studies should clarify this point. Moreover, the effects of the utterance-final /t/ aspirations on the attitudinal meanings of the stimuli showed up for constant F0 courses. Thus, further studies should investigate, how influential the aspiration cues will be if they are combined with (parallel and conflicting) F0 cues. It may be expected that F0 cues are in principle the stronger ones. On the other hand, in everyday communication these cues may be less salient and reliable, for example, as the intonation course contains mircoprosodic perturbations due to consonant articulations, and it has to account for multiple meaning dimensions simultaneously.

IV. GENERAL DISCUSSION

An acoustic analysis was performed on the basis of utterances which were produced by a naïve trained phonetician, who intended to convey the attitudinal meanings of German intonation categories like the early and late peak. The acoustic analysis revealed multidimensional differences of the utterance-final /t/ aspiration noises in early and late peak contexts. The different patterns fit in with patterns that are known from the literature to be systematically related to pitch information in friction noises. The pitch differences that may be expected for the aspiration noises after early and late peaks with regard to the literature go well with the coding profiles of the two intonation categories, which have been independently developed by Niebuhr (2007a) on an empirical basis. This parallel already suggests that the observed differences in the /t/ aspirations did not result from chance. A perception experiment, which complemented the acoustic analysis, showed that German listeners in fact project the multidimensional acoustic differences in the /t/ aspirations onto the attitudinal meaning profiles of early and late peaks, established by previous studies. In particular with regard to the latter finding, the evidence of the present investigation allows to give a first basic answer to the general research question raised in the introduction. That is, segmental friction noises of utterances *can* contribute to the signaling of meanings traditionally associated with F0-based intonation units.

Recently, the findings of Niebuhr (2006, 2007a) already suggested that the attitudinal meanings associated with the German peak categories of the KIM are better described in terms of prosodic units than in terms of intonational ones, as their identification is not only influenced by F0, but also by duration and intensity patterns of the syllables underlying the relevant F0 course. The present investigation continued this line of research and demonstrated that even the extended concept of prosodic units is still too narrow. While the investigation of spontaneous speech has demonstrated that coding (lexical) units at the traditional segmental level needs to be considered in a prosodic perspective (cf. Kohler 1990; Wesener 2001), the present study indicates that the opposite also holds from the prosodic point of view. So, the two traditionally separated coding levels, the segmental and the prosodic (which are also associated with different anatomic and physiological mechanisms and different kinds of meaning), seem to be closely intertwined in the redundant coding of meaningful speech units in various temporal domains. This also means that the segmental string does not (unilaterally) determine and restrict the coding of traditional "intonation units" or their corresponding meanings.

In total, the present study should be regarded as a first step into a neglected wide-ranging field of research which offers new perspectives on the speech code and the related

segmental and prosodic phonetics and phonology. It is the task of following steps to investigate the general research question of this study in more detail. Among others, this means paying attention to the questions that were raised in the discussion of the acoustic analysis and that are related to the question of the generalization of the findings.

Finally, the conclusion of the study may be surprising, as one might expect the aspiration noise to be used—if at all—in a simple chronological way to fill the gaps that voiceless segments tear in the F0 course and the resulting perceived pitch movements. In the present case, this would mean to continue the terminal F0 fall, which is truncated by the voiceless coda of the utterance-final accented syllable, particularly in the case of late peak productions (cf. Fig 1). If the aspiration noises had aimed at a perceptual continuation (or completion) of the terminal fall, they would have either been marked by acoustic patterns which are suitable to evoke equally low pitch impressions after early and late peaks (i.e., no differences would have been found), or they would have shown acoustic patterns that lead to even lower pitch impressions after the extensively truncated late peak falls; but the opposite was found.

Nevertheless, the speech melodies in almost all analyzed utterances actually sound terminal falling, including the ones based on extensively truncated late peaks. Comparable perceptual observations have already been noted in Grabe (1998) and Kohler (2006). So, obviously, there is something in the acoustic signal that leads to the perceptual construction of a complete terminal fall. Dombrowski (2007) demonstrated that German hearers systematically associate various musical semitone intervals with two-word utterances that show either rising-falling peak or (falling-) rising valley contours. He concludes that the range of the rise within a (nuclear) intonation pattern already contains information about the intended contour type, i.e., rise-fall or (fall-) rise. So, one potential acoustic cue triggering the perceptual completion of the terminal fall may be the range of the rise of the intonation peak category.

Moreover, despite its role in the coding of the peak-category meanings, another, possibly even more important cue may *additionally* be contained in the aspiration noise. Due to the perceptual point of view, it was only dealt with so far that the spectral energy maxima above the lower SEB were found at significantly higher frequency values for late peaks than for early peaks (cf. Fig. 2). However, the acoustic analysis also consistently revealed that the frequencies of the spectral energy maxima above the lower SEB, as well as of the lower SEB itself, *decreased* in the course of the aspiration noise. Moreover, this decrease was more pronounced after late peaks, which showed a more extensive truncation of the final F0 than early peaks. Thus, even if the aspiration noises do not directly evoke the utterance-final terminal pitch impression, the shift of the spectral energy to lower frequencies may indicate that a terminal fall was intended by the speaker; and this may then function as an acoustic release for the construction of a terminal fall in the hearer.

Further investigations should address this potential two-track function of utterance-final aspirations, e.g., by contrasting rising-falling and rising-falling-rising patterns, each with the early and the late categories. Irrespective of what will be found, it is already clear at the present stage that truncation is an acoustic and not a perceptual phenomenon. However, if further investigations confirm the two-track function of utterance-final aspirations, then the notion of truncation does not even consistently hold for the acoustic signal.

ACKNOWLEDGMENTS

I am particularly grateful to Klaus Kohler for fruitful and inspiring discussions on earlier drafts of this paper. Moreover, I would like to thank Ernst Dombrowski for taking the time to judge and to describe the aspiration noises as well as Michel Scheffers for his advice concerning the acoustic analysis. Finally, thanks are due my reviewers for their critical and useful remarks, which contributed to make this paper more straightforward and which pointed out many interesting perspectives for follow-up investigations.

Abramson, A. S. (**1972**). "Tonal experiments with whispered Thai," in *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, edited by A. Valdman (Mouton, The Hague), pp. 31–44.

Ambrazaitis, G. I. (**2005**). "Between fall and fall-rise: Substance-function relations in German phrase-final intonation contours," Phonetica **62**, 196–214.

Arvaniti, A., Ladd, D. R., and Mennen, I. (**1998**). "Stability of tonal alignment: The case of Greek prenuclear accents," J. Phonetics **26**, 3–25.

Bruce, G. (**1977**). *Swedish word accents in sentence perspective* (Gleerup, Lund).

Caspers, J., and van Heuven, V. J. (**1993**). "Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall," Phonetica **50**, 161–171.

Cavé, C., Guaitella, I., Bertrand, R., Santi, S., Harlay, F., and Espesser, R. (**1996**). "About the relationship between eyebrows movements and F0 variations," *Proceedings of International Conference on Spoken Language Processing 1996*, Philadelphia.

Chen, A. (**2004**). "How to obtain and process perceptual judgements on intonational meaning?" talk given at a Workshop on Experimental Prosody Research, Leipzig, Germany.

Dombrowski, E. (**2003**). "Semantic features of accent contours: Effects of F0 peak position and F0 time shape," *Proceedings of 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 1217–1220.

Dombrowski, E. (**2007**). "Prosodic rise and rise-fall contours and musical rising two-tone patterns," *Proceedings of 16th International Congress of Phonetic Sciences*, Saarbruecken, Germany.

Erickson, D., Iwata, R., Endo, L., and Fujino, A. (**2004**). "Effect of tone height on jaw and tone articulation in Mandarin Chinese," *Proceedings of the International Symposium on Tonal Aspects of Languages*, Beijing, China, 53–56.

Gartenberg, R., and Panzlaff-Reuter, C. (**1991**). "Production and perception of f0 peak patterns in German," AIPUK **25**, 29–115.

Grabe, E. (**1998**). "Comparative intonational phonology: English and German," Ph.D. thesis, University of Nijmegen, The Netherlands.

Grice, M., and Baumann, S. (**2000**). "Deutsche Intonation und GToBI (German Intonation and GToBI)," Ling. Berichte **181**, 1–33.

Hadding-Koch, K. (**1962**). "Notes on Swedish word tones," *Proceedings of 4th International Congress of Phonetic Sciences*, Helsinki, Finland, 630–638.

Higashikawa, M., and Minifie, F. D. (**1999**). "Acoustical-perceptual correlates of 'whisper pitch' in synthetically generated vowels," J. Speech Lang. Hear. Res. **42**, 583–591.

Hirschberg, J., and Ward, G. (**1992**). "The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English," J. Phonetics **20**, 241–251.

Kiefte, M. (**2005**). "Production and perception of whispered vowels," J. Acoust. Soc. Am. **118**, 1933.

Kohler, K. J. (**1990**). "Segmental reduction in connected speech in German; phonological facts and phonetic explanations," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, The Netherlands), pp. 69–92.

Kohler, K. J. (**1991a**). "A model of German intonation," AIPUK **25**, 295–360.
Kohler, K. J. (**1991b**). "Prosody in speech synthesis: the interplay between basic research and TTS application," J. Phonetics **19**, 121–138.
Kohler, K. J. (**2005**). "Timing and communicative functions of pitch contours," Phonetica **62**, 88–105.
Kohler, K. J. (**2006**). "Paradigms in experimental prosodic analysis: from measurement to function," in *Methods in empirical prosody research*, edited by S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, and P. Augurzky (Walter de Gruyter, Berlin), 123–152.
Kong, Y. Y., and Zeng, F.-G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**, 2830–2840.
Konno, H., Kanemitsu, H., Toyama, J., and Shimbo, M. (**2006**). "Spectral properties of Japanese whispered vowels referred to pitch," J. Acoust. Soc. Am. **120**, 3378.
Krahmer, E., and Swerts, M. (**2006**). "Hearing and seeing beats. The influence of visual beats on the production and perception of prominence," *Proceedings of the 3rd International Conference of Speech Prosody*, Dresden, Germany.
Krull, D. (**2001**). "Perception of estonian word prosody in whispered speech," *Proceedings of the 8th Nordic Prosody Conference*, 153–164.
Ladd, D. R. (**2003**). "Phonological conditioning of F0 target alignment," *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 249–252.
Ladefoged, P. (**2001**). *Vowels and consonants—an introduction to the sounds of language* (Blackwell, Oxford).
Laver, J. (**1994**). *Principles of phonetics* (Cambridge University Press, Cambridge).
Local, J. (**2003**). "Variable domains and variable relevance: Interpreting phonetic exponents," J. Phonetics **31**, 321–339.
Mack, M., and Gold, B. (**1986**). "The effect of linguistic content upon the discrimination of pitch in monotone stimuli," J. Phonetics **14**, 333–337.
McClave, E. (**1998**). "Pitch and manual gestures," J. Psycholinguist. Res. **27**, 69–89.
Meyer-Eppler, W. (**1957**). "Realization of Prosodic Features in Whispered Speech," J. Acoust. Soc. Am. **29**, 104–106.
Mücke, D., Grice, M., Becker, J., Hermes, A., and Baumann, S. (**2006**). "Articulatory and acoustic correlates of prenuclear and nuclear accents," *Proceedings of the 3rd International Conference of Speech Prosody*, Dresden, Germany, 297–300.
Nicholson, H., and Teig, A. H. (**2003**). "How to tell beans from farmers: cues to the perception of pitch accent in whispered Norwegian," Nordlyd **31.2**, 315–325.
Niebuhr, O. (**2003**). "Perceptual study of timing variables in F0 peaks," *Proceedings of 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 1225–1228.
Niebuhr, O. (**2004**). "Intrinsic pitch in opening and closing diphthongs of German," *Proceedings of the 2nd International Conference of Speech Prosody*, Nara, Japan, 733–736.
Niebuhr, O. (**2006**). "The role of the accented-vowel onset in the perception of German early and medial peaks," *Proceedings of the 3rd International Conference of Speech Prosody*, Dresden, Germany, 109–112.
Niebuhr, O. (**2007a**). "Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen," in *Language, Context, and Cognition VII*, edited by A. Steube (Mouton de Gruyter, Berlin), pp. 1–433.
Niebuhr, O. (**2007b**). "The signaling of German rising-falling intonation categories—the interplay of synchronization, shape, and height," Phonetica **64**, 174–193.
Niebuhr, O., and Ambrazaitis, G. I. (**2006**). "Alignment of medial and late peaks in German spontaneous speech," *Proceedings of the 3rd International Conference of Speech Prosody*, Dresden, Germany, 161–164.
Pierrehumbert, J. B. (**1980**). "The phonology and phonetics of English intonation," Ph.D. dissertation, MIT, Cambridge, Mass.
Pierrehumbert, J. B., and Talkin, D. (**1992**). "Lenition of /h/ and glottal stop," in *Papers in Laboratory Phonology II*, edited by G. Docherty and D. R. Ladd (Cambridge University Press, Cambridge), pp. 90–117.
Pike, K. L. (**1948**). *Tone Languages* (Edwards Brothers, Ann Arbor, Mich.).
Rossing, T. D., and Houtsma, A. J. M. (**1986**). "Effects of signal envelope on the pitch of short sinusoidal tones," J. Acoust. Soc. Am. **79**, 1926–1933.
Scarborough, R., Keating, P., Baroni, M., Cho, T., Mattys, S., Alwan, A., Auer, E., and Bernstein, L. (**2006**). "Optical cues to the visual perception of lexical and phrase stress in English," *Proceedings of the 3rd International Conference of Speech Prosody*, Dresden, Germany.
Strevens, P. (**1960**). "Spectra of fricative noise in human speech," Iowa Dent. Bull. **3**, 43–49.
't Hart, J. (**1981**). "Differential sensitivity to pitch distance, particularly in speech," J. Acoust. Soc. Am. **69**, 811–821.
't Hart, J., Collier, R., and Cohen, A. (**1990**). *A perceptual study of intonation. An experimental-phonetic approach to speech melody* (Cambridge University Press, Cambridge).
Thomas, I. B. (**1969**). "Perceived pitch in whispered vowels," J. Acoust. Soc. Am. **46**, 468–470.
Traunmüller, H. (**1987**). "Some aspects of the sound of speech sounds," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht), pp. 293–305.
van Rossum, M. A. (**2005**). "Prosody in alaryngeal speech," Ph.D. thesis, Utrecht University, The Netherlands.
Wesener, T. (**2001**). "Some non-sequential phenomena in German function words," J. Int. Phonetic Assoc. **31**, 17–27.
Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

Consonant identification in noise by native and non-native listeners: Effects of local context

Anne Cutler[a)]
Max Planck Institute for Psycholinguistics, Nijmegen 6500 AH The Netherlands and MARCS Auditory Laboratories, University of Western Sydney, Sydney 1797, Australia

Maria Luisa Garcia Lecumberri
Department of English Philology, University of the Basque Country, Vitoria, 01003, Spain

Martin Cooke
Department of Computer Science, University of Sheffield, Sheffield, S1 4DP, United Kingdom

(Received 28 December 2007; revised 14 April 2008; accepted 28 May 2008)

Speech recognition in noise is harder in second (L2) than first languages (L1). This could be because noise disrupts speech processing more in L2 than L1, or because L1 listeners recover better though disruption is equivalent. Two similar prior studies produced discrepant results: Equivalent noise effects for L1 and L2 (Dutch) listeners, versus larger effects for L2 (Spanish) than L1. To explain this, the latter experiment was presented to listeners from the former population. Larger noise effects on consonant identification emerged for L2 (Dutch) than L1 listeners, suggesting that task factors rather than L2 population differences underlie the results discrepancy.
 [DOI: 10.1121/1.2946707]

PACS number(s): 43.71.Es, 43.71.Hw, 43.71.Sy [MSS] Pages: 1264–1268

I. INTRODUCTION

Users of a second language (L2) know well that listening to speech is more disrupted by background noise in the L2 than in the native language (L1). The differential disruption has also been solidly demonstrated under laboratory conditions (Garcia Lecumberri and Cooke, 2006; Mayo *et al.*, 1997; Nábělek and Donahue, 1984; Takata and Nábělek 1990). However, it is not yet fully clear how the phenomenon should be explained.

One class of explanations locates L1/L2 differences chiefly in phoneme identification. On this account, L2 listeners use other or fewer acoustic cues for phonemes than L1 listeners; if their attention were to be drawn to the cues used by L1 listeners, their speech perception would improve both in general (Jamieson and Morosan, 1986) and in noisy conditions in particular (Hazan and Simpson, 2000). Failure to exploit the full range of phonemic identity cues will render phoneme identification less secure than that of L1 listeners; if phonemes are inaccurately recognized, then in turn no word recognition is possible, and the whole speech perception process more or less falls apart.

Another class of explanations locates differences chiefly at higher processing levels. On such an account (e.g., Bradlow and Alexander, 2007), auditory acuity of L1 and L2 listeners may be equivalent, so that noise disrupts phoneme identification comparably for both L1 and L2 input, but the crucial difference lies in the degree to which recovery from disruption is possible. In the L1, extensive experience with the distributional patterns of phonemes, words, phrases, and constructions pays off in realistic expectations of the probabilities for replacement of missing or misperceived portions of the input. In the L2, insufficient experience has been accrued for realistic expectations to be rapidly derivable. Recovery from disruption is thus too slow for speech perception to be repaired.

Prompting the present study were two specific preceding investigations, which produced different patterns of results. In both cases, phoneme identification by native English-speaking listeners and by L2 listeners was contrasted, and in both cases, the materials were American English phonemes presented in minimal nonsense contexts. Thus both studies attempted to measure phonetic identification performance in the absence of support from lexical or other higher-level information. But in one case the effect of noise on the phoneme identification performance of L1 versus L2 listeners was parallel, while in the other case, noise affected the L2 performance far more than the L1 performance.

In the first study, by Cutler *et al.* (2004) the materials were American English vowels and consonants in CV and VC syllables; the L1 listeners were American English, and the L2 listeners were Dutch; the noise mask was six-talker babble presented at three signal-to-noise ratios (SNRs): 16 dB (very low noise), 8 dB (medium noise), and 0 dB (for L2 listeners, quite high noise). No evidence was found for differential effects of noise on L2 listening; the noise affected L1 and L2 listeners to the same extent, with L2 phoneme identification staying at about 80% of L1 performance across all noise levels.

In the second study, by Garcia Lecumberri and Cooke (2006) the materials were again American English, but only consonants were identified; these were presented in an unvarying a_a context. The L1 listeners were British English, the L2 listeners Spanish. The noise mask was a single competing talker, or speech-shaped noise, or eight-talker babble,

[a)]Author to whom correspondence should be addressed. Electronic mail: anne.cutler@mpi.nl

 0001-4966/2008/124(2)/1264/5/$23.00

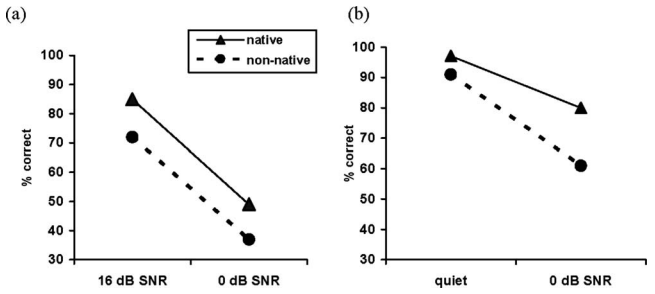


FIG. 1. Percentage of correctly recognized American English consonants in easy versus difficult listening conditions (a) for Dutch-L1 non-native listeners and American English native controls, at 16 dB and 0 dB SNR in six-talker babble; (b) for Spanish-L1 non-native listeners and British English native controls, in quiet and at 0 dB SNR in eight-talker babble.

always at 0 dB SNR, and contrasted with a quiet no-mask condition. In this case L2 performance was 92% of L1 performance without masking, but with the maskers, at 0 dB SNR, it varied from 90% (single competing talker) to 78% (multi-talker babble) – i.e., the noise affected the L2 listeners significantly more than the L1 listeners. Figure 1 contrasts the two studies' results, for consonant identification only, in the most similar conditions: quiet or low-noise versus 0 dB multi-talker babble.

Obviously, there are many differences between the two studies to which the contrast in the result patterns might be ascribed. First, one L1 group was British, the other American, while one L2 group was Spanish and the other was Dutch. The L1 difference is unlikely to be decisive, given that Australian English listeners matched the American performance with the Cutler *et al.* (2004) stimuli (Cutler *et al.* 2005). The L2 groups, however, differ in proficiency, as Garcia Lecumberri and Cooke (2006) pointed out; Dutch users of English have been repeatedly shown to perform at high standards in this L2, both in perception (Broersma, 2005, 2008; Cooper *et al.*, 2002) and production (Bongaerts, 1999). This is partly because Dutch and English are closely related languages, with similar phonology, and partly due to the wide availability of English in Dutch daily life. The Spanish L2 group was less proficient and enjoyed less regular exposure to English outside formal learning. Greater proficiency could help to insulate L2 listeners from the impact of noise beyond that affecting L1 listeners.

Further, Dutch and Spanish differ greatly in the makeup of their phoneme repertoires. Dutch has (unusually among languages, especially those of Europe) a near-balanced repertoire of 19 consonants versus 16 vowels, while Spanish has a highly unbalanced repertoire, quite typical of the world's languages: 20 consonants but only five vowels. As lexico-statistical comparisons have demonstrated (Cutler *et al.* 2004, Cutler and Pasveer 2006), phoneme repertoire structure has far-reaching consequences for the composition of vocabularies and the similarity between words, and hence for the task of word recognition and all the steps involved in it.

Thus although vowels and consonants contribute similarly to word retrieval in Dutch and Spanish (Cutler *et al.*, 2000), the difference in C:V ratio affects phoneme perception in context. In Dutch, effects of contextual (un)predictability in phoneme detection are the same for vowel and consonant detection, but in Spanish, effects of consonant context on vowel detection are greater than effects of vowel context on consonant detection (Costa *et al.*, 1998). Thus Spanish listeners recognize the greater potential for consonantal than for vocalic variation in their language, and this awareness directly affects their phonemic decision making. There are still other differences between these language groups in speech perception; for instance, the type of information used in consonant identification varies, with Spanish listeners paying greater attention to transitions in fricative identification than Dutch listeners do (Wagner *et al.*, 2006). If a particular source of information is more susceptible to masking, then such inter-group differences could also surface as differences in identification success for the affected phonemes in noisy listening conditions. In short, L2 population differences must be reckoned a likely source of the varying results patterns of the Cutler *et al.* (2004) and Garcia Lecumberri and Cooke (2006) studies.

However, differences in the experiments could also have played a role. Even the most comparable conditions, shown in Fig. 1, were not identical across the studies, although the babble talker Ns used fell in the same performance range (Simpson and Cooke, 2005), and Garcia Lecumberri and Cooke (2006) argue that the range of masking effectiveness in the two studies was similar. Cutler *et al.* (2004) participants identified both vowels and consonants, but these identification tasks were performed in separate experimental blocks, and as Fig. 1 shows, the consonant results alone showed the clear inter-experiment difference. The greatest difference was perhaps in the type of stimuli used. The syllables of Cutler *et al.* (2004) were centrally embedded in noise, and since the syllables differed in duration, the precise amount of leading noise varied across the 645 tokens. The syllables were also made from 24 consonants and 15 vowels, so that inter-token variability was high. The Garcia Lecumberri and Cooke (2006) stimuli, in contrast, had a constant leading noise and a constant preceding and following vocalic context for the target consonants.

To test the source of the different results, we presented the Garcia Lecumberri and Cooke (2006) materials to Dutch listeners as tested by Cutler *et al.* (2004). We chose the quiet and multi-talker babble conditions, which showed, respectively, the smallest and largest differences between the Garcia Lecumberri and Cooke (2006) listener groups. If the discrepancy in the Cutler *et al.* (2004) versus Garcia Lecumberri and Cooke (2006) results was due mainly to the L2 listener groups, then we will here replicate the Cutler *et al.* (2004) result (see left panel of Fig. 1): compared to GLC's L1 control group, the listeners will maintain a roughly constant performance decrement across the two conditions. If the discrepancy is due mainly to task differences, then we will here replicate the GLC result (right panel of Fig. 1): the listeners will differ more from the L1 control group in the babble condition than in quiet.

II. METHOD

A. Participants

Sixteen students at the Radboud University Nijmegen (three male; mean age 21 years) took part in return for a small payment. All were native speakers of Dutch.

B. Materials

Two of the five Garcia Lecumberri and Cooke (2006) conditions were presented to the Dutch subjects. Speech tokens were American English VCVs from Shannon *et al.* (1999) corpus. The vowel preceding and following the consonant for identification was always /ɑ/; the consonant was one of 16 possibilities /p b t d k g m n l r f v s z ʃ ʃ/. Two tokens of each VCV from each of five male talkers were used, for a total test set of 160 items. An additional two examples of each VCV formed an initial set of (unscored) practice items. In one condition the speech tokens were presented without mask. In the other condition the speech was masked with eight-talker babble produced by summing utterances by male talkers from the TIMIT corpus. The babble started 1 s before the VCV and ceased at VCV offset. The SNR of the speech in babble was 0 dB. For further details of these materials see Garcia Lecumberri and Cooke (2006).

C. Procedure

Stimulus presentation and response collection were controlled by a computer running MATLAB. Participants were told that the test involved identification of English consonants; they responded on each trial by clicking on one cell of a grid of English words or phrases representing the 16 response options (e.g., "a Path," "aLarm," etc.). Subjects were tested individually, in a sound-attenuated room; the materials were presented binaurally over Sennheiser HD 497 headphones, and the quiet condition was always presented first.

III. RESULTS

The mean percent correct responses in quiet were 93.63%, in babble noise 58.78%. Thus in this experiment the Dutch listeners performed very much worse with the noise-masked stimuli than they did in the quiet condition. The comparable results for British English L1 listeners in Garcia Lecumberri and Cooke (2006) were 98.31% and 80.35%, respectively; their Spanish L2 listeners averaged 90.95% and 62.38%. As can be clearly seen in Fig. 2, the Dutch group's performance resembles that of Garcia Lecumberri and Cooke's (2006) L2 group, not that of their L1 group.

Analyses of variance across subjects (F1) and across items (F2), comparing the performance of the present Dutch group with that of the two groups tested by Garcia Lecumberri and Cooke (2006) on the present subset of the items only, revealed a significant main effect of language (F1 and F2 $p<0.001$), and a main effect of testing condition (F1 and F2 $p<0.001$).

The major difference between the results of Cutler *et al.* (2004) and Garcia Lecumberri and Cooke (2006) was that in the data of Cutler *et al.* (2004) no interaction was observed between the language factor and the presentation

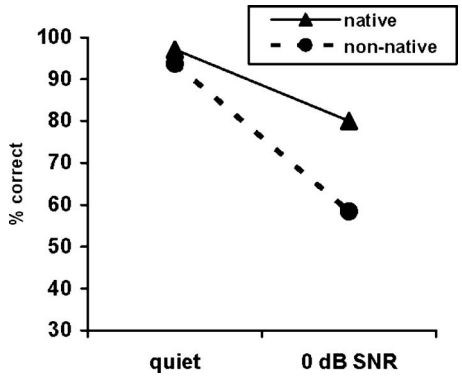


FIG. 2. Percentage of correctly recognized American English consonants for Dutch-L1 non-native listeners and British English native controls, in quiet and at 0 dB SNR in eight-talker babble.

condition (p. 3670), while in the data of Garcia Lecumberri and Cooke (2006) these effects interacted significantly (p. 2449). In the present data this interaction was significant for a comparison of the Dutch and the English groups (F1 [1,35]=153.77, $p<0.001$, F2 [1,159]=76.08, $p<0.001$).

A comparison of the Dutch and Spanish L2 groups alone revealed no significant overall difference in their performance (F1 and F2 <1); however, the small advantage (2.7 percentage points) for the Dutch listeners in quiet was significant (F1 [1,73]=5.27, $p<0.05$; $t2$ [159]=2.93, $p<0.005$), as was also the 3.6 percentage points advantage for the Spanish listeners in babble noise (F1 [1,73]=4.72, $p<0.05$; $t2$ [159]=2.16, $p<0.05$).

To assess performance across the experiment, we compared first and second halves of each condition. Both in quiet and in babble, performance of each group improved from first to second half, and though the gain was quite small, it was significant across all three groups together ($p<0.025$). Importantly, though, neither for the whole experiment nor for either condition was there any interaction of this inter-half difference with listener group.

IV. DISCUSSION

With the materials from the study of Garcia Lecumberri and Cooke (2006), Dutch L2 listeners from the Cutler *et al.* (2004) population produced performance which looked like Garcia Lecumberri and Cooke (2006) L2 results, not like Cutler *et al.* (2004) L2 results. The L2-L1 difference in the babble noise condition was much greater than the difference in quiet. This is the same asymmetry found by Garcia Lecumberri and Cooke (2006), and in other studies with more word- or sentence-like materials (e.g., Mayo *et al.*, 1997; Nábělek and Donahue, 1984). Cutler *et al.* (2004), in contrast, observed similar L1-L2 differences in difficult and easy listening conditions.

Thus the difference in results pattern in the preceding studies does not seem to stem from the difference in listener population. It is not the case that Dutch listeners to English are simply able to resist the well-known L2 disadvantage in noise. Instead, the Garcia Lecumberri and Cooke (2006) task has produced analogous results with Spanish and with Dutch

listeners. This outcome points, as we forecast above, to task differences as the source of the difference in results in the preceding studies. The task of Garcia Lecumberri and Cooke (2006) (like those of most other researchers) produces an interaction between listener group (L1, L2) and listening conditions (easy, difficult), while the task of Cutler *et al.* (2004) does not. The difference seems at least as much due to poorer performance by Cutler *et al.*'s (2004) L1 listeners as to resistance of their L2 listeners to effects of noise (compare Figs. 1 and 2). The results overall thus suggest that while masking effects for L1 and L2 speech can be equivalent, L1 listeners are definitely better at recovering from disruption. They use even the slightest low-level cues provided by a particular experimental context.

The aim of Cutler *et al.* (2004) was an experiment offering no contextual support at all for phoneme identification. In particular, the constant duration of the leading noise and the constant preceding vowel context in the present materials were not available in the materials of Cutler *et al.* (2004). We suggest that these local sources of contextual predictability can, in the absence of other contextual support, be exploited in phoneme identification tasks, at least by proficient (L1) listeners. From other research it is clear that language experience affects use of contextual information to compensate for the effects of noise masking. Mayo *et al.* (1997) presented high- and low-predictability sentences from the Speech Perception In Noise (SPIN) test to monolinguals, early bilinguals, and late learners of English as L2, and observed a significantly greater predictability benefit for the two former groups than for the last. Van Wijngaarden *et al.* (2002) found that L2 users' scores in a letter-guessing task on written text predict their speech recognition scores in noise, suggesting that individual differences in ability to exploit contextual redundancy affect relative resistance to noise masking.

Here, no higher-level (lexical or sentential) context was available to assist listeners in the consonant identification task. However, it seems that even low-level predictability, of the sort provided by a constant vocalic context and a constant duration of leading noise, can aid those listeners better able to use it. Especially in a difficult listening task such as consonant identification against a high level of babble noise, with no assistance available from any type of higher-level information, cues, of any kind, which reduce the uncertainty in the task will be very valuable. Constant vocalic contexts simplify consonant judgments in a variety of decision tasks (Costa *et al.*, 1998; Swinney and Prather, 1980). Leading noise could make the speech onset temporally predictable. Intelligibility of plosive-vowel tokens improves with continuous noise (Ainsworth and Meyer, 1994) or leading noise (Ainsworth and Cervera, 2001) compared to co-gated noise; and predictability leads to more accurate stimulus discrimination in a visual gap detection task (Rolke and Hofmann, 2007) and in an auditory pitch discrimination task (Bausenhart *et al.*, 2007).

It is plausible that both for this and for the role of vocalic context, beneficial effects will be larger for L1 listeners and detrimental effects larger for L2 listeners. Predictable onsets and predictable vocalic environment will both be better exploited by listeners with greater accrued experience of the phonemic processing task in question. We assume that total accrued amount of experience with a given language translates, *ceteris paribus*, to ability to exploit redundancy at all levels of speech processing. Although it might be reasonable to expect performance ceilings where differences between higher-ability groups disappear, in fact the literature concerning speech in noise suggests that effects of total language experience are still observable at high performance rates. Thus bilinguals will in general have accrued less experience with either of their languages separately than monolinguals will have accrued with the one language; and although Mayo *et al.*'s early bilinguals performed much better in noise than did their late learners, the early bilinguals' performance did not match that of monolinguals. Similarly, Rogers *et al.* (2006) presented monosyllabic words, without higher-level context, to monolinguals and to early bilinguals with no perceptible foreign accent in English; they found that the two groups performed equivalently in quiet, but in noise the scores of these bilinguals were much lower.

Finally, note that the full story of speech identification in noise must even allow for differing effects of relevant experience from the L1 in L2 listening. In a separate study (Cutler *et al.*, 2007) we compared the recognition of American English consonants by Dutch and Spanish and British English listeners, using different maskers than were used here, and a larger set of consonants. Crucially, the consonants included affricates, and /θ/, none of which were used in the present materials. The vowel context was constant, but the speech and noise were co-gated, and there was thus no leading noise to serve as a prior cue to the moment of stimulus onset (note, though, that there was predictability in that stimulus presentation was triggered by the participant's keypress). With most consonants, it was again the case that the performance of Dutch and Spanish listeners was parallel: both were more seriously affected by noise than L1 listeners were. However, a different pattern appeared with the fricative/affricate subset of the materials; there Dutch listeners were less seriously affected by noise than either L1 listeners or the Spanish group. This result was interpretable in the light of the cross-linguistic differences reported by Wagner *et al.* (2006); listeners attend to transitional cues for fricatives when their native phoneme inventory contains confusable pairs of fricatives. This is true of English and Spanish (both these languages contrast /f/ and /θ/) but not of Dutch. Gating studies (Wagner, 2008) showed that differential cross-language sensitivity to transitional cues does not generalize across phoneme classes (e.g., to stops). If the presence of a noise mask seriously disrupted the use of transition information, then the paradox of the better performance of the Dutch group with these consonants would be explained: in their case, their native experience with fricative identification, not relying on the cues which were fragile under noise, served them better than the other two groups' experience which required attention to the cues which had been disrupted.

ACKNOWLEDGMENTS

We are grateful to Dennis Pasveer and Eelke Spaak for research assistance. This research was supported by the NWO-SPINOZA project "Native and Non-native Listening" (AC).

Ainsworth, W. A., and Meyer, G. F. (**1994**). "Recognition of plosive syllables in noise: Comparison of an auditory model with human performance," J. Acoust. Soc. Am. **96**, 687–694.

Ainsworth, W. A., and Cervera, T. (**2001**). "Effects of noise adaptation on the perception of voiced plosives in isolated syllables," *Proc. Eurospeech 2001*, Aalborg, Denmark, pp. 371–374.

Bausenhart, K. M., Rolke, B., and Ulrich, R. (**2007**). "Knowing when to hear aids what to hear," Q. J. Exp. Psychol. **60**, 1610–1615.

Bongaerts, T. (**1999**). "Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners," in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong (Erlbaum, Mahwah, NJ), pp. 133–159.

Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Broersma, M. (**2008**). "Nonnative listeners use vowel duration less than English listeners for English final v-f," J. Acoust. Soc. Am. **123**, 3183.

Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang Speech **45**, 207–228.

Costa, A., Cutler, A., and Sebastián-Gallés, N. (**1998**). "Effects of phoneme repertoire on phoneme decision," Percept. Psychophys. **60**, 1022–1031.

Cutler, A., Cooke, M., Garcia Lecumberri, M. L., and Pasveer, D. (**2007**). "L2 consonant identification in noise: Cross-language comparisons," in *Proc. INTERSPEECH 2007*, Antwerp, pp. 1585–1588.

Cutler, A., Norris, D., and Sebastián-Gallés, N. (**2004**). "Phonemic repertoire and similarity within the vocabulary," in *Proc. 8th International Conference on Spoken Language Processing*, Jeju, Korea, Vol. **1**, pp. 65–68.

Cutler, A., and Pasveer, D. (**2006**). "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," in *Proc. 3rd International Conference on Speech Prosody*, edited by R. Hoffmann and H. Mixdorff (TUDpress, Dresden), pp. 237–240.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and Ooijen, B. van (**2000**). "Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons," Mem. Cognit. **28**, 746–755.

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

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Garcia Lecumberri, M. L., and Cooke, M. P. (**2006**). "Effect of masker type on native and non-native consonant perception in noise," J. Acoust. Soc. Am. **119**, 2445–2454.

Hazan, V., and Simpson, A. (**2000**). "The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects," Lang Speech **43**, 273–294.

Jamieson, D. G., and Morosan, D. E. (**1986**). "Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones," Percept. Psychophys. **40**, 205–215.

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**, 686–93.

Nábělek, A. K., and Donahue, A. M. (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**, 632–634.

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.

Rolke, B., and Hofmann, P. (**2007**). "Temporal uncertainty degrades perceptual processing," Psychon. Bull. Rev. **14**, 522–526.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Simpson, S. A., and Cooke, M. P. (**2005**). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778.

Swinney, D. A., and Prather, P. (**1980**). "Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts," Percept. Psychophys. **27**, 104–110.

Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.

Van Wijngaarden, S., Steeneken, H., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.

Wagner, A. (**2008**). "Phoneme inventories and patterns of speech sound perception," Ph.D. thesis, University of Nijmegen.

Wagner, A., Ernestus, M., and Cutler, A. (**2006**). "Formant transitions in fricative identification: The role of native fricative inventory," J. Acoust. Soc. Am. **120**, 2267–2277.

The combined effects of reverberation and nonstationary noise on sentence intelligibility

Erwin L. J. George,[a] Joost M. Festen, and Tammo Houtgast
VU University Medical Center, ENT/Audiology, EMGO Institute, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 29 June 2007; revised 20 February 2008; accepted 19 May 2008)

Listening conditions in everyday life typically include a combination of reverberation and nonstationary background noise. It is well known that sentence intelligibility is adversely affected by these factors. To assess their combined effects, an approach is introduced which combines two methods of predicting speech intelligibility, the extended speech intelligibility index (ESII) and the speech transmission index. First, the effects of reverberation on nonstationary noise (i.e., reduction of masker modulations) and on speech modulations are evaluated separately. Subsequently, the ESII is applied to predict the speech reception threshold (SRT) in the masker with reduced modulations. To validate this approach, SRTs were measured for ten normal-hearing listeners, in various combinations of nonstationary noise and artificially created reverberation. After taking the characteristics of the speech corpus into account, results show that the approach accurately predicts SRTs in nonstationary noise and reverberation for normal-hearing listeners. Furthermore, it is shown that, when reverberation is present, the benefit from masker fluctuations may be substantially reduced. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945153]

PACS number(s): 43.71.Gv, 43.71.An, 43.55.Hy, 43.66.Mk [MSS] Pages: 1269–1277

I. INTRODUCTION

Everyday listening situations generally include a combination of reverberation and noise. It is known that these two factors adversely affect speech comprehension. Their effects need to be taken into account when predicting speech intelligibility in a quantitative way, which can be useful, for example, in the design of public address systems, conference rooms, or public facilities, in general. Examples of objective speech intelligibility methods are the articulation index (AI; French and Steinberg, 1947; Kryter, 1962) or more advanced methods, based on the AI, such as the speech intelligibility index (SII; ANSI, 1997) or the speech transmission index (STI, Houtgast *et al.*, 1980; Steeneken and Houtgast, 1980). These methods have been shown to accurately assess the combined effects of reverberation and stationary noise on speech intelligibility (see, e.g., Bradley *et al.*, 1999, or Beutelmann and Brand, 2006).

However, many everyday backgrounds are nonstationary or "fluctuating" in nature. Normal-hearing listeners are able to make use of the relatively silent periods or "gaps" in these nonstationary backgrounds to improve their speech intelligibility (Festen and Plomp, 1990; George *et al.*, 2006). The STI and the SII methods do not take this effect of background fluctuations into account and, thus, tend to underestimate speech intelligibility. To solve this problem, Rhebergen and Versfeld (2005) introduced the extended speech intelligibility index (ESII), which is able to accurately predict speech reception thresholds (SRTs) in fluctuating noise for normal-hearing listeners. However, it remained unclear how the adverse effect of reverberation on speech intelligibility, in combination with the presence of nonstationary noise, could be assessed.

Reverberation has a double effect on the intelligibility of speech in nonstationary noise. First, reverberation adversely affects the speech modulations, an effect that is present even in silence. Second, reverberation "smears out" the temporal waveform of the masker, thus attenuating masker modulations and reducing the size of the gaps in the masker envelope. Therefore, the benefit in speech reception from masker modulations or "masking release," as often obtained by normal-hearing listeners, will be reduced when reverberation is present.

The current investigation introduces an approach that combines the STI and the ESII to predict speech intelligibility in listening conditions with both reverberation and fluctuating noise. To validate this approach, SRTs are measured in several nonstationary backgrounds. To further test the assumptions underlying the approach, the reverberation is chosen to act either on the speech only, on the noise only, or on both speech and noise, of which the latter condition—of course—most closely resembles situations in daily life. Furthermore, the consequences of reverberation for masking release are investigated.

II. PREDICTING SPEECH INTELLIGIBILITY

A. Classical prediction methods

The STI (Houtgast *et al.*, 1980; Steeneken and Houtgast, 1980) is based on the observation that speech intelligibility is related to the preservation of the envelope spectrum of speech. Noise and reverberation in the listening environment adversely affect this preservation of modulations. The preservation is characterized by the modulation transfer function

[a]Electronic mail: elj.george@vumc.nl

(MTF), which quantifies the detrimental effects of distortions on the modulations of the envelope of (band-filtered) speech. The MTF is expressed in matrix form, by determining the modulation transfer indices m, for each of 14 modulation frequencies (0.63–12.5 Hz) and for each of seven spectral octave bands (center frequencies of 125–8000 Hz). The modulation transfer indices are translated to effective signal-to-noise ratios (SNRs), averaged over modulation frequencies, and converted to an index between 0 and 1, after which a weighted average over the octave bands gives the STI.

The STI does not take individual properties of talkers or listeners into account, but is purely a measure of the transmission channel or, to be more specific, the acoustic characteristics of the environment. It has been shown to be strongly related to speech intelligibility (Houtgast and Steeneken, 1984) and is widely used as an acceptability criterion in room acoustics or telecommunication.

The SII (ANSI, 1997) is, in fact, a very similar method, but its applicability is essentially restricted to assessing the effect of noise. Instead of the MTFs, the spectra of speech and noise are used to determine the effective SNR in each spectral band. The SNRs are converted to an index between 0 and 1 and taking the weighted average over the spectral bands gives the SII. It is often interpreted as the amount of undistorted speech information available for an individual listener. Recently, the ESII (Rhebergen and Versfeld, 2005) was introduced, which makes it possible to apply the SII in non-stationary backgrounds by calculating and averaging the SIIs determined in short time frames.

In contrast to the STI, the SII does take the individual properties of talkers and listeners into account. It is possible to include the individual audiogram of the listener in the calculation, and different spectral weightings can be chosen to assess various talker styles or speech materials.

The SII, like the STI, is related to speech intelligibility: poor intelligibility is associated with SII values below 0.45 and good intelligibility assumes a SII larger than 0.60. In the assessment of speech intelligibility for individual listeners, the SRT is commonly used, defined as the SNR that the listener needs to correctly repeat 50% of the presented sentences correctly. For normal-hearing listeners, the SRT corresponds to a STI or SII of about 0.33. This means that, in situations with stationary noise without reverberation, both methods predict a SRT of about −5 to −4 dB SNR, consistent with results found in the literature (see, e.g., Festen and Plomp, 1990; Versfeld and Dreschler, 2002). Using the STI, the SRT can be predicted for all other combinations of stationary noise and reverberation, always assuming that the STI at the threshold is constant.

In summary, the STI can be used to evaluate the detrimental effects of reverberation and stationary noise on speech intelligibility, while the recently developed ESII can be applied to predict the effect of nonstationary noise on SRTs. This motivated us to derive a procedure to assess the combined effects of reverberation and nonstationary noise. This approach, based on combining the two methods, is outlined in the following section.

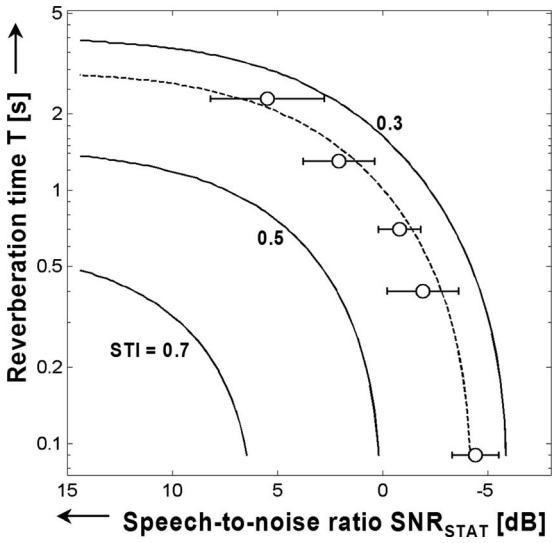


FIG. 1. Iso-STI contours for listening conditions that include a combination of reverberation and stationary noise. The dotted curve represents the .33 iso-STI contour. The data points are measurement results for normal-hearing listeners by Duquesnoy and Plomp (1980), representing the SNR at the SRT, i.e., the level at which 50% of the sentences could be correctly reproduced, for various reverberation times T. Redrawn, from Houtgast *et al.* (1980).

B. Combining the STI and the SII

The STI already combines the effects of noise and reverberation for sentence intelligibility in one method, only with the restriction that the noise should be stationary in character. Let us start with evaluating this method, after which we will generalize it to nonstationary noise.

Reverberation time T is defined as the time after which a sound has died away to a level of 60 dB below its original level. It is commonly determined by extrapolating the slope of the early decay curve. When assuming an exponentially decaying sound field, the STI can be easily calculated for any combination of reverberation time T and speech-to-noise ratio in stationary noise SNR_{stat} (see Duquesnoy and Plomp, 1980). The modulation transfer index m, as used in the STI calculation, is then the product of two terms, one reflecting the effect of reverberation and the other reflecting the effect of background noise. The result is represented in Fig. 1 in the form of iso-STI (isointelligibility) contours. These contours indicate that it is possible to find, at each value of T, a SNR_{stat} at which 50% of the presented sentences can be correctly reproduced. Results by Duquesnoy and Plomp (1980), as also displayed in Fig. 1, confirm that the SRT in different reverberant sound fields is indeed distributed along an iso-STI contour, i.e., can be represented by one single STI value. For normal-hearing listeners, the resulting STI is about 0.33, for all combinations of reverberation and stationary noise (ISO, 2002).

To generalize this approach to nonstationary backgrounds, we need to find the threshold speech-to-noise ratio in fluctuating noise (SRT_{fluc}) that gives the same intelligibility as obtained at the threshold in stationary noise (SRT_{stat}). Thus, the SII or STI at the threshold should be the same, whether the noise is stationary or fluctuating in character. Rhebergen and Versfeld (2005) showed that the extended SII may be applied to evaluate the SII in fluctuating noise. Thus,

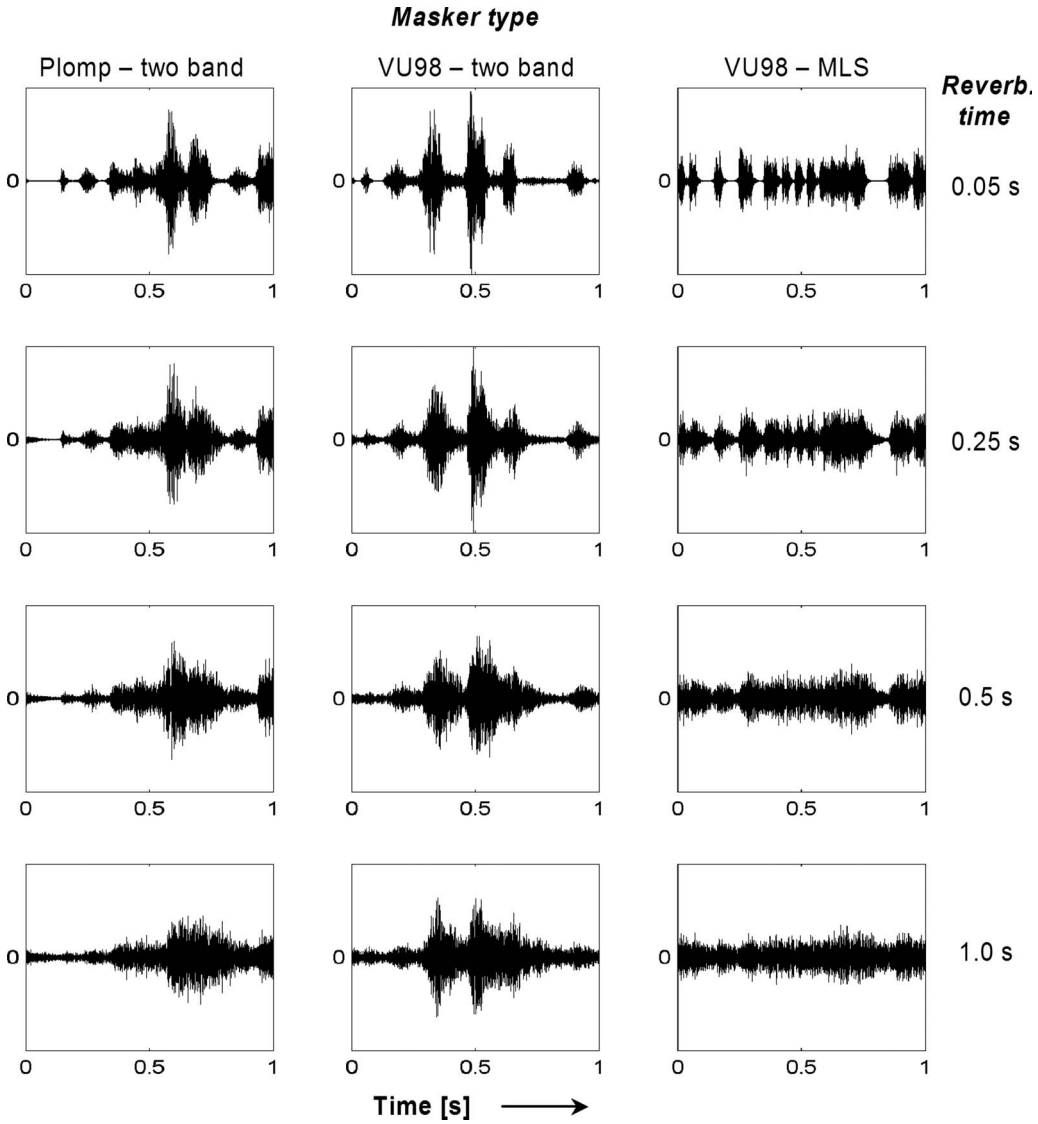


FIG. 2. The effect of simulated reverberation on the temporal waveforms of three different types of nonstationary maskers. To be able to compare the masker types in terms of visible waveform modulations, only the modulations for one representative octave band (around 2 kHz) are shown. Details on the masker types and the reverberation procedure can be found in the Sec. III.

we apply the ESII to determine the speech-to-noise ratio in fluctuating noise (SRT_{fluc}), that, in terms of intelligibility, is equivalent to the SNR in stationary noise (SRT_{stat}).

However, this approach to predict SRTs in fluctuating noise assumes that reverberation and noise independently affect speech reception. This assumption is correct in the case of stationary noise: the reverberation does not affect the characteristics of the background noise, since reverberant stationary noise is still stationary in nature. When nonstationary noise is concerned, however, the reverberation will not only adversely affect the speech, but will also change the characteristics of the noise. Some examples of the effect of reverberation on the temporal waveform of the noise are displayed in Fig. 2, for three different types of maskers. Reverberation was simulated here by convolving the nonstationary noise with synthetic impulse responses, the details of which are described in the Sec. III. As shown, the relatively silent periods or gaps in the maskers are reduced in size by the reverberation. At larger reverberation times, masker fluctuations are further reduced and eventually, there are no fluctuations or gaps present anymore and the fluctuating noise has become more or less stationary in character. A direct consequence of this effect is that the benefit from masker fluctuations in speech reception, as often observed for normal-hearing listeners, is reduced as reverberation time increases, until the point where no masking release is obtained anymore. Put differently, reduction in masker fluctuations as a consequence of reverberation directly affects the SRT in fluctuating noise.

The effect of reverberation on nonstationary noise for speech intelligibility can be quantified by applying the ESII, as shown in Fig. 3. This figure displays, for various noise types and reverberation times, the calculated amount of available speech information, expressed by the ESII, as a function of SNR. As reverberation time increases, the curves relating intelligibility (the ESII) to SNR for fluctuating maskers approach the curve for stationary noise. This effect is most prominent for the masker in the rightmost panel, for

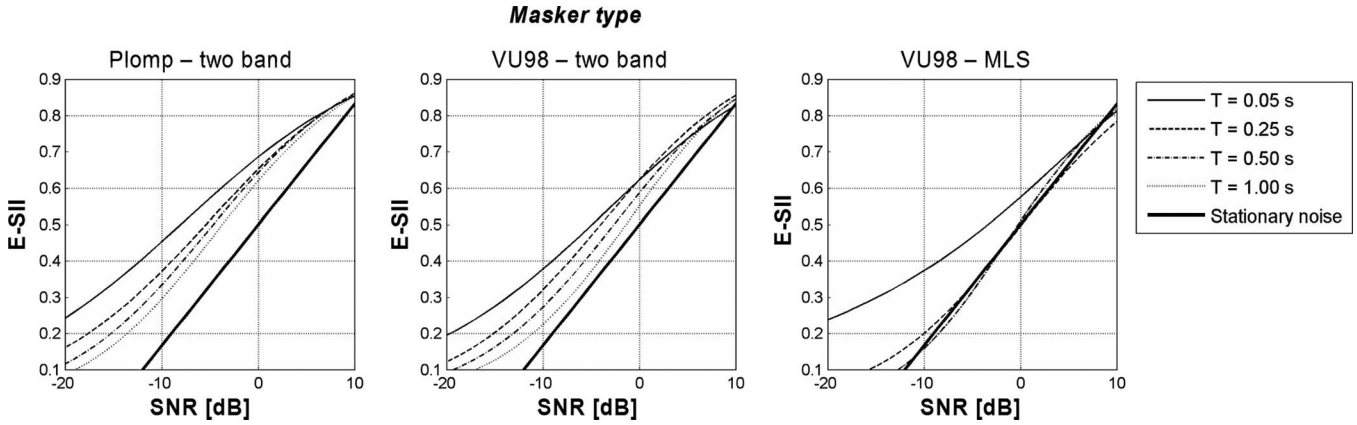


FIG. 3. ESII as a function of speech-to-noise ratio (SNR) for three different types of nonstationary maskers with reverberation time as a parameter. ESII values were calculated for each reverberant masker and fitted to the three-parameter asymmetric logistic function $\mathrm{ESII}(\mu,\sigma,n)=[1+n^{*}\exp(\mu\text{-SNR})/\sigma]^{-1/n}$ reduces to a symmetrical sigmoid for $n=1$. For stationary noise, ESII is simply the linear function $(\mathrm{SNR}+15)/30$, see ANSI (1997). Details on the masker types and the reverberation procedure can be found in the Sec. III.

which the $T=0.5$ s curve can hardly be distinguished from the curve for stationary noise anymore. Choosing the appropriate reverberation time and noise type, the resulting curve can be used to translate the speech-to-noise ratio in stationary noise ($\mathrm{SRT}_{\mathrm{stat}}$) to a SNR in nonstationary noise ($\mathrm{SRT}_{\mathrm{fluc}}$) that is equivalent in terms of the effect on intelligibility.

In summary, Fig. 4 gives an overview of the approach that is used for predicting SRTs in combinations of nonstationary noise and reverberation. The STI, in the form of iso-STI curves, is applied to determine the effect of the reverberation T on the speech modulations (see Fig. 1). This gives rise to a prediction of the SRT in stationary noise, $\mathrm{SRT}_{\mathrm{stat}}$. The effect of reverberation on the masker is also separately evaluated. Finally, the ESII is used (see Fig. 3) to translate the obtained $\mathrm{SRT}_{\mathrm{stat}}$ to an equivalent $\mathrm{SRT}_{\mathrm{fluc}}$ in the reverberant fluctuating noise.

To determine whether the proposed approach, combining the STI and the ESII, gives accurate predictions of speech intelligibility, an experiment was conducted, described in the next section, in which SRT measurements for various combinations of noise types and reverberation times were performed.

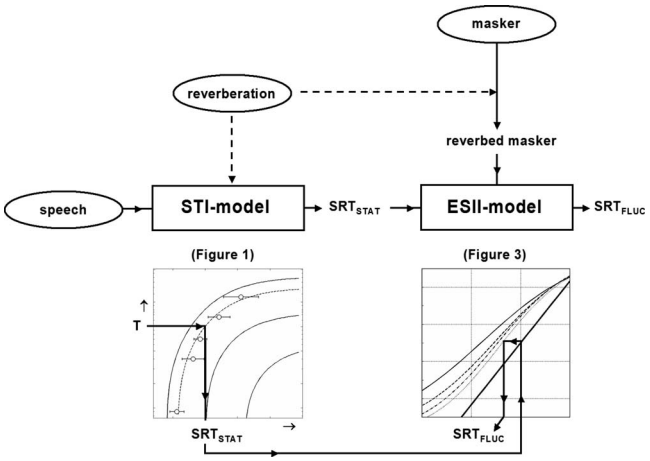


FIG. 4. Overview of the proposed prediction method. The effect of reverberation on speech is described by the STI method, while the effect of reverberation on the masker is also evaluated separately. Subsequently, the ESII is applied to determine the combined effects of both reverberation and masking noise on sentence intelligibility. The example below applies to a reverberation time of $T=1$ s and the Plomp-two band masker type.

III. EXPERIMENT AND METHOD

A. Participants

Ten young, normal-hearing listeners participated in this experiment. Eight were students from the VU University, while the other two were young university graduates. They reported no problems with their hearing or with speech reception, and were selected to have pure-tone hearing thresholds equal to or better than 10 dB HL at octave frequencies between 0.25 and 4.0 kHz, i.e., at the most relevant frequencies for speech reception. Their age ranged from 18.3 to 31.4 years, with an average of 22.9 years.

B. Method

SRT measurements were conducted by using a simple adaptive up-down procedure as described by Plomp and Mimpen (1979). In each condition, the masker and a list of 13 sentences, unknown to the listener, were presented. At a constant reverberation time, the speech-to-noise ratio was varied adaptively to estimate the SRT. Lower SRT values indicate better performance. 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 in the list were presented only once, following a one-up-one-down adaptive procedure, with a 2 dB step size. An errorless reproduction of the entire sentence was required for a correct response. The SRT was estimated as the average presentation level of sentences 4–13.

C. Stimuli

1. Sentences

Two sets of short meaningful sentences were used as speech material. The first speech corpus was developed and evaluated by Plomp and Mimpen (1979), consisting of ten

lists of 13 Dutch sentences, uttered by a female speaker. All ten lists were used in the current experiment. The second corpus used here is the "VU98" corpus (Versfeld *et al.*, 2000), which consists of 39 lists of sentences of a male talker and an equal number of lists for a female talker. The first 20 lists by the female speaker were used. Both corpora were developed to enable efficient measurement of SRTs in noise and are considered equivalent in terms of reliability, although the speaking style of the VU98 sentences is generally more informal.

2. Maskers

Sentences were presented in three different maskers; all maskers were designed to mimic the intensity fluctuations of speech. However, their specific characteristics were different. The first masker was generated using the method described by Festen and Plomp (1990), which splits up a stationary masker in a low- and a high-frequency part with 1000 Hz crossover frequency. Both parts are then modulated separately with the envelope of speech from the corresponding frequency region, after which they are added while restoring the original level ratio between the two bands. The second masker was similar to the first; the only difference was that the first masker was based on modulations in the speech corpus of Plomp and Mimpen (1979), while the second masker was based on the speech corpus of Versfeld *et al.* (2000). The third masker was constructed by multiplying a stationary noise with a time-scaled maximum-length sequence (MLS) with a value of either zero or unity. After scaling, this resulted in an on-off noise with a flat modulation spectrum between approximately 3 and 20 Hz, thus covering the spectrum of all the important envelope modulations in speech (Houtgast and Steeneken, 1985; Drullman *et al.*, 1994). The first two maskers are similar in nature and will be denoted as the "two band" masker type, while the third masker is denoted as the "MLS" type. For each of the three maskers, 10 s samples were constructed, from which a sequence of about 2–3 s was randomly chosen for each sentence presentation. Examples of the waveforms of all three maskers are displayed in Fig. 2.

Each of the three maskers was combined with the corresponding speech corpus. In all conditions, the long-term spectra of the speech and the masker were identical. The Plomp and Mimpen (1979) speech corpus was thus presented in masker one, while the VU lists were used to measure SRTs in the second and the third masker. This approach made it possible to assess the influence of speech corpus and masker type on the predictions. Since the characteristics of the first two maskers were essentially the same, the difference in predictions between the two expresses the effect of speech corpus on predictions. The influence of masker type may be assessed by comparing the SRT predictions for the second and the third masker.

3. Reverberation

Reverberation was introduced by convolving the digital signals with synthetic impulse responses. This approach made it possible to define the reverberation time in a systematic way, excluding unwanted effects of room acoustics which may be present when using real, recorded impulse responses. In terms of the MTF, the artificial reverberation was identical to purely exponentially decaying real reverberation of the same reverberation time. The impulse responses were created by multiplying white noise with the exponential decay envelope. The slope of the decay of the impulse response was determined by the desired reverberation time T, chosen to be 0.05, 0.25, 0.5, and 1.0 s. The length of the impulse response was equal to the simulated reverberation time.

These impulse responses of various reverberation times were convolved with either only the speech, or only the masker, or both the masker and the speech. This leads to three measurement conditions for each reverberation time. The short reverberation time of 0.05 was included as a reference condition and expected to be similar to "no reverberation." At this shortest T, reverberation was chosen to always act on both the speech and the noise. Finally, this resulted in a total of ten "reverberant" conditions: three modes of reverberation $\times$ three reverberation + the reference condition.

Reverberation was applied to the speech and/or to the selected masker, after which they were mixed according to the desired SNR, based on their (signal-by-signal) rms levels. The spectrum and presentation level of the resulting signal were adapted to reach octave signal levels equal to the 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 was the uncomfortable loudness level (UCL), here chosen at 110 dB SPL for all listeners. This approach is commonly used in our laboratory to assure optimal audibility for all listeners (see George *et al.*, 2006). It should be noted that, as a consequence of optimizing audibility, noise and reverberation are thus the main factors limiting speech intelligibility in the current experiment, rather than reduced audibility.

D. General method and instrumentation

A test session always started with the measurement of the listener's audiogram. Subsequently, a total of 30 SRT measurements were performed for each listener, in three blocks of ten SRTs, using one specific masker type within each block of ten reverberant conditions. Characteristics of the three maskers and conditions were described above. The order in which the three blocks were presented was randomly determined for each participant. Within each block, confounding of measurement order and sentence lists with condition effects was avoided by counterbalancing the order of conditions across subjects, according to a 10×10 diagram-balanced Latin square, while list order was kept fixed.

The experiment was run on a Dell personal computer, equipped 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 conforming to ISO 389 (1991). All measurements were performed while the listener and the investigator were seated in a sound-insulated room. SRTs were conducted monaurally, using the partici-

TABLE I. Group means (M) and standard deviations (S) for speech reception thresholds (SRT, in dB SNR) in ten reverberant conditions, for three combinations of speech corpus (Plomp and VU98) and nonstationary masker type (two band and MLS).

		Reverberation time T (s)							
		0.05		0.25		0.5		1.0	
		M	S	M	S	M	S	M	S
Plomp—two band	Both reverberant	−15.4	3.1	−9.0	1.5	−5.3	2.5	−0.6	2.4
	Only noise	⋯		−13.1	2.0	−11.2	2.2	−8.6	1.9
	Only speech	⋯		−13.8	2.2	−11.8	3.0	−8.7	2.1
VU98—two band	Both reverberant	−7.8	2.5	−2.5	2.4	+1.6	1.5	+5.1	2.2
	Only noise	⋯		−6.9	1.8	−6.0	1.6	−4.6	1.7
	Only speech	⋯		−5.8	2.1	−3.2	2.0	+3.8	4.0
VU98—MLS	Both reverberant	−13.5	4.2	−1.1	2.0	+2.1	1.7	+6.0	1.9
	Only noise	⋯		−3.9	1.5	−3.5	0.7	−3.4	0.9
	Only speech	⋯		−13.1	3.9	−8.3	4.4	−2.1	7.1

pant's best ear, which was chosen according to his or her audiogram, or, in case of equal audiograms, personal preference in telephone conversation.

IV. RESULTS

Measurement results are displayed in Table I for each of the three maskers used in the current experiment. Predictions (not shown) were derived by assuming a constant STI of 0.33 at the threshold, i.e., the "STI at the SRT" =0.33 for all conditions. The predictions appeared reasonable for the Plomp and Mimpen (1979) speech corpus, but the SRTs measured with the VU corpus were not adequately described. In all cases where the VU corpus was used, listeners obtained higher SRTs than predicted, i.e., needed more undistorted speech information to correctly reproduce half of the sentences.

These findings can be understood when considering results by Van Wijngaarden and Houtgast (2004). They showed that the classic STI, as applied to calculate the predictions, indeed underestimates the adverse effect of reverberation on speech intelligibility when conversational speech by an untrained talker is concerned. This effect was explained by the relatively stronger contributions of higher modulation frequencies in this type of speech material. They obtained better results for an adapted version of the STI method, in which the modulation frequency range was extended to 31.5 Hz, including a total of 18 modulation frequency bands instead of the classic 14. The difference in speaker styles is then expressed by differences in the STI at the SRT between the speech corpora: 0.31 for the Plomp and Mimpen (1979) corpus and 0.37 for the VU corpus.

Figure 5 shows our measurement results and predictions when the differences in speaker style are taken into account, that is, by applying the 18-band STI method to determine the effect of reverberation on speech. Predictions, represented by curves, were derived here by choosing the STI at the threshold, for each speech corpus, such that an optimal fit "least squares" was obtained with the measurements. The sums of squared (SSQ) residuals are 10.1, 37.3, and 37.2, for the left, the middle, and the right panels, respectively.

Figure 6 displays a scatter plot of the observed SRTs and the predictions, for all the conditions from the current experiment. It can be seen that the observed SRTs and the predictions are in good agreement. The standard deviation of the data from the predictions is 1.7 dB, while the maximum deviation is 4.2 dB.

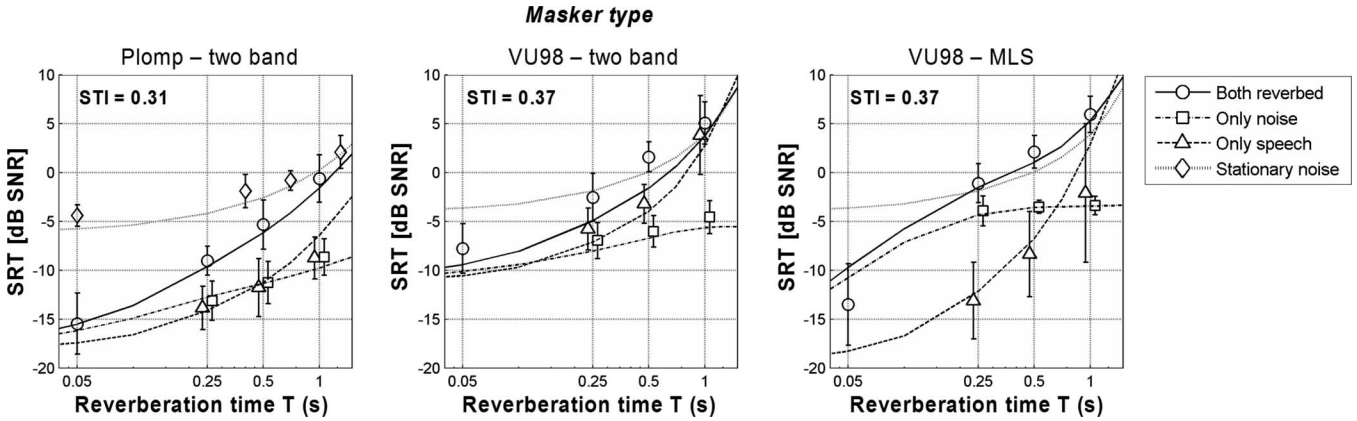


FIG. 5. SRTs as a function of reverberation, for three combinations of speech corpus (Plomp and VU98) and nonstationary masker type (two band and MLS). Predictions, represented by curves, were determined by applying the STI method with 18 modulations bands, as suggested by Van Wijngaarden and Houtgast (2004). For each speech corpus (Plomp or VU98), the STI at threshold was chosen to give an optimal fit (least squares) between measurements and predictions. The leftmost panel also displays results from SRT measurements in stationary noise, taken from Duquesnoy and Plomp (1980).

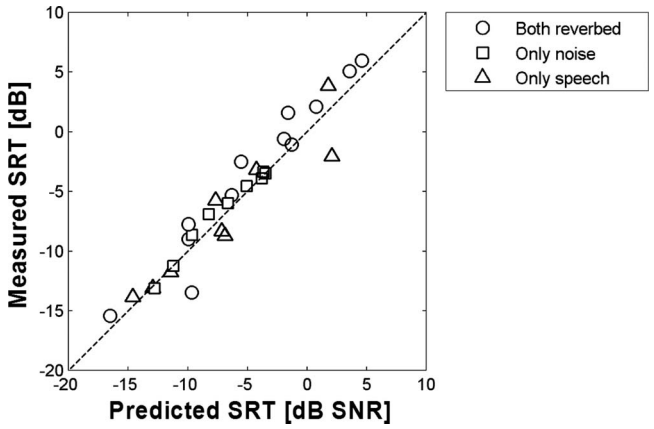


FIG. 6. Scatter plot of the observed SRTs and the predicted SRTs, for all combinations of speech corpus, masker type, and reverberation.

V. DISCUSSION

A. Improving predictions of speech intelligibility

An optimal fit was obtained between the observed and predicted SRTs by choosing a different STI at the threshold, for each speech corpus (Plomp or VU98). This fitting procedure gave rise to a STI of 0.31 for the Plomp speaker and a STI of 0.37 for the VU98 speaker. These values are consistent with the results of the measurements by Van Wijngaarden and Houtgast (2004). The difference between the obtained optimal STI values can be accounted for by the effect of speaker style, which is more informal for the VU98 corpus.

For the VU98 speech corpus, the chosen STI, giving rise to an optimal fit between data and predictions, was based on data from both the VU98/"two band" masker combination and the VU98/"MLS" masker combination. The fit could be further improved by distinguishing between the two combinations, i.e., by fitting an optimal STI value for each combination separately. In that case, a STI value of 0.35 was obtained for the VU98/MLS masker combination, while for the combination of the VU98 speaker and the two band masker, an optimal fit was obtained using a STI value of 0.40. The values of the SSQ residuals then decrease from around 37 for both combinations to 10.2 (two band) and 25.6 (MLS), respectively. The standard deviation of the data improved from 1.7 to 1.3 dB, while the maximum deviation decreased to 2.6 dB.

The obtained STI value of 0.35 for the VU98/MLS masker combination is still consistent with the results of Van Wijngaarden and Houtgast (2004). In contrast, the obtained STI value of 0.40 for the combination with the two band masker seems rather high. Apparently, listeners find it hard to understand speech in this specific combination of masker type and speaker style, giving rise to an elevated STI value. This may be brought about by effects related to informational masking.

In theory, an alternative explanation may be that the STI, the ESII, or the combination of both is not working optimally and should be adapted for this specific combination of speaker and masker. However, these explanations are not very probable. The SII model was validated with various combinations of speech and masker types (Rhebergen and Versfeld, 2006). The adapted STI, including a larger range of modulation frequencies, has been shown to be robust to reverberation, giving rise to a constant "STI at the threshold," also for the VU98-speech corpus (Van Wijngaarden and Houtgast, 2004). Therefore, the authors see no reasonable rationale to question the applicability of the adapted STI for this particular combination of speech type and masker.

B. Effect of reverberation on speech modulations and on nonstationary maskers

It can be seen in Fig. 5 that, in all conditions, more reverberation gives rise to a higher SRT. This was to be expected, since, as mentioned earlier, reverberation reduces the modulations present in speech and also reduces the modulations in the masker. The figure also shows that, as expected, the curves for the conditions in which only the noise was reverberant approach the limit of the SRT in stationary noise (around −5 to −4 dB SNR), for all three masker types. The curves for the conditions in which only the modulations of the speech were affected by reverberation theoretically reach a SRT of +15 dB at reverberation times around 2–3 s. In that case, reverberation affects the speech modulations so severely that the masker can no longer be tolerated.

The curves for the conditions where both the noise and the speech were reverberant can be regarded as "weighted sums" of the other two curves. At lower reverberation times, speech intelligibility performance is dominated by the benefit that a listener can derive from "listening in the gaps," i.e., the benefit from masker fluctuations. At larger reverberation times, the fluctuating maskers become more stationary in character: the differential effect on masker fluctuations between a 0.5 s reverberation and a 1.0 s reverberation is fairly small, as shown by the "noise-only" data. This means that, at larger reverberation times, SRT differences between the "both reverbed" conditions must be brought about by the effect of reverberation on speech modulations, which is still fairly large, as may be derived from Fig. 1.

In the left panel and especially in the middle panel, the SRT is mainly affected by the adverse effect of the reverberation on the speech modulations, while the effect of reduced masker modulations on intelligibility remains reasonably limited. For the rightmost panel—the MLS masker—the situation seems the other way around: the effect of reduced speech modulations is relatively small for low reverberation times and the effect of the reverberation on the masker is dominant in affecting the SRTs. Particularly when a small amount of reverberation is introduced (T=0.05 s), the effect of the reverberant noise on speech intelligibility appears substantial. Data from a pilot experiment (unpublished) showed an average SRT of −20.8 dB SNR (standard deviation of 2.7 dB) in this type of masker when no reverberation is present, while the currently measured SRT at T=0.05 s is −13.5 dB SNR, a difference of about 7 dB. The fact that even a small amount of reverberation has such a large effect on SRT in the MLS masker can be explained by the masker's modulation spectrum. Since its modulation spectrum is flat,

the MLS masker contains a relatively large amount of fast modulations, up to 20 Hz. In addition, the modulation depth of the MLS masker was 100%. These fast and deep masker fluctuations are more vulnerable to reverberation (more easily "smeared") than modulations with lower frequencies and smaller modulation depths, as mostly present in the two band masker type.

C. Effect of reverberation on masking release

Finally, consider the difference between the predictive curves for the conditions where both the noise and the speech were reverberant and the curves for stationary noise (Fig. 5). This difference gives an estimate for the obtained masking release, that is, the benefit in speech reception due to masker modulations, at a specific reverberation time. At reverberation times below 0.1 s, masking release is clearly present (up to 10 dB), see Festen and Plomp (1990) or George *et al.* (2006). When reverberation increases, masking release is reduced and appears not present anymore at reverberation times near 1.0 s in the leftmost panel. In the middle panel, masking release is also reduced substantially by reverberation, approaching zero at a reverberation time around 0.5 s. In the rightmost panel, a reverberation time of 0.25 s is already enough to fully eliminate the obtained benefit. Thus, when the effects of reverberation on the characteristics of both the masker and the speech are taken into account, the benefit that listeners obtain from masker modulations is largely reduced.

These differences between the panels concerning the reduction of masking release can be understood when considering the effects of reverberation on the speech modulations or on the masker, as explained in the previous section. However, the reduction of masking release by reverberation in nonstationary noise may have large consequences. In everyday life, the style of the speaker is often not very clear and fairly informal, i.e., comparable or even worse in quality than the speaker used in the middle and right panels of Fig. 5. Moreover, reverberation times around 0.4–0.5 s, or even larger, are very commonly found in daily listening environments such as living rooms, offices, or hospitals. The results from the current experiment indicate that masking release in even these common everyday situations may be substantially reduced or even absent.

It should be noted, however, that the presented predictions only hold when the listener is in the indirect sound field, that is, for distances to the speaker larger than about 0.2 $(V/T)^{1/2}$. For a living room of 75 m^3 and $T=0.5$ s, this distance is 2.5 m (see Duquesnoy and Plomp, 1980). When the listener approaches the speaker more closely, the effect of reverberation is reduced and the obtained masking release in fluctuating noise is likely to be larger than predicted. Nevertheless, when the listener is not very close to the speaker, the obtained benefit from masker modulations may be considerably reduced in situations with noise and reverberation.

D. Applicability and limitations

It has been shown that the "STI at the threshold" values from literature (Van Wijngaarden and Houtgast, 2004) gives rise to an optimal fit between predictions and obtained data, i.e., lead to optimal predictions. Apparently, the fitted values of the STI are specifically those values that a typical normal-hearing listener needs for 50% sentence intelligibility. This means that, by assuming that a normal-hearing listener needs a STI equal to this value, and applying the proposed method, the SRT can be predicted reasonably well, for all combinations of fluctuating noise and reverberation. Thus, the current method extends the applicability of the STI to listening conditions with nonstationary maskers, which is interesting since fluctuating backgrounds are very common in everyday situations.

While the STI value for normal-hearing listeners may be expected to be fairly constant across listeners, the STI that hearing-impaired listeners need to reach 50% speech intelligibility may differ from person to person. The amount that the STI is elevated for a specific hearing-impaired listener depends not only on the individual's hearing loss but is also related to individual suprathreshold deficits in auditory processing (see, e.g., George *et al.*, 2006). Although the methods to estimate the amount that the STI is elevated, based on hearing loss, have proven to be fairly successful (see Humes *et al.*, 1986; Payton *et al.*, 1994; Halling and Humes, 2000), one should be careful to apply the current prediction method to individual hearing-impaired listeners, considering the large variety of difficulties that hearing-impaired listeners experience in complex listening situations.

Finally, a limitation of the proposed method is that the background masker should be known to make accurate predictions. In common daily situations, background noises can vary greatly. This means that, to predict speech intelligibility, assumptions will have to be made regarding the presentation level, the spectrum, and the modulation spectrum of the background noise. In addition, the speech-to-noise ratio and the applied reverberation time in the current experiment were fixed, while in everyday listening situations, they may change, both as a function of time and of frequency.

VI. CONCLUSION

An approach has been introduced that combines the ESII and the STI to assess the combined effects of reverberation and nonstationary noise on speech intelligibility. After taking the characteristics of the speech corpus into account, the proposed method accurately predicts the SRT in nonstationary noise and reverberation for normal-hearing listeners. Further consideration of the predictions shows that the masking release, as observed in the absence of reverberation, may be substantially reduced in everyday listening situations with noise and reverberation.

ACKNOWLEDGMENTS

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands. Thanks are due to Johannes Lyzenga for his contribution to the discussions underlying this paper and for providing the MLS masker.

ANSI (**1997**). ANSI S3.5–1997, "American national standard methods for the calculation of the Speech Intelligibility Index," American National Standards Institute, New York.

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.
Bradley, J. S., Reich, R. D., and Norcross, S. G. (**1999**). "On the combined effects of signal-to-noise ratio and rooms acoustics on speech intelligibility," J. Acoust. Soc. Am. **106**, 1820–1828.
Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.
Duquesnoy, A. J., and Plomp, R. (**1980**). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. **68**, 537–544.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal-hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
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.
Halling, D. C., and Humes, L. E. (**2000**). "Factors affecting the recognition of reverberant speech by elderly listeners," J. Speech Lang. Hear. Res. **43**, 414–431.
Houtgast, T., Steeneken, H. J. M., and Plomp, R. (**1980**). "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics," Acustica **46**, 59–72.
Houtgast, T., and Steeneken, H. J. M. (**1984**). "A multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria," Acustica **54**, 185–199.
Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.
Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (**1986**). "Application of the Articulation Index and the Speech Transmission Index to the recognition of speech by normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **29**, 447–462.
ISO (**1991**). ISO 389:1991(E), "Acoustics—Standard reference zero for the calibration of pure-tone air conduction audiometers," International Organization for Standardization, Geneva, Switzerland.
ISO (**2002**). ISO/FDIS 9921, "Ergonomics—assessment of speech communication," International Organization for Standardization, Geneva, Switzerland.
Kryter, K. R. (**1962**). "Methods for the calculation and use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689–1697.
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.
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.
Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
Van Wijngaarden, S. J., and Houtgast, T. (**2004**). "Effect of talker and speaking style on the speech transmission index," J. Acoust. Soc. Am. **115**, 38–41.
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.
Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.

Perception of silent-center syllables by native and non-native English speakers[a)]

Catherine L. Rogers[b)] and Alexandra S. Lopez
Department of Communication Sciences and Disorders, University of South Florida, 4202 E. Fowler Avenue, PCD1017, Tampa, Florida 33620

(Received 8 May 2007; revised 23 April 2008; accepted 7 May 2008)

The amount of acoustic information that native and non-native listeners need for syllable identification was investigated by comparing the performance of monolingual English speakers and native Spanish speakers with either an earlier or a later age of immersion in an English-speaking environment. Duration-preserved silent-center syllables retaining 10, 20, 30, or 40 ms of the consonant-vowel and vowel-consonant transitions were created for the target vowels /i, ɪ, eɪ, ɛ, æ/ and /ɑ/, spoken by two males in /bVb/ context. Duration-neutral syllables were created by editing the silent portion to equate the duration of all vowels. Listeners identified the syllables in a six-alternative forced-choice task. The earlier learners identified the whole-word and 40 ms duration-preserved syllables as accurately as the monolingual listeners, but identified the silent-center syllables significantly less accurately overall. Only the monolingual listener group identified syllables significantly more accurately in the duration-preserved than in the duration-neutral condition, suggesting that the non-native listeners were unable to recover from the syllable disruption sufficiently to access the duration cues in the silent-center syllables. This effect was most pronounced for the later learners, who also showed the most vowel confusions and the greatest decrease in performance from the whole word to the 40 ms transition condition.
 [DOI: 10.1121/1.2939127]

PACS number(s): 43.71.Hw, 43.71.Es [MSS] Pages: 1278–1293

I. INTRODUCTION

To attain "nativelike" proficiency in a second language implies mastering a robustness in speech processing that allows native listeners to adapt to a wide range of listening conditions—from optimal to adverse—that are encountered every day. These daily challenges may result from environmental factors such as noise or reverberation (Miller *et al.*, 1951; Moncur and Dirks, 1967), talker-related factors such as dialect differences or speech impairments (Clopper and Pisoni, 2004; Kent *et al.*, 1990; Hudgins and Numbers, 1942; Monsen, 1983), or linguistic factors such as speaking style (Payton *et al.*, 1994; Lindblom, 1996). Relatively little research on second-language (L2) speech perception has compared native and non-native listeners' responses to these sources of variation in the speech signal (cf., however, Bradlow and Bent, 2002 and Bradlow and Alexander, 2007, with regard to the effects of speaking style on speech perception by non-native listeners).

One area that has been more extensively investigated is the effects of noise on speech perception by native and non-native listeners. There is accumulating evidence that even early learners of a second language, who may speak their L2 with little or no foreign accent and who perform similarly to monolinguals in quiet, may experience greater difficulty than monolingual listeners in processing speech in noise (Mayo *et al.*, 1997; Meador *et al.*, 2000; Rogers *et al.*, 2006).

Both Mayo *et al.* (1997) and Meador *et al.* (2000) studied non-native English speakers with a first language (L1) of Spanish and found that even early learners of English recognized words in noise presented in sentence context less well than monolinguals. Mayo *et al.* (1997) also found that both early bilingual and monolingual listeners, but not later learners of English, recognized words in noise more accurately in sentences with high semantic predictability than in sentences with low semantic predictability.

Rogers *et al.* (2006) compared the recognition of isolated words presented in noise and in reverberation by monolingual English speakers and non-native English speakers with Spanish L1 and an age of onset of immersion in an English-speaking environment of age six or earlier. They also collected accentedness and self-ratings of language dominance for speaking, listening, reading, and writing for the non-native English speakers. Rogers *et al.* (2006) found that even these early learners of English, who were rated as having little or no foreign accent in English and rated themselves as English dominant or balanced for the skills of listening, speaking, reading, and writing, still recognized significantly fewer words in noise and reverberation than did the monolingual English-speaking listeners. It should also be noted that in all three of the above-mentioned studies, the early learners performed similarly to monolinguals in quiet or at very favorable signal-to-noise ratios (SNRs), while proficient non-natives with a later age of immersion have been shown to perform more poorly than monolinguals in quiet or at favorable SNRs of +15 dB or more (cf. Cutler *et al.*, 2004 and Mayo *et al.*, 1997).

[a)]Portions of these data were presented at the 147th meeting of the Acoustical Society of America [J. Acoust. Soc. Am., **115**, 2605 (2004)].
[b)]Electronic mail: crogers@cas.usf.edu

One potential explanation for even relatively early learners' increased difficulty in processing speech in noisy environments is a reduced ability to process speech sounds based on partial information (as may occur in the masking of speech sounds in noisy environments). The hypothesis that learners of a second language may have greater difficulty than monolinguals in identifying speech sounds based on partial acoustic information is compatible with evidence suggesting that some phoneme categories of even experienced non-natives may be intermediate between those of L1 and L2, producing potential mismatches between native and non-native speakers' phoneme categories, perhaps in the form of differences in phoneme boundary locations or differences in cue weighting (cf. Flege, 1995; Imai *et al.*, 2005; and Flege and MacKay, 2004). To investigate the hypothesis that even proficient non-natives may be less able to identify speech sounds based on partial acoustic information, the present study used a silent-center syllable perception task (Strange *et al.*, 1983; Parker and Diehl, 1985) to compare the ability of native and non-native listeners to identify consonant-vowel-consonant (CVC) syllables from which varying degrees of the vowel center had been removed.

The silent-center syllable-identification paradigm typically employs stop-vowel-stop sequences in which the vowel center (as measured from the release of the initial voiced stop to the onset of closure for the final stop) is reduced to silence while the consonant-vowel (CV) and vowel-consonant (VC) transitions are retained unchanged (Strange *et al.*, 1983). Strange *et al.* (1983) found that native listeners could identify target vowels from which most or all of the vowel "steady states" had been removed nearly as well as the full CVC syllable. Strange *et al.* (1983) (cf. also Strange, 1989; Jenkins and Strange, 1999; Jenkins *et al.*, 1999; and Jenkins *et al.*, 1994) interpreted this result to imply that native listeners use dynamic information related to articulatory trajectories into and out of the vowel target for vowel identification, rather than the vowel's target articulatory position (cf., however, Andruski and Nearey, 1992, for an alternative interpretation).

The theoretical interpretation of these data allowed Strange *et al.* (1983) to account for successful vowel identification in the presence of target undershoot. Since then, several studies and authors have used the silent-center paradigm to determine the amount of the vowel center that can be removed before vowel identification is seriously impaired (Parker and Diehl, 1985) and to compare vowel identification performance of different populations of native English speakers (Fox *et al.*, 1992; Sussmann, 2001; Kirk *et al.*, 1992; Murphy *et al.*, 1989). Fox *et al.* (1992), for example, found that older adults' recognition accuracy for silent-center syllables was significantly lower than that of younger adults, despite similar performance for whole words.

The interpretation of the silent-center results of Strange *et al.* (1983) does not, however, necessarily imply that all L1s place an equally strong weight on the dynamic information presented in CV and VC transitions. It may be that dynamic information (either vowel intrinsic or that contained in formant transitions) is particularly important for languages such as English with a relatively crowded vowel space, in which converging sources of acoustic information may be important to differentiate vowels that are close neighbors in phonological space or for vowels that are strongly inherently dynamic (cf. Kewley-Port and Goodman, 2005). Thus, the ability to identify a vowel based on formant transitions alone may not be equally accessible to all learners of English as a second language and a comparison of these abilities between native and non-native listeners may help us to explain how even proficient non-natives' ability to identify speech sounds may be more strongly degraded in noise than that of monolingual listeners.

Therefore, the present study used a silent-center syllable gating paradigm to compare syllable identification from CV and VC transitions by monolingual English speakers and non-native English speakers with a first language of Spanish and either an early or a later age of immersion in an English-speaking environment. Spanish was chosen as the L1 for comparison because Spanish and English differ markedly in their vowel inventories, Spanish having 5 (Dalbor, 1969) and American English approximately 12 stressed vowels, excluding diphthongs (Ladefoged, 1982), and because non-native English speakers with a first language of Spanish constitute a large and rapidly growing minority in the U.S. (approximately 28×10^6 persons at the 2000 Census; United States Census Bureau, 2000). The silent-center paradigm is attractive because it allows for manipulation of the degree of target acoustic information presented without the use of synthesized speech, which may be perceived less accurately by non-native listeners for reasons other than those we wished to investigate, such as the naturalness or appropriateness of the speech cues themselves (cf. Logan *et al.*, 1989 and Hillenbrand and Nearey, 1999).

Specifically, the present study sought to compare the performance of monolingual listeners and two groups of non-native listeners differing in age of onset of immersion in (1) identifying target vowels when varying degrees of the CV and VC transitions were presented and (2) using vowel duration as a cue to vowel identification in silent-center syllables. Six target vowels (/i, ɪ, eɪ, ɛ, æ/ and /ɑ/) that span the vowel space from high to low and occupy a region of the vowel space that is considerably more crowded in English than in Spanish (cf. Dalbor, 1969 and Ladefoged, 1982) were selected.

Performance on the individual target vowels and vowel confusions (vowels selected other than the target vowel) was also compared across the three listener groups. Specific hypotheses about confusion patterns for individual target vowels were not made because the acoustic differences and perceptual assimilation patterns from Spanish to American English vowels have not been extensively investigated and because a wide variety of dialects of Spanish are spoken in the Tampa Bay area, potentially affecting vowel assimilation patterns in different ways. It was hypothesized, however, that the late learners and perhaps the early learners would show greater decreases in performance than the monolingual listeners when vowel duration was not available as a cue. It was also anticipated that the late learners would show the lowest overall performance and the greatest number of confusions for some target vowels, due to more poorly defined vowel

categories, and that different confusion patterns would be observed when vowel duration was available as a cue than when it was not, perhaps for both the early and late learners of English.

II. METHOD

A. Participants

Three groups of participants were recruited: monolingual native English speakers, relatively early learners of English as a second language (age of onset of immersion of 12 years or earlier), and late learners of English as a second language (age of onset of immersion of 18 years or later). Participants were screened to include only persons between the ages of 18 and 50 with no history of speech or hearing disorders. For the monolingual group, potential participants who reported fluency in a second language or exhibited a regional accent that differed strongly from that of the Tampa metropolitan area were also excluded. All potential non-native English-speaking participants were required to be native speakers of Spanish. No speakers of Peninsular varieties of Spanish were included; however, regional variation within New World varieties of Spanish was not controlled. Potential non-native participants who reported fluency in a language other than Spanish or English were also excluded. All participants were also required to pass a basic hearing screening (20 dB hearing level (HL) at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz) prior to participation in the experiment.

Participants were recruited from flyers posted around the campus of the University of South Florida and from newspaper advertisements. Non-native participants were also recruited from among the second author's personal contacts. Prior to participation, all potential participants were required to fill out a language background questionnaire. Forms for both native and non-native English-speaking participants included items related to participants' age, dialect background, history of speech or hearing disorders, and languages spoken. In addition, the form for the potential non-native participants included items probing parents' native languages, age of onset of learning English (AOL), age of onset of immersion in an English-speaking environment (AOI), number of years living in the U.S., and self-ratings of language dominance (more proficient in English or Spanish) for the skills of listening, speaking, reading, and writing.

Participants were compensated by payment of $8.00 per hour of participation, or by an equivalent-value gift certificate. All participants were required to have sufficient proficiency in English to read and understand the consent forms and language background questionnaire, which were printed in English. The second author, a native Spanish speaker with an early age of immersion in an English-speaking environment, was able to assist some participants if they had trouble understanding particular words or phrases, although this assistance was seldom needed.

According to the criteria outlined above, 13 monolingual English speakers (MO), 10 early learners of English as a second language (EL), and 8 late learners of English as a second language (LL) completed the tasks. Data for two monolinguals who did not appear to understand or take the task seriously were removed from the analysis. Data for one additional monolingual speaker were removed from the analysis due to the need to complete counterbalancing of listening conditions across listeners (see below), leaving ten monolinguals whose data were included in statistical analyses.

Although participants of both genders were recruited, volunteers were primarily female, resulting in unbalanced groups (nine females and one male in the MO group, ten females and zero male in the EL group, and five females and three males in the LL group). A literature search revealed several studies showing gender differences in various aspects of speech production and intelligibility, but few that examined the effects of listener gender on speech perception were found. Two studies that indicated no effect of listener gender on the processing of male versus female voices, either behaviorally (Mullenix *et al.*, 1995) or neurophysiologically (Lattner *et al.*, 2005) were found. Thus, when an examination of the individual data for the male listeners did not reveal patterns of performance that were markedly different from the means for the respective groups, data for the male listeners were retained.

Following the listening tasks described below, all participants were recorded in a sound-attenuating booth as they read three English sentences selected from the Harvard sentences, which contain five key words and are semantically appropriate but not highly predictable (IEEE, 1969 and Egan, 1948). All participants were recorded using an Audio-Technica AT4033a microphone. The output of the microphone was routed through a preamplifier and recorded to a Roland VS890 digital recorder at a sampling rate of 44.1 kHz [16 bit analog-to-digital (A/D) converter]. The recorded sentences were digitally transferred to computer and then isolated to file. Each sentence file was amplitude equalized using the rms amplitude of the entire sentence file (which always contained about 10 ms of silence at the beginning and end), in order for the sentences to be presented to the listeners at an approximately equal presentation level.

As part of a related study (Crawford, 2006), 15 adult monolingual English-speaking females heard the three sentences spoken by each of the 28 participants in the present study, presented in random order. The raters were asked to rate each sentence for foreign accentedness on a nine-point scale, with one representing no foreign accent and nine representing a very strong foreign accent. For the present study, the listeners' accentedness ratings were averaged across sentences and listeners to provide a measure of proficiency beyond that provided by the participants' self-ratings of language dominance.

The average ages of the participants in the MO, EL, and LL groups were 26.4, 27.1, and 26.0 years, respectively. The participants ranged in age from 19 to 48 years; the standard deviations (SDs) for the participants' ages were 5.4, 10.1, and 6.5 years for the MO, EL, and LL groups, respectively. The average AOI was 5.6 years (SD=3.3) for the EL group and 25.1 years (SD=6.5) for the LL group. The average number of years in the U.S. was 21.0 (SD=11.1) for the EL group and 1.4 (SD=2.3) for the LL group. Three of the ten EL participants (EL06, EL07, and EL08) were born and

TABLE I. Demographic data for individual participants who were either early (EL) or late (LL) learners of English as a second language. Data are displayed for gender; age, country of origin (of listener or listener's parents if born in the U.S.); age of onset of immersion (AOI); number of years spent living in the U.S.; self-ratings of language dominance (E=English, S=Spanish, and B=balanced) for the skills of listerning, speaking, reading, and writing; and average foreign accentedness ratings from a nine-point scale, with one indicating little or no accent and nine indicating a very strong accent.

Listener	Gender	Age	Country	AOI	Years in U.S.	Listen	Speak	Read	Write	Accent
EL01	F	48	Cuba	5	42	E	E	E	E	2.56
EL02	F	25	Colombia	5	7	S	E	S	S	1.63
EL03	F	23	Mexico	10	13	E	E	E	E	1.40
EL04	F	44	Cuba	8	36	B	E	B	E	2.38
EL05	F	22	Cuba	3	21	E	E	E	E	3.37
EL06	F	21	Cuba	3	21	E	E	E	E	2.33
EL07	F	21	Cuba-Colombia	4	21	S	E	E	E	2.22
EL08	F	25	Cuba	0	25	E	E	E	E	1.46
EL09	F	22	Puerto Rico	8	14	S	B	E	E	3.27
EL10	F	20	Puerto Rico	10	10	E	E	E	E	1.38
LL01	F	23	Peru	22	1	E	E	S	S	6.94
LL02	M	24	Colombia	21	8	S	S	S	S	6.71
LL03	F	41	Colombia	40	1	S	S	S	S	7.77
LL04	F	25	Colombia	25	<1	S	S	S	S	5.70
LL05	F	27	El Salvador	27	<1	S	S	S	S	7.87
LL06	M	23	Nicaragua	23	<1	S	S	S	S	7.03
LL07	M	26	Canary Islands	24	<1	S	S	S	S	6.62
LL08	F	19	Colombia	19	<1	S	E	S	S	5.87

raised in the U.S., but only one (EL08) reported being immersed in an English-speaking environment since birth. The other two participants who were born in the U.S. reported AOIs of 3 and 4 years, but some exposure to English via television and parents' interactions outside the home is likely to have occurred before this age for these two talkers. All three received all of their formal schooling in English. Table I displays the following data for the individual EL and LL participants: (1) age; (2) AOI; (3) country of origin (or country of origin of parents if the listener was born in the U.S.); (4) number of years spent living in the U.S.; [(5)–(8)] self-ratings of language dominance for listening, speaking, reading, and writing; and (9) average ratings of foreign accentedness across the three sentences recorded.

As can be seen from Table I, the EL participants typically rated themselves as English dominant in most domains; eight out of ten EL participants rated themselves as English dominant for reading and nine out of ten rated themselves as English dominant for writing and speaking; only six out of ten rated themselves as English dominant for listening, however. Most of the EL listeners received much or all of their schooling in the U.S., so their self-ratings of English dominance are not surprising, especially for the reading and writing domains. Six of the eight LL participants rated themselves as Spanish dominant in all four domains.

The average accentedness rating was 1.47 (SD=0.22) for the MO participants, 2.20 (SD=0.74) for the EL participants, and 6.87 (SD=0.78) for the LL participants. The range of ratings across participants within each group was 1.22–2.03 for the MO participants, 1.38–3.37 for the EL participants, and 5.70–7.87 for the LL participants. Thus, all of the MO and EL participants received an average accentedness rating in the lower third of the scale (1.0–3.67), indicating little or no foreign accentedness. Furthermore, the scores of the MO and EL participants overlapped substantially, with four of the ten EL participants obtaining accentedness ratings within the range of scores obtained by the MO participants, suggesting a native or near-native degree of proficiency in spoken English. None of the average ratings for the individual LL participants fell within the range of scores obtained by either the MO or EL participants. All of the scores for the LL participants fell in the upper half of the scale (5.0–9.0), indicating at least a moderate to strong degree of foreign accent for all the talkers.

The accentedness rating data also support the retention of data for two participants whose demographic data do not otherwise fit the pattern of data obtained for the participants within their respective groups: EL02 and LL02 (see Table I). Participant EL02 reports Spanish dominance in three of the four domains queried, but obtained an accentedness rating of 1.63, within the range of scores obtained for the native English speakers, suggesting a native or near-native degree of proficiency in speaking English. This participant also obtained an overall score of 100% correct on perception of the whole words, one of the two highest scores for participants in this group, again indicating nativelike proficiency on this task.

On the other hand, participant LL02 reports a much

longer time of residence in the U.S. than the other LL participants, but obtained a mean accentedness rating (6.71) that was just below the group mean of 6.87 and obtained an average perception score of approximately 69% correct on the whole-word identification condition, which was about 10% below the overall average score for the LL group on this condition. These data would suggest that the perception and production skills of this listener do indeed fit with those of the other LL participants, despite his long time of residence in the U.S. Note also that this participant reports only 3 years of immersion in an English-speaking environment, despite a much longer length of residence in the U.S. Such situations are not unusual for Spanish L1 late learners in Florida, where extensive Spanish-speaking communities exist in cities such as Miami and Tampa, underscoring the need for detailed questionnaires if participants' true age of immersion is to be recorded.

B. Stimuli

1. Target words

Target words in /bVb/ context were used, as in Strange *et al.* (1983); however, only the following six target vowels were selected: /i, ɪ, eɪ, ɛ, æ/ and /ɑ/.

2. Speakers and recording procedure

Three monolingual native speakers of American English (two males and one female) with an accent typical of that of persons from the Tampa metropolitan area were recorded saying the target words and nonwords ("beeb, bib, babe, bebb, babb" and "bob") in the following carrier phrase: "I say ———on the tape." Speakers were instructed to speak at their normal speaking rate. Twenty repetitions of each target phrase were recorded. Recordings from the female talker were used to create example and practice stimuli; recordings from the two male talkers were used to create the experimental stimuli.

Talkers were seated in a sound-attenuating booth and read the target sentences from a sheet of paper placed on a stand. A boom-mounted Audio-Technica AT4033 microphone was placed at a 45° angle and approximately 6 in. from the talker's lips. Output from the microphone was routed through an Applied Research and Technology Professional Tube Mic preamplifier and digitized at a 44.1 kHz sampling rate (16 bit A/D converter) using a Roland VS890 digital recorder. Sound files stored on the digital recorder were transferred directly to computer using the digital input of a high-quality sound card (M-Audio Audiophile 2496) and a signal editing software program (COOLEDIT 2000, 2000).

3. Word isolation and creation of whole-word stimuli

Three clear tokens of each of the six target words that were judged to maintain some audible differences from token to token were selected for each of the two male talkers. The female talker's recordings were used only for the creation of example and practice stimuli and therefore only one clear token of each target word was selected for this talker. Target words were isolated as follows. First, the release times of the initial and final /b/ consonants were identified from the wave form, based on the small burst of noise energy associated with the release of the /b/. Next, the 15 ms of energy preceding the initial /b/ release and the 15 ms of energy following the final /b/ release were preserved; all energies preceding (in the case of initial /b/) or following (in the case of the final /b/) these 15 ms buffers were silenced. Linear on-ramping and off-ramping of the first 2 ms of the initial 15 ms buffer and the final 2 ms of the final consonant 15 ms buffer were used to eliminate any clicks associated with the abrupt onset or cessation of energy created by the silencing described above. Thus, up to 15 ms of prevoicing for initial /b/ and a clear release of the final /b/ were preserved, where present. Finally, 15 ms of the silence created at the beginning and at the end of the word were preserved; all other energies were deleted and the token was saved to file.

The resulting 36 sound files for the two male talkers (2 talkers × 3 tokens × 6 target vowels) served as the whole-word stimuli and as the basis for the creation of the silent-center stimuli. Six whole-word sound files were created in the same manner for the female talker. To ensure that all these whole-word stimuli would be presented at approximately the same overall intensity, the root-mean-square (rms) amplitude of each target word file was adjusted to equal 15 dB less than the maximum amplitude using an automated procedure (COOLEDIT 2000, 2000). Stimuli were screened for peak clipping following the amplitude adjustment procedure; no stimuli were found to be peak clipped. Figure 1(A) shows an example "whole word" wave form for the syllable "bebb," spoken by talker 1.

In addition to the times of the onset of release of the initial and final /b/, the time of closure for the final /b/ was measured and used to compute target vowel durations for the two male talkers. Vowel duration was measured as the time from the release of the initial /b/ to the closure for the final /b/ (cf. Strange *et al.*, 1983). The onset of closure for the final /b/ was measured from the wave form by locating the point in time at which voicing ceased or the point in time at which the wave form of the voicing cycles changed from more complex to more sinusoidal, indicating the onset of low-pass filtering created by lip closure.

Fundamental frequency ($F0$) and the frequencies of the first and second formants ($F1$ and $F2$) at vowel midpoint were also measured for the 36 syllables selected as stimuli. All frequency analyses were made by the first author using PRAAT (Boersma and Weeknik, 2006). Values for $F0$ were made using autocorrelation analysis with a pitch range between 100 and 500 Hz in most cases. For five of the stimuli for talker 1, the $F0$ was below or within 5 Hz of the 100 Hz minimum, and the low end of the pitch range was therefore set to 80 Hz to ensure appropriate measurement of the $F0$. Formant values were measured using linear predictive coding (LPC) analysis, with an analysis range 0–5500 Hz, a 20 ms analysis window, and between five and seven formants used for LPC tracking. Formant tracks were overlaid on a wide-band spectrogram of the target syllable, and the number of formants used for tracking was adjusted up or down from a default of six, until a good visual match with the formants observed on the spectrogram was obtained. A

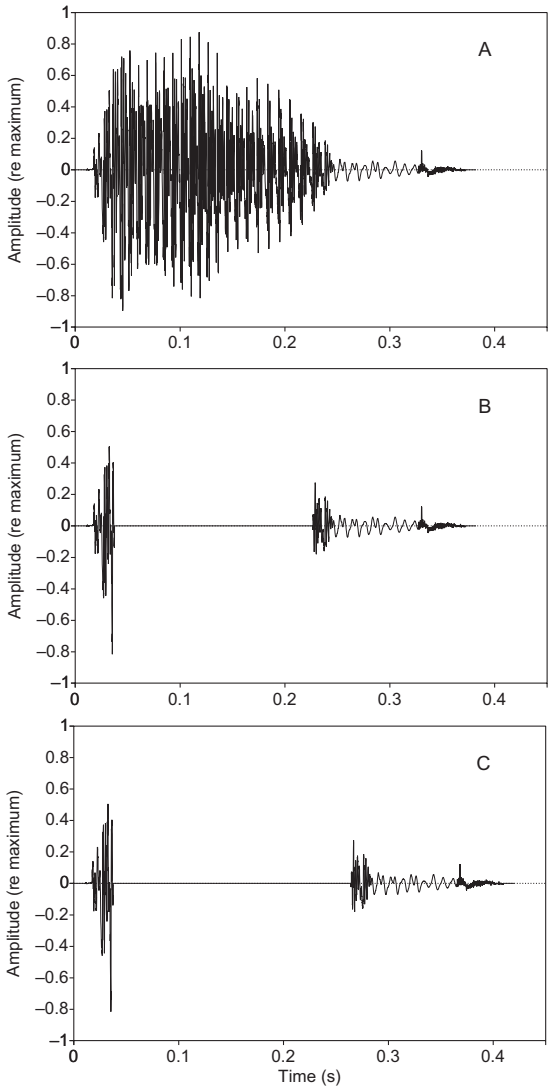


FIG. 1. Wave forms showing an example of the target syllable "bed," spoken by talker 1 and edited to create the following three listening conditions: whole-word (A), 20 ms DP (B), and 20 ms DN (C).

good visual match was obtained in all cases and no analysis by hand of *F*0, *F*1, or *F*2 values was judged to be needed.

Table II shows the average vowel duration, *F*0, *F*1, and *F*2 values for each talker and target vowel, as well as the average across talkers and the average across vowels. These durations indicate a somewhat slow rate of speech for these two male talkers. The *F*0, *F*1, and *F*2 values are largely compatible with values expected for these vowels spoken in General American English for a male talker, although the *F*0 for talker 2 is somewhat higher than average.

4. Creation of duration-preserved silent-center stimuli

Four silent-center versions of each target word were created, with 10, 20, 30, and 40 ms of the initial CV and final VC information preserved. These versions or "gates" were selected based on pilot testing that showed a ceiling effect for gates of 35 ms or greater (15, 25, 35, and 45 ms gates were tested). Silent-center tokens were created by first selecting the desired gate duration (e.g., 20 ms) immediately following the release of the initial /b/. Next, the closure of the final /b/ was located and the desired gate duration (e.g., 20 ms) prior to the release of the final /b/ was selected. All energies between the initial and final selections were then silenced. The edges of the initial and final preserved portions of the syllable were then off-ramped and on-ramped, respectively, using a 2 ms linear ramp, as described above for the initial and final portions of the word. This process was repeated for each of the four desired gates, resulting in 144 additional stimuli (2 talkers $\times$ 3 tokens $\times$ 6 target vowels $\times$ 4 gates). These were termed the "duration-preserved (DP)" silent-center stimuli because no modification of the duration of the silent center was performed. Figure 1(B) shows an example of a 20 ms DP silent-center wave form for the syllable "bebb," spoken by talker 1.

Six additional DP silent-center stimuli (one for each target word) were created from the female talker's utterances, using the 30 ms silent-center gate only. These were used as practice stimuli.

5. Creation of duration-neutral silent-center stimuli

To create the duration-neutral (DN) silent-center tokens, the vowel duration of each of the 144 DP silent-center stimuli was adjusted to equal the average vowel duration across all tokens of the target words spoken by the two male talkers (267 ms). This was done by noting the measured duration of the vowel in question (from initial /b/ release to final /b/ closure) and then inserting or deleting the appropri-

TABLE II. Mean duration (in s) and *F*0, *F*1, and *F*2 values (in Hz) for each of the six target vowels, with the average across vowels for each talker and the average across talker for each vowel.

Target vowel	Talker 1				Talker 2				Average across talkers			
	Dur	*F*0	*F*1	*F*2	Dur	*F*0	*F*1	*F*2	Dur	*F*0	*F*1	*F*2
/i/	0.291	120	243	2327	0.293	184	349	2507	0.292	152	296	2417
/ɪ/	0.222	125	407	1803	0.183	167	506	1993	0.202	146	457	1898
/eɪ/	0.323	126	386	2100	0.275	170	493	2473	0.299	148	440	2286
/ɛ/	0.237	110	586	1600	0.216	171	640	1907	0.227	141	613	1753
/æ/	0.285	103	718	1595	0.280	165	724	1987	0.282	134	721	1791
/ɑ/	0.295	101	677	1087	0.309	172	766	1323	0.302	137	721	1205
Average across vowels	0.276	114	503	1752	0.259	171	580	2032	0.267	143	541	1892

ate duration of silence in the silent-center portion so that the resulting vowel duration was equal to the average vowel duration across the two male talkers. This procedure resulted in an additional 144 DN silent-center stimuli. No DN silent-center stimuli were created from the female talker's utterances. Figure 1(C) shows an example of a 20 ms DN silent-center wave form for the syllable "bebb," spoken by talker 1. Prior to presentation to listeners, the sample rate of all stimuli was adjusted to 48.8 kHz to accommodate the software program used for stimulus presentation (ECOS/WIN, 1999).

C. Main experiment procedure

1. Listening environment

Listeners were seated in groups of up to four in a quiet, sound-treated room with individual carrels for each listener. Each carrel was separated by a divider and an empty carrel separated each listener from the next. Each listener's carrel was equipped with a flat-screen monitor, keyboard, and mouse. The CPU controlling each independent listening station was located outside the room. The stimuli were presented binaurally over headphones (Sennheiser HD265) at approximately 70 dB sound pressure level (SPL) (based on the whole-word stimuli). Presentation level was controlled using the programmable attenuators (PA5) of the Tucker-Davis Technologies (TDT) Psychoacoustics System III (2001) hardware.

2. Calibration

A 1000 Hz tone was used for calibration of stimulus presentation level. The rms amplitude of the tone was adjusted to match that of the whole-word stimuli (15 dB below maximum amplitude), prior to the calibration procedure. During calibration, the tone was played out without attenuation through the headphones. As the tone was played out, the right and left headphones were placed in turn onto the coupler of a sound-level meter (Bruel & Kjaer Model 2235) and the level of input to the sound-level meter was measured in dB SPL and noted. The amount of attenuation needed for a presentation level of 70 dB SPL (average across the two headphones) was then computed and the measurement procedure was repeated with this setting to confirm a presentation level of 70 dB. The attenuation levels of the PA5s for the four listening stations were adjusted accordingly within the software used for presentation of stimuli (ECOS/WIN, 1999).

3. Listener task

Prior to beginning the experiment, the listeners were familiarized with the pronunciation and spelling used for all of the target words and nonwords. For all trials (practice and main experiment), one of the six target words was presented over the headphones and six alternatives were displayed within boxes on the screen (two rows and three columns) in the following order (clockwise): "beeb, bib, babe, bebb, babb" and "bob." To focus the listeners' attention on the target vowels and to ensure accurate reading of the nonwords, a common word with the same vowel as the target word was displayed on the screen below each target word ("feed, crib, tape, red, crab" and "dog"). Listeners were informed that the more common words were displayed for reference purposes only and that the word presented would always be one of the target (/bVb/) words; they all reported familiarity with the pronunciation of the key words used.

Listeners were verbally familiarized with the task prior to participation. In addition, written instructions were provided on the monitor prior to each set of trials (example, practice, silent center, and whole word). On each trial, listeners were instructed to choose which word they had heard and responded by left clicking with the mouse. Listeners were informed that some items would be more difficult than others and to make their best guess if unsure. The order of presentation of all stimuli and collection of listener responses were controlled automatically using ECOS/WIN (1999). All trials were self-paced; the next item was presented approximately 1 s following the listener's response.

4. Example and practice trials

The six whole-word stimuli created for the female talker were used to create six example trials. On these trials, all six whole-word stimuli were presented in the following order: "beeb, bib, babe, bebb, babb" and "bob." Listeners were informed of the order prior to the beginning the task. For the example trials, the correct response was highlighted in green following the listener response in order to provide visual reinforcement to the listener. Six silent-center *practice* trials were created using six 30 ms gate silent-center stimuli created from the utterances of the female talker. Feedback was not provided on these trials and the words were presented in random order.

5. Main experiment trials

Listeners completed three blocks of trials in the main experiment: 144 DP trials, 144 DN trials, and 36 whole-word trials. To control for any practice effects on the silent-center trials, half of the listeners in each group completed the DP trial block first and half of the listeners completed the DN trial block first. All listeners completed the whole-word trials last, to avoid overfamiliarization with the stimuli prior to presentation of the silent-center trials. Stimuli were presented in random order in all three blocks and no feedback was provided. A required 5 min break was provided prior to the whole-word block; listeners were allowed to take a break following any trial block. The entire experiment, including completion of forms, hearing screening, and all experimental trials, took between 1 and 1.5 h.

The ECOS/WIN program (1999) automatically scored listener responses as correct or incorrect and recorded information on the alternative chosen on each trial. These data were imported to a spreadsheet and the number of correct responses was computed for each target vowel at each of the nine listening conditions (2 listening conditions $\times$ 4 gates + the whole-word condition). Confusion matrices showing the number of items correct and the alternatives chosen for incorrect responses were computed for each listener for both the DP and DN conditions. Prior to data analysis, percent-

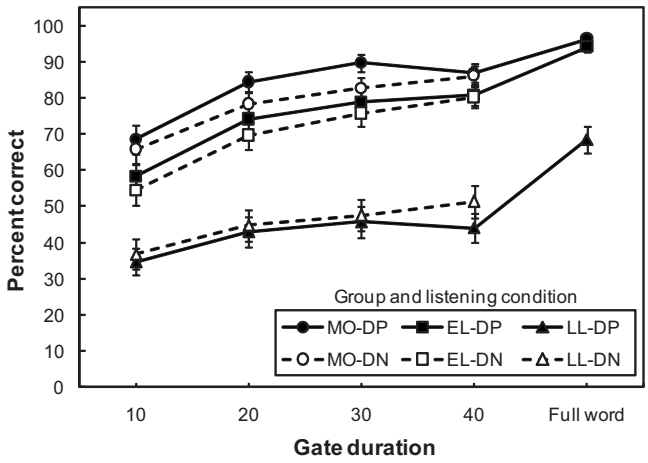


FIG. 2. Percent-correct identification performance by listener group and listening condition, averaged across target vowels. The solid lines with filled symbols indicate the performance on duration-preserved (DP) conditions and the dashed lines with open symbols indicate the performance on duration-neutral (DN) conditions. The performance of monolingual listeners (MO) is indicated by circles, the performance of early learners of English as a second language (EL) is indicated by squares, and the performance of late learners (LL) is indicated by triangles. Error bars indicate one standard error of the mean.

correct scores were converted to rationalized arcsine transform unit (RAU)-transformed scores, which correct for correlation of variances with the mean that can occur when proportional data are used, help to correct for ceiling and floor effects, and yield values that are reasonably interpretable with respect to the corresponding percent-correct scores (Studebaker, 1985).[1]

III. RESULTS

A. Whole-word performance

Figure 2 displays percent-correct performance for each listener group on each of the nine conditions (2 duration conditions × 4 gates + the whole-word condition), averaged across target vowels. As shown in the figure, the monolingual (MO) and early learners of English as a second language (EL) listeners performed nearly identically and nearly perfectly on the whole-word condition (96% and 94%, respectively, or 100 and 97 RAU); the performance of the late learners (LL) on the whole-word condition was about 27% (or 30 RAU) below that of the other two listener groups (approximately 69% correct or 68 RAU).[2]

While the overall performance of the LL listeners on the whole-word condition is substantially poorer than that of the other two groups, it is also about 52% above chance performance (17%) for this six-alternative forced-choice task. Thus, while their phonetic categories for the target vowels may be less well formed than those of the other two groups, these data do suggest that the whole-word stimuli were reasonably well identified by the majority of the LL listeners.

B. Effects of partial vowel information (whole-word versus 40 ms duration-preserved syllables)

Despite the lower overall performance of the LL group, performance appears to differ more between the whole-word to the 40 ms DP condition for this group (about a 25% or 24 RAU difference in performance between the whole-word and the 40 ms DP condition) than for the MO and EL groups (approximately 9% and 13% or 11 and 15 RAU differences in performance, respectively). To examine just the effects of presenting partial vowel information on performance for the three listener groups, a three-way mixed design analysis of variance (ANOVA) was performed using SPSS (2006), with listener group (three levels) as the between-subjects variable and listening condition (whole word vs. 40-ms DP) and target vowel (six levels) as the within-subjects variables. RAU-transformed (Studebaker, 1985) percent-correct identification performance was the dependent variable. The whole-word condition could not be included in a larger ANOVA with the effects of gate and listening condition because there was no DN whole-word condition. Therefore, this limited analysis was considered the best way of determining whether removing *any* vowel information at all affected one group more than another.

Partial eta-squared (η^2_P) effect-size statistics provided by SPSS (2006) were converted to generalized eta-squared (η^2_G) values, using the formulas provided by Bakeman (2005). According to Bakeman (2005), comparisons of effect sizes between studies with between-subjects variables and studies with any within-subjects variables are not appropriate when η^2_P effect-size values are used because η^2_P can be much larger in within-subjects or mixed designs than in between-subjects designs showing similar effect sizes for a given factor. Bakeman (2005) recommended η^2_G as a measure of effect size that is interpretable as the proportion of variance in the dependent variable accounted by the effect, that is appropriate for designs with within-subjects variables, and that allows for comparisons across studies with between-subjects, within-subjects, and mixed designs.

As recommended by Bakeman (2005), type I sum of squares values were used in the ANOVA and the computation of effect sizes and power, due to the unequal group sizes in the between-subjects variable. Table III shows values of F, degrees of freedom, η^2_P, η^2_G, and power for each of the main effects and interactions, along with the classification of effect sizes as negligible, small, medium, or large, according to the guidelines suggested by Bakeman (2005).

As can be seen from Table III, the three-way ANOVA showed significant main effects of listener group, listening condition, and target vowel; based on η^2_G values, the effect sizes of these main effects were categorized as large, medium, and small, respectively. Two interactions were significant: target vowel by listening condition and listener group by target vowel by listening condition; both of these effects were categorized as small. Statistical power reached the generally accepted criterion level of 0.8 or above for all but two effects: the two-way interaction between listener group and listening condition and the two-way interaction between listener group and target vowel. The effect size for the two-way interaction between listener group and listening condition was classified as negligible, while that for the two-way interaction between listener group and target vowel fell at the low end of the range classified as small. Thus, the only two nonsignificant interactions that were substantially underpowered

TABLE III. F values, degrees of freedom, p values, power and effect size data for the three-way ANOVA examining the effects of listener group, listening condition (whole word vs DP 40 ms silent-center) and target vowel on RAU-transformed percent-correct syllable identification performance. Both partial eta-squared (η^2_P) and generalized eta-squared (η^2_G) effect size statistics are provided, as well as the recommended classification for η^2_G (cf. Bakeman, 2005). Bold type indicates effects that reached significance.

Effect	F	df	p	Power	η^2_P	η^2_G	η^2_G classified as
			Main effects				
Listener group	**48.50**	**2,23**	**<0.0005**	**1.00**	**0.81**	**0.40**	**Large**
Listening condition	**64.12**	**1,23**	**<0.0005**	**1.00**	**0.74**	**0.09**	**Small**
Target vowel	**13.51**	**5,115**	**<0.0005**	**1.00**	**0.37**	**0.13**	**Medium**
			Two-way interactions				
Listener group X listening condition	1.89	2,23	0.175	0.35	0.14	0.01	Negligible
Listener group X target vowel	0.80	10,115	0.625	0.40	0.07	0.02	Small
Listening condition X target vowel	**6.99**	**5,115**	**<0.0005**	**1.00**	**0.23**	**0.05**	**Small**
			Three-way interaction				
Listener group X listening condition X target vowel	**2.11**	**10,115**	**0.029**	**0.88**	**0.16**	**0.03**	**Small**

were also classified as negligible to small, suggesting that no important effects failed to reach significance, despite the relatively small number of participants in each of the three listener groups.

A Tukey HSD *post hoc* analysis of the main effect of group showed no significant difference in the performance of the MO and EL listener groups ($p=0.333$) and significantly lower performance for the LL group than for the other two groups ($p<0.0005$ for both comparisons). On average, the EL listeners identified the target syllables about 4% (or 4 RAU) less accurately than the MO listeners and the LL listeners identified the target syllables about 32% (or 32 RAU) less accurately than the EL listeners.

An examination of the main effect of target vowel showed that the six target vowels were perceived in the following order, from most to least accurately perceived, across the groups, and listening conditions: /ɑ/ (94% or 97 RAU), /æ/ (89% or 91 RAU), /eɪ/ (86% or 88 RAU), /i/ (81% or 83 RAU), /ɛ/ (74% or 75 RAU), and /ɪ/ (71% or 71 RAU). Simple main effects of comparisons with Bonferroni adjustment for the 15 comparisons among pairs of target vowels showed that target /ɑ/ was perceived significantly more accurately than all of the other vowels except /æ/; target /æ/ was perceived significantly more accurately than /ɪ/ and /ɛ/ but not differently from /i/ or /eɪ/; target /eɪ/ was perceived significantly more accurately than /ɪ/, but not differently from /i/ or /ɛ/; performance for target /i,ɛ/ and /ɪ/ did not differ significantly. This order did not differ dramatically in the interactions and will not be discussed further in this section.

Figure 3 compares performance across the levels of the three factors in the significant three-way interaction: listener group, target vowel, and listening condition (whole word versus DP 40 ms). Performance for each listener group is shown as a separate panel. The three-way interaction was explored by pairwise comparisons of listeners' performance on the two listening conditions at each level of group and target vowel and by pairwise comparisons of the performance of the three listener groups at each level of listening condition and target vowel. Bonferroni adjustment for the number of comparisons at each level was used.

As shown in the figure, similar patterns of performance were obtained for the MO and EL listener groups. Performance differed significantly between the whole-word and the DP 40 ms conditions for the target vowels /ɪ/ and /ɛ/ for both the MO and EL groups [by about 20%–28% or 23–31 RAU; see Figs. 3(A) and 3(B)]. For target /i/, performance for the MO but not the EL group differed significantly between the whole-word and the DP 40 ms conditions (about a 13% or 16 RAU difference for the MO listeners and 10% or 11 RAU for the EL listeners), while for target /eɪ/ performance for the EL but not the MO group differed significantly between the whole-word and the DP 40 ms conditions (a 12% or 15 RAU difference for the EL group but only 2% or 3 RAU for the MO group). Performance for target /æ/ and /ɑ/ did not differ significantly between the whole-word and the DP 40 ms conditions for either the MO or the EL listener group (with differences of at most 2% or 4 RAU in each case).

Performance for the LL group was lower than for the other two groups and typically more variable, even on the whole-word condition [see Fig. 3(C)]. Performance for the LL group differed significantly between the whole-word and the DP 40 ms conditions for the target vowels /i,ɪ/ and /ɑ/, for which performance differed by 56%, 36%, and 11% or 51, 34, and 11 RAU, respectively between the two conditions.

Comparisons of group within each level of listening condition and target vowel revealed relatively minor differences. First, the performance of the MO and EL groups did not differ significantly for any target vowel, although the

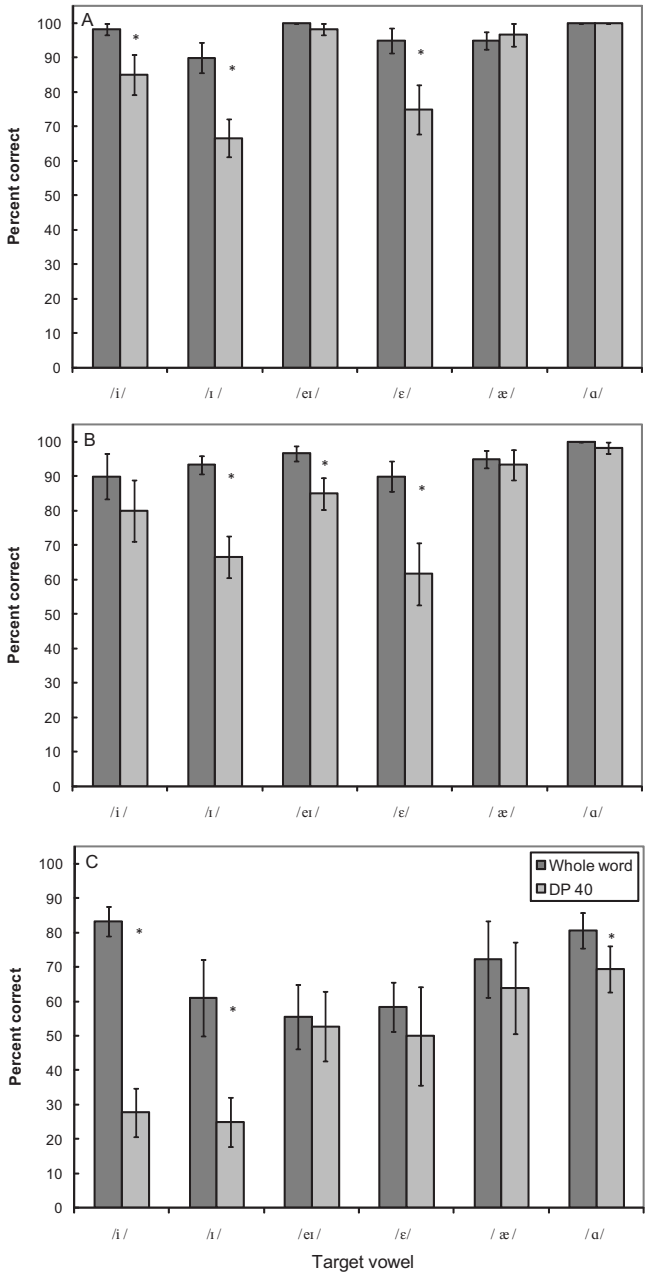


FIG. 3. Percent-correct identification performance by vowel and information condition, for the whole-word and DP 40 ms conditions and each of the three listener groups: monolinguals (A), early learners (B), and late learners (C). In each panel, darker gray bars indicate the whole-word condition and lighter gray bars indicate the DP 40 ms condition. Error bars indicate one standard error of the mean, and asterisks indicate conditions that differed significantly in performance within each target vowel.

difference approached significance for target /eɪ/ in the DP 40 ms condition ($p=0.064$), in which performance for the MO group was about 13% or 18 RAU higher than the performance for the EL group [see Figs. 3(A) and 3(B)]. Second, the performance of both the MO and EL groups was significantly higher than that of the LL group for all target vowels and conditions, except target /ɛ/ in the DP 40 ms condition and target /i/ in the whole-word condition, for which no group differences were significant [see Figs. 3(A)–3(C)]. The difference in the performance between the MO and LL listeners did approach significance for target /i/ in the whole-word condition ($p=0.056$), in which the performance for the MO group was about 15% or 22 RAU higher than that of the LL group. The performance of the LL listener group ranged from 19% to 57% (or 23 to 57 RAU) lower than that of the other two listener groups for the target vowels showing significant differences. Together, these data show that the EL listeners were able to perform as well as the MO listeners when the whole syllable was provided, or when the complete CV and VC transitions were available. The LL listeners performed more poorly overall and showed overall larger decreases in performance from the whole-word to the DP 40 ms condition than did the other two groups.

C. Effects of group, gate, duration neutralization, and target vowel

To address the main research question regarding the effects of varying transition information on vowel identification, a four-way mixed design ANOVA was used to compare the effects of gate, duration neutralization, and target vowel across listener groups. Listener group (three levels) was the between-subjects variable and gate (four levels), and duration condition (DP versus DN) and target vowel (six levels) were within-subjects variables. As in the first ANOVA, RAU-transformed percent-correct identification performance was the dependent variable; η^2_P values were converted to η^2_G values; and type I sum of squares values were used in the computation of effects, effect sizes, and power. Table IV shows values of F, degrees of freedom, η^2_P, η^2_G, and power for each of the main effects and interactions, along with the classification of effect sizes as negligible, small, medium, or large, according to the guidelines suggested by Bakeman (2005).

As shown in Table IV, the four-way ANOVA showed significant main effects of listener group, target vowel, and gate; based on η^2_G values, the effect sizes of these main effects were categorized as large, medium, and small, respectively. The main effect of duration condition was not significant. Significant two-way interactions were found for the listener group by duration condition, the listener group by gate, and gate by target vowel; all three of these effects were categorized as small or negligible in size (see Table IV). No other interactions were significant, although the listener group by gate by target vowel interaction approached significance ($p=0.061$). Statistical power failed to reach the generally accepted criterion level of 0.8 or above for several effects or interactions, but the effect size for all but one was categorized as negligible. The remaining underpowered interaction (listener group by target vowel) approached significance ($p=0.098$) but its effect size fell at the low end of the range classified as small. Thus, as in the first ANOVA, all of the nonsignificant interactions that were substantially underpowered were also classified as negligible to small, suggesting that no important effects failed to reach significance, despite the relatively small number of participants in each of the three listener groups.

Unlike in the analysis comparing performance on the whole-word and DP 40 ms conditions, a Tukey HSD *post hoc* analysis of the main effect of group showed significant

TABLE IV. F values, degrees of freedom, p values, power and effect size data for the four-way ANOVA examining the effects of listener group, gate, duration condition (DP vs DN syllables), and target vowel on RAU-transformed percent-correct syllable identification performance. Both partial eta-squared (η^2_P) and generalized eta-squared (η^2_G) effect size statistics are provided, as well as the recommended classification for η^2_G (cf. Bakeman, 2005). Bold type indicates effects that reached significance.

Effect	F	df	p	Power	η^2_P	η^2_G	η^2_G classified as
			Main effects				
Listener group	**54.12**	**2,25**	**<0.0005**	**1.00**	**0.81**	**0.36**	**Large**
Duration condition	2.22	1,25	0.149	0.30	0.08	0.00	Negligible
Target vowel	**26.17**	**5,125**	**<0.0005**	**1.00**	**0.51**	**0.22**	**Medium**
Gate	**76.71**	**3,75**	**<0.0005**	**1.00**	**0.75**	**0.07**	**Small**
			Two-way interactions				
Listener group X duration condition	**4.20**	**2,25**	**0.027**	**0.68**	**0.25**	**0.00**	**Negligible**
Listener group X target vowel	1.66	10,125	0.098	0.77	0.12	0.03	Small
Listener group X gate	**2.80**	**6,75**	**0.017**	**0.86**	**0.18**	**0.01**	**Negligible**
Duration condition X target vowel	1.97	5,125	0.088	0.65	0.07	0.00	Negligible
Duration condition X gate	2.29	3,75	0.085	0.56	0.08	0.00	Negligible
Gate X target vowel	**6.82**	**15,375**	**<0.0005**	**1.00**	**0.21**	**0.03**	**Small**
			Three-way interactions				
Listener group X duration condition X target vowel	1.22	10,125	0.282	0.61	0.09	0.00	Negligible
Listener group X duration condition X gate	0.23	6,75	0.964	0.11	0.02	0.00	Negligible
Listener group X gate X target vowel	1.46	30,375	0.061	0.98	0.10	0.01	Negligible
Duration condition X gate X target vowel	0.61	15,375	0.868	0.40	0.02	0.00	Negligible
			Four-way interaction				
Listener group X duration condition X gate X target vowel	1.38	30,375	0.093	0.97	0.10	0.01	Negligible

differences in performance between all three groups ($p=0.032$ for MO versus EL and $p<0.0005$ for LL versus MO and EL). Overall, the MO listeners identified the target syllables about 8% (or 10 RAU) more accurately than the EL listeners, and the EL listeners identified the target syllables about 28% (or 29 RAU) more accurately than the LL listeners. Thus, overall, even the EL listeners were found to identify the syllables significantly more poorly than the monolingual listeners when partial vowel information was provided. The main effect of gate will not be discussed separately because it was changed by its interactions with other variables.

Similar to the first analysis, an examination of the main effect of target vowel showed that the six target vowels were perceived in the following order of accuracy, from highest to lowest, across the groups, and listening conditions: /ɑ/ (87% or 90 RAU), /æ/ (78% or 79 RAU), /eɪ/ (73% or 74 RAU), /i/ (57% or 58 RAU), /ɪ/ (53% or 53 RAU), and /ɛ/ (51% or 52 RAU). Pairwise comparisons using Bonferroni adjustment for the number of comparisons among the target vowels revealed significantly higher performance for target /ɑ/ than all of the other target vowels except /æ/ and significantly higher performance for targets /æ/ and /eɪ/ than for /i,ɪ/ and /ɛ/ (p values ranged from <0.0005 to 0.004 for the significant comparisons). Overall listener performance did not differ significantly between the target vowels /æ/ and /eɪ/ or among the target vowels /i,ɪ/ and /ɛ/. This order of performance for the target vowels did not differ dramatically across gates in the significant target vowel by gate interaction and therefore will not be discussed further in this section.

1. Group by duration condition effect

To address the question of whether the groups differed in their ability to benefit from duration information in identification of silent-center syllables, pairwise comparisons of performance on the DP and DN conditions at each level of the listener group variable were used to examine the significant listener group by duration neutralization interaction. Contrary to our hypothesis, only the MO listener group showed significantly higher performance (by about 4% or 4 RAU; $p=0.039$) for the DP than for the DN condition. Although the performance of the EL group was also higher in the DP than in the DN condition by a similar amount (by about 3% or 4 RAU), the difference only approached significance for this group ($p=0.087$). The performance of the LL group was *lower* in the DP than in the DN condition (by

about 3% or 4 RAU overall), but this difference was not significant ($p=0.113$).

A comparison of groups at each level of the listening condition variable was also made, using Bonferroni adjustment for the number of comparisons among groups at each level of duration neutralization. All groups differed significantly from one another in their performance on the DP condition. The MO listeners identified the target vowels significantly more accurately than the listeners in the other two groups ($p=0.05$ for MO versus EL and $p<0.0005$ for MO versus LL), and the EL listeners identified the target vowels significantly more accurately than the listeners in the LL group ($p<0.0005$). In the DN condition, listeners in both the MO and EL groups identified the target vowels significantly more accurately than the listeners in the LL group ($p<0.0005$ in both cases), but the MO and EL listener groups did not perform significantly differently from one another, although the difference did approach significance ($p=0.066$). In both the DP and DN conditions, the difference in performance between the MO and EL listener groups was 9% (or 10 RAU) or less, while the difference in performance between the LL listener group and the other two groups ranged from 25% to 40% (or 25 to 42 RAU) across the two listening conditions. In summary, the MO listeners were able to benefit to some degree from the vowel duration cues provided in the DP condition, but a similar magnitude benefit for the EL listeners failed to reach significance. The LL listeners showed no evidence that they were able to benefit from the vowel duration cues provided in the DP condition.

2. Group by gate effect

Figure 2 shows percent-correct performance for each listener group as a function of increasing gate duration. As shown in the figure, performance for the MO and EL groups increases by about 15% from the 10 to the 20 ms gate condition and to a lesser degree (about 4% and 9%, respectively) from the 20 to the 40 ms gate condition. Performance for the LL group, on the other hand, improves by only about 8% from the 10 to the 20 ms gate, and improves by about another 5% from the 20 to the 40 ms gate (when averaged across DP and DN). Thus, the EL group's vowel identification performance improves the most across the four gates (about 24%) and the LL group's performance improves the least (about 13%).

To determine the specific gates at which group differences were found, pairwise comparisons of performance were made between the groups at each level of the gate variable in the significant group by gate interaction. Bonferroni adjustment for the number of group comparisons at each level of gate was used. The MO listener group identified the syllables significantly more accurately than the EL group (by about 11% or 11 RAU) for the 10 ms gate condition only ($p=0.007$), although similar magnitude differences did not reach significance for the 30 and 20 ms gate conditions ($p=0.066$ and $p=0.117$, respectively). Both the MO and EL listener groups identified the syllables significantly more accurately than the LL group (by 21%–40% or 21–42 RAU) at all four gate conditions ($p<0.0005$ in all cases).

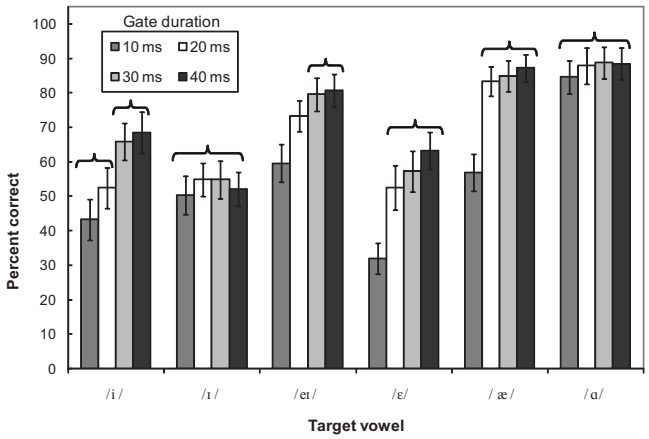


FIG. 4. Percent-correct identification performance by target vowel and gate, averaged across listener groups and duration conditions (DP and DN). Performance at the 10, 20, 30, and 40 ms gates is indicated by medium gray, white, light gray, and dark gray bars, respectively. Error bars indicate one standard error of the mean, and braces indicate sets of conditions for which performance did not differ significantly.

Pairwise comparisons of performance across the gates within each level of the listener group were used to determine the gate(s) for which performance differed significantly for each group. Again, Bonferroni adjustment for the number of comparisons at each level of group was used. All three listener groups showed similar patterns of significant effects across the gates. For the MO listener group, performance for the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate condition (by 14%–19% or 14–20 RAU; $p<0.0005$ in each case), but performance did not differ significantly among the 20, 30, and 40 ms gate conditions. For the EL listener group, performance on the 40 ms gate condition was significantly higher than the performance in the 10 and 20 ms conditions (by 9% and 24% or 9 and 24 RAU, respectively; $p=0.007$ and $p<0.0005$, respectively). Performance on the 20 and 30 ms gate conditions was also higher than the performance on the 10 ms condition for the EL listener group (by 21% and 16% or 21 and 15 RAU, respectively; $p<0.0005$ in both cases). The difference in performance between the 30 and 20 ms gate conditions (about 5% or 6 RAU) also approached significance ($p=0.061$) for the EL group. For the LL listener group, performance on the 30 and 40 ms gate conditions was significantly higher than the performance on the 10 ms gate condition (by 12% and 11% or 11 and 10 RAU, respectively; $p<0.0005$ and $p=0.005$, respectively), but no other comparisons reached significance.

3. Target vowel by gate effect

Although the three-way interaction between group, gate, and target vowel was not significant, the significant target vowel by gate interaction indicates that the six vowels exhibited different patterns of performance across the gates, but that this pattern did not differ significantly across the groups. Thus, Fig. 4 shows the performance for each target vowel on each of the four gate conditions and averaged across the DP and DN conditions *and* across listener groups. Pairwise comparisons of performance with Bonferroni adjustment for the

TABLE V. Percent-correct performance for each listener group on each target vowel for the DP and DN conditions, rounded to the nearest whole percentage. For each target vowel, the most frequent (i.e., those with greater than 5%) confusions, or vowels chosen instead of the target vowel are shown, with the confusion percentage in parentheses.

		DP		DN	
Listener group	Target vowel	Percent correct	Confusions (pct)	Percent correct	Confusions (pct)
MO	/i/	78	/eɪ/ (17)	81	/eɪ/ (15)
	/ɪ/	72	/eɪ/ (16), /ɛ/ (12)	59	/eɪ/ (28), /ɛ/ (13)
	/eɪ/	92	/ɛ/ (5)	91	/ɪ/ (5)
	/ɛ/	65	æ/ (26), /ɪ/ (5)	51	/æ/ (38), /ɪ/ (9)
	/æ/	89	/ɛ/ (8)	88	/ɛ/ (10)
	/ɑ/	99		100	
EL	/i/	60	/eɪ/ (20), /ɪ/ (18)	51	/ɪ/ (25), /eɪ/ (21)
	/ɪ/	65	/eɪ/ (20), /ɛ/ (11)	63	/eɪ/ (25), /ɛ/ (10)
	/eɪ/	76	/ɪ/ (12), /ɛ/ (8)	74	/ɪ/ (16), /ɛ/ (10)
	/ɛ/	57	æ/ (24), /ɪ/ (13)	52	/æ/ (30), /ɪ/ (12)
	/æ/	85	/ɛ/ (8)	85	/ɛ/ (8)
	/ɑ/	96		96	
LL	/i/	33	/ɪ/ (51), /eɪ/ (9), /ɛ/ (7)	32	/ɪ/ (51), /eɪ/ (9), /ɛ/ (6),
	/ɪ/	26	/ɛ/ (27), /eɪ/ (26), /i/ (18)	23	/eɪ/ (29), /ɛ/ (24), /i/ (20)
	/eɪ/	46	/ɛ/ (19), /ɪ/ (18), /i/ (15)	51	/ɛ/ (22), /ɪ/ (14), /i/ (12)
	/ɛ/	35	/æ/ (34), /eɪ/ (15), /i/ (6), /ɪ/ (6)	42	/æ/ (21), /eɪ/ (20), /i/ (12)
	/æ/	55	/ɛ/ (21), /ɑ/ (9), /eɪ/ (7)	58	/ɛ/ (22), /ɑ/ (10), /eɪ/ (8)
	/ɑ/	57	/æ/ (40)	66	/æ/ (29)

number of comparisons across the gates within each level of target vowel were used to explore the significant interaction between target vowel and gate. Performance did not differ significantly between any gates for targets /ɑ/ and /ɪ/, which showed the highest and second lowest rates of correct identification, respectively. As shown in the figure, performance is relatively flat across the four gates for these two target vowels.

For target /æ/, which showed the second highest overall rate of correct identification, performance on the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate (by 26%–30% or 27–31 RAU; $p<0.0005$ in all three cases), but performance did not differ significantly across the three longer gates. Similarly, for target /eɪ/, performance on the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate (by 14%–21% or 12–20 RAU; $p<0.0005$ to $p=0.032$), but performance on the 40 ms gate was also significantly higher than the performance on the 20 ms gate (by 7% or 9 RAU; $p=0.025$); no other gates differed significantly for this vowel. For target /i/, performance on the 30 and 40 ms gates was significantly higher than the performance on both the 10 and 20 ms gates (by 15–25 RAU or 13%–25%; $p<0.0005$ to $p=0.013$); no other gates differed significantly for this vowel. For target /ɛ/, for which the lowest overall level of performance was obtained, performance on the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate (by 19–30 RAU or 21%–31%; $p<0.0005$ to $p=0.02$); no other gates differed significantly for this vowel (see Fig. 3).

D. Confusion analyses

The analysis of the confusion matrices was used to compare the groups in terms of the identity and number of vowels they perceived other than the target vowel in both the DP and DN listening conditions. Table V summarizes percent-correct performance and percent confusions for each target vowel, averaged across the four gates. Results are shown separately for each listener group and listening condition (DP versus DN). Within each group and for each target vowel, a percentage is given for each confusion vowel (vowel selected other than the target vowel) that received greater than 5% of listener responses for each target vowel.

As anticipated, the LL group showed a greater number of confusions than the other two groups; one to two more confusion vowels exceeded 5% confusions than for either the MO or EL group for all six target vowels and both listening conditions. Thus, the performance of the LL group was not only lower overall, but these listeners also appeared to be more uncertain of which vowel they had heard in the silent-center conditions. The EL group also showed one more confusion vowel exceeding 5% than for the monolingual group for two target vowels (/i/ and /eɪ/) in both listening conditions.

Although the order of confusions changed relatively little from the DP to the DN listening condition within each listener group, some interesting patterns did emerge. For the MO listener group, the greatest decreases in performance from the DP to the DN condition were seen for the target vowels /ɪ/ and /ɛ/ (13% and 14%, respectively), but the order of confusions did not change in either case. In fact, the pat-

tern changed for the MO listeners from the DP to the DN for only one target vowel (/eɪ/); for this condition, the only confusion vowel equal to or greater than 5% was /ɛ/ in the DP condition and /ɪ/ in the DN condition, which is not surprising considering the substantial shortening of /eɪ/ that occurred for the creation of the DN condition. Thus, although the MO listener group showed the greatest (and only significant) decrease in performance from the DP to the DN condition, the pattern of confusions for these listeners changed very little across the two conditions.

For the EL listener group, the greatest decreases in performance from the DP to the DN condition were seen for the target vowels /i/ and /ɛ/ (5% and 9%, respectively). In the case of target /i/, the most frequent confusion vowel changed from /eɪ/ to /ɪ/ from the DP to the DN condition, as would be expected for the removal of a duration cue to these two neighboring vowels and partially supporting the hypothesis stated in the Introduction that different confusion patterns might be seen for this listener group when duration was not available as a cue. No other confusion vowels change order from the DP to the DN condition for this listener group, however.

For the LL listener group, the order of confusions varied between the DP and DN conditions for only one target vowel (/ɪ/), for which the most frequent confusion was /ɛ/ in the DP condition and /eɪ/ in the DN condition. These data do not support the hypothesis stated in the Introduction that different confusion patterns would be seen for this listener group when duration was not available as a cue. It is notable, however, that the most frequent *response* for target /i/ (not just the most frequent confusion) was /ɪ/ in both the DP and DN listening conditions. Figure 3(C) shows that percent-correct performance for target /i/ decreased by over 50% (from over 80% correct to less than 30% correct) from the whole-word to the DP 40 ms condition. These results suggest that the disruption caused by the silencing of the center made the target vowel /i/ unidentifiable for this listener group in either duration condition.

IV. DISCUSSION

The results of the present study generally support the research hypothesis that even relatively early Spanish L1 learners of English as a second language may have greater difficulty identifying L2 speech sounds based on partial acoustic information, compared to monolingual English speakers. Although the performance of the monolingual and early learner groups was nearly identical for the whole-word condition and did not differ significantly for the DP 40 ms condition, the EL listeners showed significantly lower performance overall on the silent-center conditions. In the *post hoc* analysis of the significant listener group by gate interaction, the performance of the MO listeners was significantly higher than that of the EL listeners for the 10 ms condition and approached significance for the 30 ms condition. Both the MO and the EL listeners performed with significantly higher accuracy than the LL listeners on most conditions, including the whole-word condition.

Furthermore, only the MO listener group appeared to be able to use the vowel duration information provided in the DP silent-center syllables effectively because this was the only group that showed significantly higher performance in the DP than in the DN listening condition. At first glance, this result would seem to differ from previous results that suggest that vowel duration information is often weighted as heavily or more heavily by speakers of English as a second language, even when vowel duration is not a contrastive cue in the L1 (Bohn, 1995). However, it is possible that the EL and the LL listeners *would* have made as much (or more) use of the duration information as the MO listeners in the intact syllable, but that they were unable to overcome the disruption of the syllable sufficiently to *use* the duration information effectively in the DP condition.

This interpretation would suggest that the phonemic representations of the EL listeners may not be substantively different from that of the MO listeners, but rather that their representations may be less *robust* than those of the MO listeners. In fact, the EL listeners did not perform significantly more poorly than the MO listeners in the DN condition; rather, their performance failed to *improve* significantly in the DP condition, although the difference did approach significance. The LL listeners, on the other hand, showed no evidence of improved identification rates from the DN to the DP condition and performed significantly more poorly than the other two groups in both the DP and DN conditions. Performance for the LL group also declined more dramatically than for the other two groups from the whole-word to the DP 40 ms condition; this result is partially accounted for by a much greater decline for this listener group for target vowel /i/ than for the other two groups. These data suggest that the LL listeners may have a reduced ability to recover from the loss of *any* target vowel information, compared to the other two groups. Once the initial disruption in the syllable occurred, however, the differences among the groups did not increase dramatically as the amount of information presented decreased (i.e., from the 40 to the 10 ms gate conditions).

A comparison of the number of confusions above 5% (shown in Table V) shows one to two more confusion vowels for the LL listeners than for the MO and EL listeners for all target vowels. Although the performance of the LL listeners on the whole-word condition is well above chance for each target vowel, the increased number of confusions in the silent-center conditions suggests that these listeners' category representations may be much more fragile than those of the other two groups, in that when the syllable was disrupted they appeared to be less certain of what vowel they heard than the MO and EL listeners.

The level of performance attained by the monolingual English-speaking listeners in the present study is similar to that of adult listeners in other studies of silent-center syllable perception, supporting the validity of the present data. The 30 ms DP and 30 ms DN conditions in the present study are most similar to two experimental conditions in Strange (1989) because (1) the target words in Strange (1989) were produced in a carrier phrase (although at a faster rate than the present study); (2) the initial and final portions in Strange

(1989) contained 26 and 34 ms of the CV and VC transitions, respectively; and (3) Strange (1989) also examined perception of syllables with DN silent centers. Strange's (1989) task did, however, employ more target vowels and a ten-alternative forced-choice task. Nevertheless, the performance by the present group of monolingual listeners and that of Strange's listeners is quite similar, with 96% correct performance on the whole-word data, 90% correct performance for the 30 ms DP stimuli, and a 4% reduction in performance on the 30 ms DN stimuli in the present study, compared to 98% correct whole-word performance, 84% silent-center performance, and a 5% reduction in performance on the DN stimuli in Strange's (1989) study. The present results also parallel those of Strange *et al.* (1983), in that the monolingual listeners in the present study showed greater decreases in identification performance for the short vowels (/ɪ/ and /ɛ/) and the midlength vowel (/i/) than for the longer vowels (/eɪ, æ, and ɑ/), despite the relatively long duration obtained for the "midlength" vowel /i/ found in the present study.

Parker and Diehl (1985) attempted to neutralize duration by the use of similar duration syllable triads as their alternatives and used four CV and VC transition durations (60%, 70%, 80%, and 90% vowel deletions, corresponding roughly to the 10, 20, 30, and 40 ms duration CV and VC conditions used in the present study). Despite the difference in number of alternatives in the forced-choice task [six in the present study and three in Parker and Diehl (1985)], the performance of the monolingual listeners in the present study was consistently within about 4%–8% of that of listeners in Parker and Diehl's (1985) study.

Fox *et al.* (1992) used a number of different syllable types and an open-set identification task to compare the performance of older and younger monolingual listeners on whole-word and silent-center stimuli. As in the present study, they found no significant difference between the listener groups in the performance on the whole-word condition but significantly lower performance (by about 7%) on the identification for older adults than for younger adults in the silent-center condition. In the present study, no significant difference between the performance of the MO and EL listeners was found in the whole-word condition, but the EL listeners' performance was consistently and significantly less accurate (by about 6%–10%) than that of the MO listeners across most of the silent center conditions. Thus, the effects of early learning of a second language and the effects of aging on the listeners' ability to identify syllables based on partial vowel information would appear to be similar in magnitude, although they may have quite different origins. This similarity is intriguing and direct comparison of silent-center syllable perception by non-native and elderly listeners may yield interesting results.

V. CONCLUSION

The results of the present study indicate that even relatively early learners of English as a second language may have more difficulty in identifying speech sounds based on partial acoustic information. This difference may therefore account for a portion of the increased difficulty that even early learners of English appear to have in processing speech in noise and/or reverberation. The source of this difference at the phonetic level may lie in a reduced robustness in recovering from syllable disruption, less flexibility in switching perceptually to focus on the available speech cues when some cues are obscured, differences in phoneme boundaries or cue weighting, or other potential explanations. Whatever the reason for the differences observed in the present study between the native and non-native listeners' ability to identify the target vowels based on CV and VC formant transitions alone, it is likely that they only account for a portion of the difficulty that non-natives experience processing speech in difficult environments in the real world. As suggested by the results of both Bradlow and Alexander (2007) and Cutler *et al.* (2004), differences may well be observed between native and non-native listeners at each level of linguistic processing and from both top-down and bottom-up sources. Thus, further systematic comparisons of speech processing using different tasks and investigating different levels of processing are necessary to understand this problem. Furthermore, in light of the small but significant differences found in the present study for even early learners with little or no observable foreign accent, future research should detail language background variables carefully.

Despite the significant results, the small number of participants in each listener group is a limitation of the present study. Thus, replication and further analysis of the use of partial information for both vowels and consonants with larger numbers of listeners from varying L1s and L2s is necessary to generalize from the present data. Replication of the present study using consonants with lingual articulations (e.g., /d/ or /g/) would be particularly interesting because the effects of CV and VC transitions on syllable identification might be expected to be stronger for these consonants, due to the greater degree of coarticulation between consonant and vowel needed for consonants with lingual articulations.

Despite its limitations, the results of the present study suggest that later learners of a second language and perhaps even some early learners may benefit from perceptual training that may help listeners to develop greater flexibility in using alternate sources of phonetic information when some information is unavailable. Such training methods may be helpful in enabling non-native speakers to perform more like monolingual listeners in challenging listening environments such as noise or reverberation.

ACKNOWLEDGMENTS

This research was supported by NIH-NIDCD Grant No. 5R03 DC005561 to Catherine L. Rogers and by a University of South Florida Graduate Student Travel Award to Alexandra S. Lopez. We thank Stefan A. Frisch, Joseph Constantine, Gail S. Donaldson, and Theresa H. Chisolm for their helpful suggestions.

[1]The RAU-transformed scale extends from −20 to 120, rather than 0% to 100%, but the scale is designed so that scores in the middle of the percent-correct range change very little when transformed to RAUs. The maximum possible performance (six out of six correct) translated to approxi-

mately 105 RAU in the present case, rather than the theoretical maximum RAU of 120.

[2]Data for two LL listeners were dropped in this analysis because raw data files were lost due to computer failure, but data for all eight LL participants were available for the analysis of gated conditions and overall data for the whole-word condition indicate similar performance on that condition for these two participants.

Andruski, J. E., and Nearey, T. M. (**1992**). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables," J. Acoust. Soc. Am. **91**, 390–410.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods Instrum. Comput. **37**, 379–384.

Boersma, P., and Weenik, P. (**2006**). *PRAAT: Doing phonetics by computer* (version 4.4.30), retrieved September 7, **2006** from http://www.praat.org.

Bohn, O.-S. (**1995**). "Cross-language speech perception in adults: First language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore, MD), pp. 279–304.

Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Clopper, C. G., and Pisoni, D. B. (**2004**). "Some acoustic cues for the perceptual categorization of American English regional dialects," J. Phonetics **32**, 111–140.

COOLEDIT 2000 (version 1.1) (**2000**). Syntrillium, Inc., Phoenix, AZ.

Crawford, K. E. (**2006**). "The relationship between degree of foreign accentedness and vowel perception of Spanish-English bilinguals," thesis, University of South Florida, Tampa, FL.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Dalbor, J. B. (**1969**). *Spanish Pronunciation: Theory and Practice* (Holt, Reinhart and Winston, New York, NY).

ECOS/WIN (version 1.3) (**1999**). AVAAZ Innovations, Inc., London, Ontario.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings and problems." in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore, MD), pp. 233–277.

Flege, J. E., and MacKay, I. R. A. (**2004**). "Perceiving vowels in a second language," Stud. Second Lang. Acquis. **26**, 1–34.

Fox, R. A., Wall, L. G., and Gokgen, J. (**1992**). "Age-related differences in processing dynamic information to identify vowel quality," J. Speech Hear. Res. **35**, 892–902.

Hillenbrand, J., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

Hudgins, C. V., and Numbers, F. C. (**1942**). "An investigation of the intelligibility of the speech of the deaf," Genet. Psychol. Monogr. **25**, 289–392.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements." IEEE Trans. Audio Electroacoust., **AU-17**, 225–246.

Egan, J. P. (**1948**). "Articulation testing methods," Laryngoscope **58**, 955–991.

Imai, S., Walley, A. S., and Flege, J. E. (**2005**). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," J. Acoust. Soc. Am. **117**, 896–907.

Jenkins, J. J., and Strange, W. (**1999**). "Perception of dynamic information for vowels in syllable onsets and offsets," Percept. Psychophys. **61**, 1200–1210.

Jenkins, J. J., Strange, W., and Miranda, S. (**1994**). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**, 1030–1043.

Jenkins, J. J., Strange, W., and Trent, S. A. (**1999**). "Context-independent dynamic information for the perception of coarticulated vowels," J. Acoust. Soc. Am. **106**, 438–448.

Kent, R. D., Weismer, G., Sufit, G., Rosenbek, J. C., Martin, R. E., and Brooks, B. R. (**1990**). "Impairment in speech intelligibility in men with amyotrophic lateral sclerosis," J. Speech Hear Disord. **55**, 721–728.

Kewley-Port, D., and Goodman, S. S. (**2005**). "Thresholds for second-formant transitions in front vowels," J. Acoust. Soc. Am. **118**, 3252–3260.

Kirk, K. I., Tye-Murray, N., and Hurtig, R. R. (**1992**). "The use of static and dynamic vowel cues by multichannel cochlear implant users," J. Acoust. Soc. Am. **91**, 3487–3498.

Ladefoged, P. (**1982**). *A Course in Phonetics*, 2nd ed. (Harcourt, Brace, Jovanovich, New York).

Lattner, S., Meyer, M. E., and Friederici, A. D. (**2005**). "Voice perception: Sex, pitch and the right hemisphere," Hum. Brain Mapp **24**, 11–20.

Lindblom, B. (**1996**). "Role of articulation in speech perception: Clues from production," J. Acoust. Soc. Am. **99**, 1683–1692.

Logan, J. S., Greene, B. G., and Pisoni, D. B. (**1989**). "Segmental intelligibility of synthetic speech produced by rule," J. Acoust. Soc. Am. **86**, 566–581.

Mayo, L., 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. A. (**2000**). "Factors affecting the recognition of words in a second language," Bilingualism: Lang. Cognit. **3**, 55–67.

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.

Moncur, J., and Dirks, D. (**1967**). "Binaural and monaural speech intelligibility in reverberation," J. Acoust. Soc. Am. **10**, 186–195.

Monsen, R. B. (**1983**). "The oral speech intelligibility of hearing-impaired talkers," J. Speech Hear Disord. **48**, 286–296.

Mullenix, J., Johnson, K., Topcu-Durgun, M., and Farnsworth, L. (**1995**). "The perceptual representation of voice gender," J. Acoust. Soc. Am. **98**, 3080–3095.

Murphy, W. D., Shea, S. L., and Aslin, R. N. (**1989**). "Identification of vowels in 'vowelless' syllables by 3-year-olds," Percept. Psychophys. **46**, 375–383.

Parker, E. M., and Diehl, R. L. (**1985**). "Identifying vowels in CVC syllables: Effects of inserting silence and noise," Percept. Psychophys. **36**, 369–380.

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.

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.

SPSS (version 15.0) (**2006**). SPSS, Inc., Chicago, IL.

Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.

Studebaker, G. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 494–509.

Sussman, J. E. (**2001**). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?," J. Acoust. Soc. Am. **109**, 1173–1180.

TDT SYSTEM III (**2001**). Tucker-Davis Technologies, Inc., Gainesville, FL.

United States Census Bureau (**2000**). 2000 U.S. Census. Retrieved May 8, **2007**, from http://www.factfinder.census.gov.

The effect of age on auditory spatial attention in conditions of real and simulated spatial separation

Gurjit Singh and M. Kathleen Pichora-Fuller[a)]
Department of Psychology, University of Toronto, 3359 Mississauga Road North, Mississauga, Ontario L5L 1C6, Canada and Toronto Rehabilitation Institute, University of Toronto, 3359 Mississauga Road North, Mississauga, Ontario L5L 1C6, Canada

Bruce A. Schneider
Department of Psychology, University of Toronto, 3359 Mississauga Road North, Mississauga, Ontario L5L 1C6, Canada

(Received 29 May 2007; revised 28 May 2008; accepted 29 May 2008)

The contributions of auditory and cognitive factors to age-dependent differences in auditory spatial attention were investigated. In conditions of real spatial separation, the target sentence was presented from a central location and competing sentences were presented from left and right locations. In conditions of simulated spatial separation, different apparent spatial locations of the target and competitors were induced using the precedence effect. The identity of the target was cued by a callsign presented either prior to or following each target sentence, and the probability that the target would be presented at the three locations was specified at the beginning of each block. Younger and older adults with normal hearing sensitivity below 4 kHz completed all 16 conditions (2-spatial separation method × 2-callsign conditions × 4-probability conditions). Overall, younger adults performed better than older adults. For both age groups, performance improved with target location certainty, with *a priori* target cueing, and when location differences were real rather than simulated. For both age groups, the contributions of natural spatial cues were most pronounced when the target occurred at "unlikely" spatial listening locations. This suggests that both age groups benefit similarly from richer acoustical cues and *a priori* information in difficult listening environments. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2949399]

PACS number(s): 43.71.Lz, 43.66.Pn, 43.71.Es, 43.66.Qp [RLF] Pages: 1294–1305

I. INTRODUCTION

Older adults, even those with clinically normal audiograms in the speech range, often report difficulty understanding speech in noisy, reverberant, and multitalker environments (e.g., CHABA, 1988; Dubno *et al.*, 1984; Duquesnoy, 1983; Frisina and Frisina, 1997; Gordon-Salant and Fitzgibbons, 1995; Helfer, 1992; Pichora-Fuller *et al.*, 1995; Tun and Wingfield, 1999). Their problems in complex listening situations seem to arise from a combination of age-dependent differences in auditory and cognitive processing (for reviews, see Pichora-Fuller, 1997, 2003, 2007; Schneider *et al.*, 2002). In a typical everyday multitalker environment, listeners rely on auditory processing to localize sources originating from different locations and on cognitive processing to attend to the target(s) and ignore distracting sounds from other sources. Age-dependent differences in binaural processing have been found in older listeners with normal hearing sensitivity in the speech range (for reviews, see Grose, 1996; Koehnke and Besing, 2001), and these differences may contribute to the disproportionate problems experienced by older listeners when they must understand speech in spatially distributed multitalker situations. In addition, age-dependent differences in attention could also play a role. Some of the task-relevant attentional factors important for spatial listening include selective attention (i.e., attending to a single talker), inhibition (i.e., ignoring irrelevant talkers), divided attention (i.e., attending to multiple streams simultaneously), and spatial attention switching (i.e., shifting the allocation of attentional resources over time to different points in space). Previous research has found that, relative to younger adults, older adults do not demonstrate deficits on tasks specific to selective attention (Verhaeghen and Cerella, 2002), but that age-dependent deficits are observed on tasks related to inhibition (Hasher and Zacks, 1988) and divided attention (for a review, see Pashler, 1993). Mixed results have been observed for studies examining spatial attention switching for visual tasks (Folk and Hoyer, 1992; Greenwood and Parasuraman, 1994; Madden *et al.*, 1994; for a review, see McDowd and Shaw, 2000). Currently, we know of no research that has examined the auditory spatial attention abilities of older adults in multitalker environments.

A. Auditory and cognitive factors in spatial listening

The interplay of auditory and cognitive factors in multitalker listening situations was described by Cherry (1953) as the ability of a listener to selectively attend to and identify the speech from a single talker among a mixture of background conversations and noises in a "cocktail party" situation (for reviews, see Bregman, 1990; Yost, 1997; Bronkhorst, 2000; Ebata, 2003; Haykin and Chen, 2005).

[a)]Author to whom correspondence should be addressed. Tel.: 905-828-3865. FAX: 905-569-4326. Electronic mail: k.pichora.fuller@utoronto.ca

 0001-4966/2008/124(2)/1294/12/$23.00

Auditory scene analysis (Bregman, 1990) provides a useful framework within which to consider the cocktail party problem and the specific processes by which the human auditory system enables the decomposition of incoming complex signals into separate perceptual representations. When listening in noisy backgrounds, auditory input is partitioned based on the acoustic properties of the stimulus (i.e., bottom-up or data-driven processing) and/or categorized by making use of stored auditory object representations that are formed by prior knowledge and experience (i.e., top-down or prediction-driven processing). The assumption is that in order to distinguish a target from a masker, listeners will use a combination of auditory and cognitive processes.

Cherry (1953) suggested that spatially separating a target from a masker improves target recognition. This benefits listeners because there is then less neural competition between a target and masker in the auditory periphery (i.e., release from energetic masking) and because there is reduced competition at higher levels of processing (i.e., release from perceptual or informational masking) (e.g., Watson, 1987; Durlach *et al.*, 2003; Hornsby *et al.*, 2006; Li *et al.*, 2004; Wu *et al.*, 2005). Several mechanisms account for spatial unmasking in complex listening environments, and it may be useful to consider them as described in an information processing framework that incorporates both signal-driven bottom-up and cognitively meditated top-down factors.

Spatial separation of target and masker provides a number of monaural and binaural auditory cues that could facilitate speech identification in noise. Foremost among the monaural cues is that a separation of target and masker will change the signal-to-noise ratio (SNR) at an ear. Compare a situation in which both target and masker are located to the listener's left to a situation in which the target is on the right and the masker is on the left. Moving the target from left to right dramatically improves the SNR at the right ear due to the sound shadow cast by the listener's head. In addition, a shift in target position will change the spectral profile of the target at one ear due to diffractive and reflective properties of the pinna, head, and torso, but the role of monaural spectral profile cues in promoting intelligibility appears to be comparatively minor relative to monaural changes in SNR (Wightman and Kistler, 1997). Hence, spatially separating the target from the masker produces cues, which when processed monaurally could aid speech identification in noise.

Speech identification could also be enhanced by binaural processing. Spatial separation of a target from a masker leads to an interaural level difference (ILD) that is different between target and masker because of the head shadow, as well as an interaural time difference (ITD) that is different between target and masker because of the different distances a sound must travel to reach each ear (for a review, see Blauert, 1997; Bronkhorst, 2000). Binaural processing of these interaural differences enables higher perceptual systems to take advantage of the subtle spectro-temporal differences between the target and masker signals arriving at each ear (e.g., Colburn *et al.*, 2006; Culling *et al.*, 2004; Duquesnoy, 1983; Plomp, 1976). Hence, it is possible that age-dependent changes in either monaural or binaural processing could reduce the effectiveness of spatial separation in releasing a target from masking.

In addition to bottom-up, signal-driven processes, cognitive, or prediction-driven, mechanisms are also likely to be engaged, or become more effective, when the target and masker are physically separated. One such top-down process may involve the listener's ability to focus attention along a spatial vector to a target (Spence and Driver, 1994).[1] Mondor and Zatorre (1995) found evidence that the auditory system is able to exploit acoustic spatial expectations, with faster response times being observed on trials where auditory targets were preceded by cues providing accurate information about the spatial location of a target compared to response times on trials preceded by cues providing inaccurate target location information. They proposed that a gradient model best describes auditory spatial attention where performance declines with increasing distance from the spatial center of attentional focus. There is also evidence that auditory spatial attention operates in a manner consistent with Broadbent's (1958) spotlight model of attention (Best *et al.*, 2006). Accordingly, physical separation of the target and masker should facilitate attentional processes by distancing the masker from the center of the spotlight.

Investigations of auditory spatial attention have found that benefits in performance are not limited to faster response times for cued locations, but that they also include improved discrimination and recognition abilities (Sach *et al.*, 2000; Sach and Bailey, 2004). In one experiment, participants who were instructed to focus attention on one of seven loudspeakers using the probe-signal method were better able to identify signals for cued relative to uncued spatial locations (Arbogast and Kidd, 2000). Larger effect sizes were found in a subsequent investigation completed in a complex multitalker listening environment including conditions with a high degree of uncertainty about the target location (Kidd *et al.*, 2005). Specifically, in a three-talker listening environment, speech intelligibility on the coordinate response measure (Bolia *et al.*, 2000) ranged from 31% to an impressive 92% correct when listeners were provided with prior information about the identity and location of the target.

B. Real and simulated spatial separation

A listener's ability to benefit from spatial separation when listening to multiple sound sources could depend to varying degrees on different acoustical cues. The roles of ILD and ITD cues for the direct wave front were described above. However, in realistic listening situations there are usually reflected wave fronts as well. Over short distances and brief time intervals, the precedence effect occurs, whereby the human auditory system fuses direct sound waves and early reflections into a single auditory event, rather than a percept followed by an echo (Wallach *et al.*, 1949; for reviews, see Zurek, 1980; Blauert, 1997; Litovsky *et al.*, 1999; Li and Yue, 2002). By manipulating the time delay between the presentation of a signal from two equidistant loudspeakers located in the right and left hemifields of a listener, it is possible to simulate different spatial locations. By presenting a sentence simultaneously from each loud-

speaker, a listener will perceive the sentence as originating from a location between the two loudspeakers. If a sound from one of the loudspeakers lags the sound from the other loudspeaker by a few milliseconds, the resulting percept is localized near the location of the leading loudspeaker. In this way, one can simulate spatial separation between a target and masker when both sentences are presented from both loudspeakers. This methodology has been previously used to simulate spatial separation and is particularly well suited to the study of auditory spatial attention insofar as the perception of spatial separation is achieved while minimizing the contribution of some of the acoustical differences (changes in monaural SNRs, interaural correlation, etc.) that are present when there is real spatial separation (e.g., Freyman *et al.*, 1999; Rakerd *et al.*, 2006; and the Appendix in Li *et al.*, 2004).

In the real spatial separation case, the target is delivered only from a loudspeaker at one location, and the competitor is delivered only from a loudspeaker at a different location. Consequently, multiple interaural cues are available, and the average target-to-competitor ratio at one ear is not equal to that at the other ear because of head shadow effects. In contrast, when the precedence effect is used to simulate spatial separation, the target and competitor are delivered at equal presentation levels from each of two loudspeakers with time delays introduced to induce the perception of spatial separation. Consequently, interaural level difference cues due to head shadow are minimal, and the average target-to-competitor ratio is the same at each ear. There are, of course, other differences between the spectra of signals presented in simulated spatial separation conditions compared to those presented in real spatial separation conditions, the most prominent of which are comb-filtering effects. However, the effects of comb-filtering are unlikely to affect performance when the onset delay between the left and right loudspeakers is long (>2 ms) and tests are conducted in a nonanechoic sound-attenuating chamber (Li *et al.*, 2004; see also Brungart *et al.*, 2005). At the perceptual level, sounds whose perceived locations are induced using the precedence effect (playing the sounds over two loudspeakers with one sound leading the other) are perceived to be more diffuse and are less precisely localized than the same sound played over a single loudspeaker (Blauert, 1997). The degree to which these perceptual differences between a precedence-induced spatially located sound and a sound presented from single spatial location will affect performance will likely depend on the precision of localization demanded by a listening task.

C. Aging

Older adults typically have more difficulty than younger adults when speech identification is tested using monaurally presented speech-in-noise tests, and their difficulties may be explained, at least partially, by age-dependent declines in temporal processing (for reviews see CHABA, 1988; Divenyi and Simon, 1999; Pichora-Fuller and Souza, 2003). For older listeners with near normal pure-tone thresholds, there are discrepant findings concerning their ability to use ILD and ITD cues. For example, Herman *et al.* (1977) found ILD limens to be the same for younger and older listeners; however, Pichora-Fuller and Schneider (1992, 1998) found age-dependent binaural masking-level differences that were explained in terms of an age-dependent increase in the amount of temporal jitter in the binaural system (see also Dubno *et al.*, 2002; for a review, see Grose, 1996). Thus, age-dependent differences in monaural and/or binaural auditory processing could compromise the ability of older listeners to follow speech in complex multitalker situations.

The current investigation is designed to extend the findings from Kidd *et al.* (2005) and Li *et al.* (2004) to explore possible age-dependent differences in auditory spatial attention in a multitalker situation where the acoustic cues are either fully available or reduced using the precedence effect. Using the precedence effect, Li *et al.* (2004) presented semantically meaningless target sentences at a location to the right of younger and older adult listeners, while manipulating the simulated location of a single masker (right, left, or in front). In their experiment, the location of target and masker sentences was 100% certain. Although older adults needed a better SNR to perform equivalently to younger adults, there were no age-dependent differences in masking release. Using real spatial separation of target and maskers, the study of Kidd *et al.* (2005) demonstrated that for younger listeners, certainty about the location of the target and prior information about the callsign cue to the target utterance are important factors contributing to the allocation of attention and improvement in word identification performance. Age-dependent differences in unmasking may not have been found in the study of Li *et al.* (2004) because the attentional demands placed on listeners were limited insofar as the location of the target was always known. In vision, previous research suggests that there is an effect of age when attentional demands are high (for a review, see McDowd and Shaw, 2000). Therefore, by manipulating the attentional demands of the listening task by varying location certainty, age-dependent differences in performance might be revealed.

By comparing the word identification accuracy of younger and older age listeners in conditions of real spatial separation (following Kidd *et al.*, 2005), we set out to determine if age-dependent differences would be observed when the attentional demands of the task were varied. Second, we used the precedence effect (following Freyman *et al.*, 1999) to simulate spatial separation to determine if age-dependent differences would be exacerbated when the monaural and binaural cues to spatial location were reduced. By comparing the results between the two presentation methods, we hoped to gain a better understanding of how bottom-up processes interact with top-down processes to enhance word identification when a target is physically separated from a masker and to determine the extent to which this interaction is age dependent.

II. METHODS

A. Participants

Eight younger adults (aged 21–30 years; mean=24.38; SD=3.02) and eight older adults (aged 66–78 years; mean =70.38; SD=3.89) participated in the study. Participants were recruited from the local university and community. All participants were native English speakers and in good overall

TABLE I. Left (L) and right (R) ear hearing thresholds (dB HL) for the eight younger adults.

	Frequency (kHz)							
SS	0.250	0.500	1	1.5	2	3	4	8
L1	15	5	0	5	5	15	20	15
L2	15	10	5	0	5	5	5	15
L3	10	5	0	0	0	10	25	10
L4	−10	−10	−10	−5	−10	−10	−5	−10
L5	0	−5	0	−5	−5	−5	−5	0
L6	−5	5	−10	−5	5	0	15	−5
L7	0	5	5	10	5	15	10	0
L8	0	0	−5	−5	−5	0	5	0
$\bar{x}$	3.13	1.88	−1.88	−0.63	0.00	3.75	8.75	3.13
SD	9.23	6.51	5.94	5.63	5.98	9.16	10.94	9.23
R1	15	10	10	5	5	0	0	5
R2	5	10	5	0	0	−5	−5	10
R3	5	0	0	0	5	5	10	0
R4	−5	−10	−5	−5	−5	−5	−5	0
R5	5	5	−5	−5	−5	−5	5	−5
R6	0	0	5	0	0	−5	−5	0
R7	5	5	0	0	0	5	10	5
R8	0	0	0	0	0	0	−5	5
$\bar{x}$	3.75	2.50	1.25	−0.63	0.00	−1.25	0.63	2.50
SD	5.82	6.55	5.18	3.20	3.78	4.43	6.78	4.63

health. All listeners had clinically normal pure-tone air-conduction thresholds (less than or equal to 25 dB HL) from 0.25 to 3 kHz in both ears (see Tables I and II). All participants provided informed consent and were paid $10 per hour of testing.

B. Stimuli

The stimuli consisted of sentences from the coordinate response measure (CRM) corpus spoken by the four male talkers (Bolia *et al.*, 2000). The sentences have the format: "Ready [callsign] go to [color] [number] now." The CRM corpus contains sentences with all possible combinations of eight callsigns (arrow, baron, charlie, eagle, hopper, laker, ringo, and tiger), four colors (red, white, blue, and green), and eight numbers (one, two, three, four, five, six, seven, and eight).

TABLE II. Left (L) and right (R) ear hearing thresholds (dB HL) for the eight older adults.

	Frequency (kHz)							
SS	0.250	0.500	1	1.5	2	3	4	8
L1	0	0	5	5	15	20	40	10
L2	10	−5	5	5	10	10	30	20
L3	15	10	15	10	15	25	40	25
L4	5	5	15	10	15	0	25	25
L5	5	5	15	15	10	15	5	5
L6	5	15	20	15	20	20	25	25
L7	10	0	10	5	15	10	25	20
L8	10	10	0	0	5	0	40	25
$\bar{x}$	7.50	5.00	10.63	8.13	13.13	12.50	28.75	19.38
SD	4.63	6.55	6.78	5.30	4.58	9.26	11.88	7.76
R1	5	−5	5	5	10	25	20	30
R2	5	0	5	5	20	5	10	40
R3	20	20	5	15	10	25	20	25
R4	0	10	10	0	5	0	15	30
R5	5	10	10	15	15	15	15	10
R6	5	15	20	15	15	15	20	30
R7	5	0	5	10	15	15	30	15
R8	10	10	5	0	5	10	15	25
$\bar{x}$	6.88	7.50	8.13	8.13	11.88	13.75	18.13	25.63
SD	5.94	8.45	5.30	6.51	5.30	8.76	5.94	9.43

C. Equipment

All testing was performed in a 3.3 m^2 (approximately 10.8 ft^2) single-walled sound-attenuating Industrial Acoustics Company (IAC) sound booth. Presentation of the stimuli was controlled via custom software developed on a VISUAL BASIC platform. All stimuli were routed from a computer to a Tucker-Davis Technologies (TDT) System III to a Harmon/Kardon amplifier (model HK3380). Sentences were converted to analog form using two TDT System III RP2.1s at a 24.414 kHz sampling rate by a 24 bit Sigma-Delta digital-to-analog converter. The analog outputs were attenuated using two programmable attenuators (TDT System III PA5) for the simulated spatial separation condition, and three attenuators for the real spatial separation condition. Signals were then conditioned by stereopower amplifiers (TDT System III SA1) and presented over Grason-Stadler Inc. (Catalog No. 1761-9630) loudspeakers. Visual cueing and feedback was displayed on a Compar (model 1760nx) 17 in. NEC touch screen monitor on a 0.46 m high table located in front, but below the shoulder level of the listener. Calibration was performed using a Brüel & Kjær (B&K) modular precision sound level meter (type 2260) with a 0.5 in. B&K condenser microphone (type 4189). The microphone was placed at the location that the listener's head would occupy, and measurements were collected using the *A*-weighting equivalent. Measurements were taken separately for each sentence presented alone from each loudspeaker. During calibration, concatenated sentences from the CRM corpus were presented at a level such that each loudspeaker, playing alone, would produce an average sound pressure of 60 dBA at the location corresponding to the center of the listener's head.

D. Design

The main dependent variable was accuracy of word identification. Three independent variables were systematically manipulated in this experiment. Two of the variables, callsign cue and location certainty, were identical to those studied in the experiment of Kidd *et al.* (2005). A new variable, presentation method, was also manipulated. Word identification was measured for all participants in each age group for 16 conditions (2 callsign cue conditions $\times$ 4 location probability specifications $\times$ 2 presentation methods). All participants completed every condition. The order of testing using the two presentation methods was counterbalanced, with half of the younger and older participants starting with one presentation method and the remainder starting with the other. For each presentation method there were 8–16 testing sessions, usually with two sessions completed in a 1–2 h visit. The order of testing the two callsign cue conditions was counterbalanced for each participant; the data in each session were collected in eight blocks with a sequence of cue conditions alternating every two blocks (i.e., two blocks of one cue condition followed by two blocks of the other cue condition). Each block consisted of 30 trials. The starting callsign cue condition was randomly determined for each listener. Within each session, for each callsign condition, the four probability specifications were assigned randomly without replacement so that a different probability certainty condition was assigned for each block. Each participant minimally completed 3840 trials, with a minimum of 240 completed trials for each of the 16 conditions.

E. Procedures

The listener's task was to identify the color and number in the target sentence that was presented simultaneously with two masker sentences. All sentences were presented at 60 dBA and were selected randomly from the CRM corpus. On a given trial, the target and masker sentences differed with respect to the color, number, callsign, and talker of the sentence. Listeners pressed on a touch screen to indicate which one of the four possible colors and which one of the eight possible numbers they heard in the target sentence. Both the correct color and number were required for a correct response. Feedback (*correct* or *incorrect*) was provided after every trial and the percentage of correct trials was provided at the end of each block.

At the start of each visit, participants completed either one or two practice blocks using the presentation method to be tested. Within any given practice block, a randomly chosen callsign cue condition combined with a randomly chosen location certainty condition was completed.

In the test phase, within any given block, participants were informed in advance of the probability specification for the block and they were cued to the identity of the callsign that began the target utterance either 1 s before (callsign cue before condition) or immediately following (callsign cue after condition) each trial in the block. The probability specification appeared on the monitor 1 s before and throughout the block. There were four probability specifications indicating the proportion of trials that the target would be presented from the left, center, and right spatial locations (0-100-0, 10-80-10, 20-60-20, and 33-33-33). For example, when participants were provided with a 10-80-10 cue, they were instructed that there was a 10% chance that the target would be presented from the left location, an 80% chance that the target sentence would be presented from a center location in front of them, and a 10% chance that the target would be presented from the right location. Thus, a cue of 0-100-0 indicated certainty that the target would be presented from a location in front and the 33-33-33 cue indicated that the location of the target sentence would be determined randomly. On any given trial, the location of the target sentence was randomly selected from the left, center, and right locations, with the limitation being that the probability cue was accurate across a block of 30 trials. Participants were instructed to face directly ahead for the duration of the stimulus presentation.

Two different presentation conditions were used, a real spatial separation condition and a simulated spatial separation condition. Following Kidd *et al.* (2005), in the real spatial separation condition, three loudspeakers were used to present stimuli. The target and two masker sentences were presented simultaneously, but only one sentence was played from each loudspeaker. Each loudspeaker was located 1.83 m (approximately 6 ft) from the participant's head. As shown in Fig. 1, they were positioned at 0° and $\pm$54° azimuth in the horizontal plane.

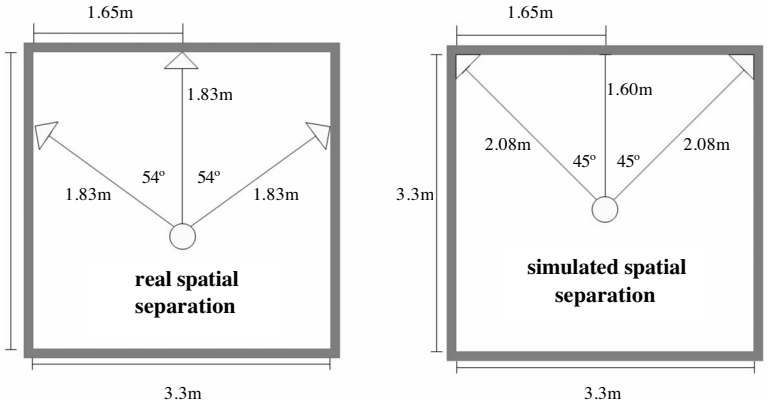


FIG. 1. Schematic of loudspeaker configuration for real (left) and simulated spatial separation (right). The circle indicates the position of the participant and the triangles indicate the position of the loudspeakers.

In the simulated spatial separation condition, we took advantage of the precedence effect (e.g., Freyman *et al.*, 1999) and achieved simulated spatial separation of the target and competitors by manipulating time delays between the presentations from only two loudspeakers. Each loudspeaker was located 2.08 m (approximately 6.8 ft) from the listener's head and was positioned at ±45° azimuth. The height of all the loudspeakers was approximately the same height as the listener's head when seated. Although all three sentences were presented from both loudspeakers, an utterance appeared to come from the left spatial location when the signal from the left loudspeaker led the signal from the right loudspeaker by 3 ms. Similarly, an utterance appeared to come from the right spatial location when the left loudspeaker lagged the right by 3 ms. Finally, an utterance appeared to come from in front of the participant when a sentence was presented from both loudspeakers at the same time. Thus, we achieved simulated spatial separation of the target and competing sentences using the precedence effect, but all three sentences were presented from both the left and right loudspeakers (see Fig. 1 for a schematic of the soundbooth configurations).

III. RESULTS

The overall results from the experiment are shown in Fig. 2 for both age groups. In general, we observed four main findings of interest. First, although older adults performed worse than younger adults, there was a similar pattern of results across all conditions for both age groups. Second, performance improved when listeners were more certain about the location of the target. Third, performance was better when the callsign cue was known in advance. Fourth, the effect of location certainty and callsign cue were more pronounced in the real compared to the simulated spatial separation conditions. These effects were tested with a $2\times2\times2\times4$ repeated-measures analysis of variance (ANOVA) where age (younger versus older) was a between-subjects variable and presentation method (real versus simulated spatial separation), callsign cue (before versus after), and location certainty (1.0, 0.8, 0.6, and 0.33) were within-subjects variables.

A. Age

As shown in Fig. 2, younger adults outperformed older adults in almost every condition of the study. This was confirmed by a significant main effect of age [$F(1,14)=7.60, p<0.05$]. Collapsing across all conditions, the mean word identification accuracy of the younger adult group was approximately 8.2 percentage points better than that observed in the older adult group. Most importantly, we did not

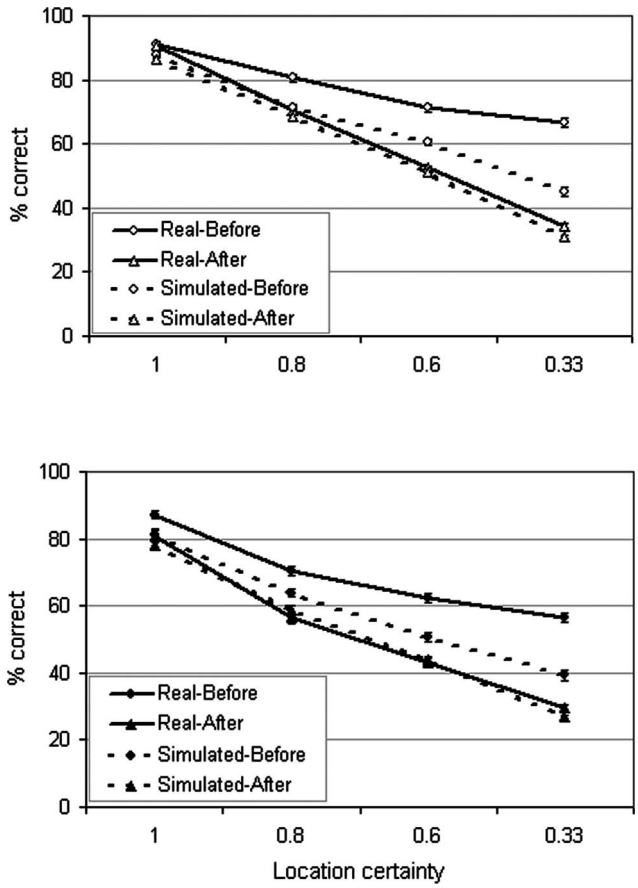


FIG. 2. Mean percent correct identification scores and standard errors of the mean for younger (top, unfilled symbols) and older (bottom, filled symbols) adults for the four location certainties. Scores on the *y*-axis were calculated as the percentage of trials where participants correctly identified both the color and number associated with the target callsign. Solid lines indicate real spatial separation, dashed lines indicate simulated spatial separation, circles indicate callsign cue before conditions, and triangles indicate callsign cue after conditions.

observe any significant interactions of age with presentation method, callsign cue, or location certainty (all $p > 0.05$), suggesting that younger and older adults did not differentially make use of information about target identity and location probability under either real or simulated spatial separation presentation methods.

B. Location certainty

The single largest factor which improved performance was *a priori* target location information, as confirmed by the significant main effect of location certainty [$F(3,42) = 526.37$, $p < 0.001$]. On average, when listeners were most uncertain about the location of the target, correct identification was 41.6% whereas it was 85.4% when the target was presented at a fixed location, yielding an improvement of 44.2 percentage points. However, as depicted in Fig. 2, the deleterious effect of location uncertainty was offset when participants had prior knowledge of the identity of the target callsign, particularly for presentations using real spatial separation (i.e., in Fig. 2, the slopes of the functions are shallower for the callsign-before compared to the callsign-after conditions). Nevertheless, even with this offset, performance improved by 27.5 percentage points when location certainty was increased from 0.33 (where performance was 61.7%) to 1.00 (where performance was 89.2%). This pattern of interactions was confirmed with a significant three-way interaction of presentation method, callsign cue, and location certainty [$F(3,42) = 16.68$, $p < 0.001$]. We also observed significant two-way interactions of location certainty and callsign cue [$F(3,42) = 47.80$, $p < 0.001$], location certainty and presentation method [$F(3,42) = 17.02$, $p < 0.001$], and callsign cue and presentation method [$F(1,14) = 100.79$, $p < 0.001$]; however, we will not further discuss these particular analyses as the relationships between location certainty, callsign cue, and presentation method can more precisely be characterized through consideration of the three-way interaction.

C. Callsign cue

Word identification improved with prior knowledge of the target callsign. Specifically, performance improved by 11.5 percentage points when the callsign cue was known prior to the start of a trial (67.8%) compared to when it was cued after stimulus presentation (56.3%). This was confirmed by a significant main effect of callsign cue [$F(1,14) = 58.10$, $p < 0.001$]. As illustrated in Fig. 2 and confirmed by the significant three-way interaction discussed above, the effect of knowing the callsign cue prior to the start of the signal presentation was most pronounced when location certainty was reduced and when there was real spatial separation. Whereas there was a 7 percentage point benefit associated with advance knowledge of the callsign for simulated spatial separation, there was a 15 percentage point improvement with real spatial separation, suggesting that the removal of some of the auditory cues associated with real spatial separation lessens, but does not completely eliminate the benefit of knowing the identity of the target callsign in advance of hearing the sentence.

D. Real versus simulated spatial separation

Relative to the listening situation where spatial location was simulated via the precedence effect, overall performance in the real spatial separation condition was better by an average of 6.3 percentage points. This was confirmed by a significant main effect of presentation method [$F(1,14) = 36.04$, $p < 0.001$]. However, again evidenced by the significant three-way interaction, the benefit realized when sentences were generated from three independent spatial locations was most noticeable when the callsign cue preceded stimulus presentation and the target location was less certain. Notably, no benefit from real versus simulated spatial separation was observed when the callsign cue was presented after stimulus presentation.

E. Spatial listening expectations

The significant three-way interaction discussed above indicates that the drop in performance with decreasing certainty is offset by prior knowledge of the target callsign when listening in real spatial separation conditions. One possibility is that prior knowledge of the target callsign enables the listener to draw on the interaural cues available with actual spatial separation in order to selectively attend to targets originating from unexpected locations. To further investigate this possibility, we directly compared performance for targets appearing in expected and unexpected locations. For this analysis, we focused on the callsign cue conditions where target identity was known before stimulus presentation and the conditions in which location certainty was less than 1.0 but more than chance (0.33). Whereas the callsign cue established listener expectations regarding target identity, by choosing the intermediate location certainty conditions it was possible to compare trials where the target was presented at the more likely central location or a less likely side location. For example, when the probability of the target being presented at the center location was 0.80 or 0.60, the listener would need to allocate some attentional resources from the expected central location to an unexpected side location on some trials. A "likely" trial would occur if the target callsign occurred at the center location and an "unlikely" trial would occur if the target callsign occurred at either the left or right spatial location. The ability of listeners to allocate spatial attention was gauged by comparing their word identification performance on trials where the target was presented at the likely central location versus the unlikely side locations. The difference in performance between likely and unlikely trials should be smaller for listeners who are better able to reallocate attention from expected to unexpected spatial locations. If there are age differences in the ability to reallocate spatial attention on this task, then there would be a greater difference in performance between likely and unlikely trials for older compared to younger listeners.

As shown in Fig. 3, for both age groups, word identification was approximately 45 percentage points better on likely than unlikely trials. Importantly, although younger adults performed better than older adults by an average of nine percentage points when the results are collapsed across

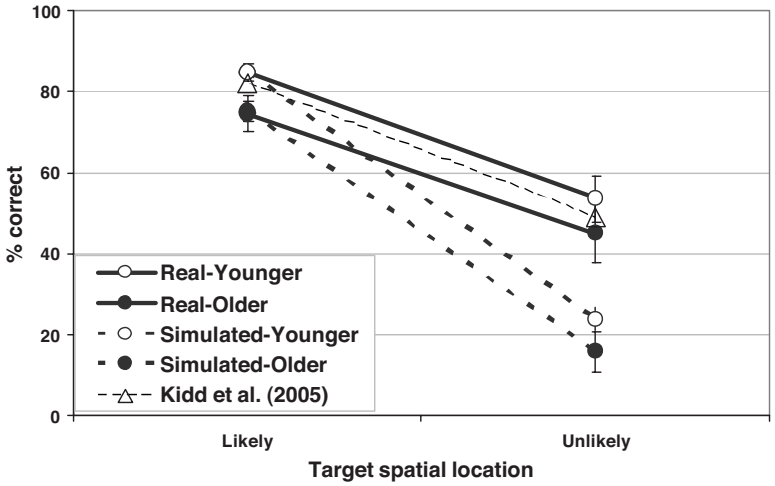


FIG. 3. Mean percent correct identification scores and standard errors of the mean for real (solid lines) and simulated spatial separation (dashed lines) presentation conditions, depicted for the "likely" or "unlikely" spatial locations. Unfilled and filled circles represent data collected on younger and older adults, respectively. Triangles represent the results of Kidd *et al.* (2005) for younger adults when performance was averaged over conditions in which the likely location was left, center, or right. Data are collapsed across location probabilities 0.80 and 0.60.

the two presentation methods, the cost of reallocating attention from a likely to an unlikely spatial location was similar for older (44%) and younger adults (46%).

In order to statistically confirm these descriptions, a repeated-measures ANOVA was conducted with age (younger versus older) as the between-subjects variable, and presentation method (real versus simulated spatial separation) and cue validity (likely versus unlikely) as within-subjects variables. We found a significant main effect of cue validity [$F(1,14)=152.53$, $p<0.001$] indicating that performance was better for likely than unlikely spatial locations. Although we found a borderline main effect of age [$F(1,14)=4.14$, $p=0.06$], we did not observe a significant cue validity by age interaction [$F(1,14)=0.05$, $p>0.05$], suggesting that older adults did not exhibit more difficulty than younger adults reallocating attentional focus.

F. The role of natural spatial cues

The influence of the richness of the interaural cues on the cost of reallocating attention from a likely to an unlikely spatial location was examined by comparing performance in the real compared to the simulated spatial separation conditions for the callsign cue conditions where target identity was known before stimulus presentation and where location certainty was 0.80 or 0.60. As shown in Fig. 3, the richness of the interaural cues strongly affected word identification performance for the unlikely trials, with scores being 29 percentage points higher in the real compared to the simulated spatial separation condition. However, the availability of rich interaural cues did not affect performance on the likely trials, with there being less than a one percentage point difference between the real and simulated spatial separation conditions. Again, the same pattern was observed for both age groups. The description of this pattern was confirmed with a significant main effect of presentation method [$F(1,14)=98.33$, $p<0.001$] as well as a presentation method by cue validity interaction [$F(1,14)=59.34$, $p<0.001$]. *Post hoc* Student–Newman–Keuls tests revealed no significant difference between the real and simulated spatial separations conditions for likely trials ($p>0.05$), but significant differences between the two presentation methods for unlikely trials ($p<0.01$), suggesting that the benefit resulting from the availability of rich interaural cues in the real spatial separation conditions is contingent upon a listener's spatial expectations. Finally, we failed to obtain a significant presentation method × cue validity × age interaction [$F(1,14)=0.00$, $p>0.05$], suggesting that younger and older adults are equally disadvantaged from a reduction in auditory cues associated with simulated spatial separation across both likely and unlikely spatial listening locations. No other significant effects were obtained.

IV. DISCUSSION

The goal of the current investigation was to determine if there are age-dependent differences in word identification that could be attributable to the use of acoustic cues and/or auditory spatial attention. First, we discuss our replication of previous findings (Kidd *et al.*, 2005) for younger adults in the real spatial separation conditions, and then we compare the results we found for younger and older listeners in the real and simulated spatial separation conditions.

A. Performance of younger listeners

On average, the results we found for younger adults tested in the real spatial separation conditions are a very good replication of the findings reported previously (Kidd *et al.*, 2005). Overall, the mean word identification score was three percentage points higher for younger listeners in the present study than for those in the prior study. When the callsign cue was specified prior to sentence presentation, the scores for our participants and those in the previous study were 77% and 75%, respectively. When the callsign cue was specified following sentence presentation, the corresponding scores were 62% and 59%. Furthermore, as shown in Fig. 4, the pattern of results was almost identical across the different conditions that were tested.

B. The effect of age on performance

When there was real spatial separation between the sound sources, in every condition the older adult group demonstrated significantly poorer word identification performance compared to the younger adult group. Importantly, based on the lack of significant interactions between age and

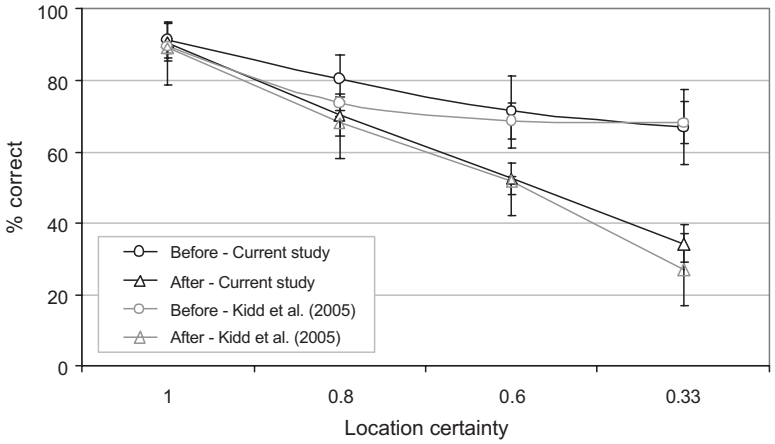


FIG. 4. Mean percent correct word identification scores, and standard deviations of the means, from the younger adults listening with real spatial separation in the current study (darker lines) and in the study of Kidd *et al.* (2005) (lighter lines) for the four location certainties. Circles indicate callsign cue before conditions and triangles indicate callsign cue after conditions.

the other independent variables (presentation method, callsign cue, or location certainty), the pattern of findings indicates that there are general, but no condition-specific age-dependent differences in the ability to use auditory spatial attention to follow target speech when there are three simultaneous talkers. Furthermore, comparing the results obtained in the real and simulated spatial separation conditions, we found that both younger and older adults performed significantly better when there was real spatial separation than when there was simulated spatial separation, but the older age group was not disproportionately impaired in the simulated spatial separation condition.

The finding that older listeners were not disproportionately disadvantaged in the simulated spatial separation conditions indicates that there were no relevant age-dependent differences in the precedence effect. In previous research examining the precedence effect in which loudspeaker delays were manipulated, older listeners had more difficulty fusing clicks compared to younger listeners for time delays less than 0.7 ms (Cranford *et al.*, 1990; Cranford *et al.*, 1993). However, our time delay was much longer (3 ms) and our findings are consistent with a previous psychoacoustic study that failed to observe age-dependent differences in the precedence effect for longer time delays (Schneider *et al.*, 1994; Roberts and Lister, 2004).

No doubt there are multiple mechanisms underpinning the overall poorer word identification abilities of the older adults compared to younger adults. One possible explanation that must be considered is that the older adults could be in the early stages of presbycusis. Although both age groups had normal audiometric thresholds for frequencies in the speech range, at higher frequencies (4 and 8 kHz), the average differences between the thresholds of the older and younger groups were 14 and 17 dB, respectively (see Tables I and II). Note that of the 32 thresholds measured above the speech range in the older listeners (two ears × eight participants × two frequencies), only nine thresholds were classified as clinically abnormal, but all of these fell within the mild hearing loss category. Nevertheless, a high-frequency hearing loss could have contributed to the main effect of age observed in this study. Previous research suggests that listeners with impaired high-frequency hearing fail to fully benefit from spatial separation between a target and noise because the improvement in SNR in the higher frequencies arising from head shadow effects is reduced or eliminated (Bronkhorst and Plomp 1989), or because poorer binaural interaction reduced the benefit that would normally be gained from spatial separation when listening with one or two ears (Bronkhorst and Plomp, 1988). Consequently, the poorer performance observed in our older listeners may reflect an inability to take advantage of head shadow effects or binaural interaction. However, we argue that this explanation is unable to account for the overall pattern of results in the present study. If the better performance of the younger adults compared to the older adults were due to the superior high-frequency hearing thresholds and ability of the younger listeners to use high-frequency cues to take advantage of spatial separation, then there should have been a greater age-dependent deficit in the real spatial separation conditions than in the simulated spatial separation conditions where these cues were virtually eliminated. As shown in Fig. 2, younger and older listeners exhibited similar reductions in performance when comparing the results of real and simulated spatial separation conditions, suggesting that the high-frequency audiometric deficits exhibited by our older adults cannot account for the pattern of results related to the benefit arising from spatial separation.

Our finding of a main effect of age in the real spatial separation condition is consistent with previous studies where speech-in-noise tests were conducted in conditions of spatial separation (Dubno *et al.*, 2002; Gelfand *et al.*, 1988; Li *et al.*, 2004). When considering the benefit from spatial separation, however, there is little evidence suggesting that older adults benefit less from spatial separation than younger adults. Gelfand *et al.* (1988) using the revised speech perception in noise test (Bilger *et al.*, 1984) found no age-dependent differences in benefit from spatial separation; however, Dubno *et al.* (2002) using the hearing in noise test (Nilsson *et al.*, 1994) found that the benefit from spatial separation was approximately 1 dB less for older adults with normal hearing relative to younger adults. In our study, we found that older adults benefited as much as younger adults from spatial separation either using real (Gelfand *et al.* (1988)) or simulated (Li *et al.* (2004)) spatial separation. Hence, the available evidence suggests that there is little (or no) age-dependent difference in the ability to use spatial separation for individuals with normal audiometric thresholds.

C. Auditory perception of spatial separation

One possible explanation that may account for the observed differences in performance between the real and simulated spatial separation conditions is the difference in the precise nature of the perceived location of sound sources between the two conditions. For example, the simulated spatial sound sources may have appeared to sound further or closer than the sentences presented in the corresponding real spatial separation conditions. However, the pattern of results found by Li *et al.* (2004) suggests that this explanation is unlikely to account for the differences in question. In their study, the simulated location of the target was to the right of the listener and the simulated locations of the maskers were to the right, center, or left of the listener. Relative to the condition where the masker was simulated on the right, Li *et al.* (2004) observed a similar masking release of 4.8 dB regardless of whether the simulated masker was located at the center or left position. Hence, it seems unlikely that the minor differences in perceived spatial location of the sound sources between the real and simulated conditions could explain the pattern of results we observed. Nevertheless, to our knowledge no previous experiment has attempted to simulate three simultaneous, spatially separated sentences using the precedence effect, and it would be prudent to interpret the results with caution. Future research should more carefully investigate to what extent differences observed between presentation methods using real and simulated spatial separations are due to a loss of natural binaural cues and which differences are due to failures of the precedence effect to assist in the formation of clearly localized auditory objects.

In the current investigation, the precedence effect was used to simulate spatial separation in order to isolate the contribution of the rich, natural binaural cues that are available with real spatial separation. Although no benefit was observed when the callsign was cued after stimulus presentation,[2] the advantage of full binaural cues became apparent when the callsign was cued before stimulus presentation and was increasingly apparent when the certainty about the location of the target decreased for both age groups.

D. Selective auditory spatial attention

In order to further examine the extent to which performance is governed by auditory and/or attentional factors, we conducted an analysis focusing on the relationship between spatial listening expectations and the benefit derived from auditory cues associated with spatial separation. Specifically, we compared the results obtained in the real and simulated spatial separation conditions using only trials on which the target was presented from a likely or unlikely spatial location. Performance was similar when stimuli were presented with real or simulated spatial separation (<1 percentage point difference) at likely locations, but there was a considerable performance difference (>29 percentage points) between real and simulated spatial separation conditions for targets presented at unlikely spatial locations (see Fig. 3). Seemingly, the cognitive influence of location expectation may modulate the importance of the natural auditory binaural cues associated with real spatial separation. The usefulness of full binaural cues may be more critical in ambiguous and/or dynamic listening situations in which the listener is required to selectively reallocate spatial attention because the availability of full binaural cues facilitates sound localization, thus allowing for greater precision with which to focus attentional resources.

At unlikely listening locations, all listeners performed more poorly when listening with simulated rather than real spatial separation. When the likely location of the target is at the center (as it always was in our study), and the target is presented from an unlikely side location, there is a monaural SNR advantage in the real spatial separation conditions for side targets that is not available in the simulated spatial separation conditions. The availability of this SNR advantage could explain why performance was poorer in the simulated than it was in the real spatial separation conditions when the target appeared in an unlikely position. However, because we did not vary the location at which the target was likely to appear (as did Kidd *et al.*, 2005), we were not able to determine the extent to which performance at the unexpected locations in the real spatial separation condition varied as a function of the expected location of the target. Nevertheless, because Kidd *et al.* (2005) only reported results for likely and unlikely locations that were averaged over the different possible locations at which the target was expected, we can only compare their averaged data to our results. Figure 4 shows that our results in the real spatial separation conditions are only slightly better than those of Kidd *et al.* (2005), with the difference between studies being only 3 percentage points at the expected and 4.5 percentage points when the target is presented at an unexpected location. Hence, although we cannot resolve whether the pattern of results found here for real spatial separation conditions would vary with the location at which the target was expected, we note that our results replicate the data (averaged over the different locations at which the target was expected) of Kidd *et al.* (2005)

Assuming our view that the data truly reflect differences between likely and unlikely listening locations and do not arise from a monaural SNR advantage for side locations in conditions using real spatial separation, then several mechanisms may potentially account for this pattern of results. One possibility is that the listener's probability of attending to a spatial position matches the probability that the target will appear at that spatial position (for a review, see Vulkan, 2000). However, Kidd *et al.* (2005) rejected this alternative because if listeners are using a probability rule, then one would expect that the proportion of responses from an expected location in callsign-after conditions would closely approximate the associated location probability. Since Kidd *et al.* (2005) found that under such conditions listeners made substantially more responses from expected locations than would be predicted by a probability-matching strategy, they concluded that there is little evidence that listeners adopt such a strategy. A second possibility is that observers always begin each trial focused on the likely location and then switch attention to the unlikely location. A third possibility is that as uncertainty increases, listeners broaden the spatial extent of locations to which attention is allocated or increase

the width of the attentional "beam." Based on the methodology used in this study, it is currently unclear if listeners (a) switch a more tightly focused attentional beam from likely to unlikely locations, (b) more broadly tune their auditory attentional filter to encompass more spatial locations, or (c) perform the listening task by incorporating elements of both attention switching and broadened attentional focus. Future research should more clearly delineate between these possible alternatives.

E. Interaction of auditory and attentional factors

We have shown that the performance deficits observed in both age groups at unlikely spatial listening locations are modulated by the availability of full binaural cues, with the deficits being substantially larger in the simulated than in the real spatial location conditions. Currently, it is unclear whether the poorer performance at unlikely locations for simulated spatial locations relative to real spatial locations is primarily due to disruptions of cues arising from (a) monaural signal-to-noise differences, (b) ILDs, or (c) ITDs. Because the current methodology is unable to definitively identify the relative contributions of particular acoustic cues to word identification performance at unlikely locations in a spatially complex listening situation, future research will need to develop testing situations that are better able to do so.

As noted above, the relative contributions of location expectations and the natural binaural cues associated with real spatial separation were similar for younger and older adults. Although younger adults outperformed older adults by an average of approximately nine percentage points on word identification performance across the likely and unlikely spatial listening locations, the loss of full binaural cues did not disproportionately impair the older listeners. To our knowledge no previous research has examined the role of binaural cues in auditory spatial attention in older adults. The general finding in vision studies of attention and aging is that disproportionate age-dependent differences are found for slower, more controlled behaviors, but not for faster, more automatic behaviors that incorporate spatial attention switching (for a review, see McDowd and Shaw, 2000). It may be that speech perception in auditory spatial displays follows the pattern of age-dependent results found in vision for faster, reflexive behaviors.

V. CONCLUSIONS

The results of the current investigation replicate and extend the findings of Kidd *et al.* (2005). Although older adults demonstrated significantly poorer word identification performance compared to younger adults, both age groups benefited from information about target callsign identity and location certainty. Comparing real and simulated spatial separations, performance was reduced by the removal of the binaural cues associated with real spatial separation, but older adults were not more impaired by the removal of these cues than were younger adults. We also observed that for both age groups, the richness of the available binaural cues had no impact on word identification at "likely" spatial locations, but that there were significant performance deficits observed at "unlikely" spatial listening locations when full binaural cues were not available. It seems that natural binaural cues are increasingly important in more ambiguous listening environments. Overall, these findings indicate that the benefit from spatial separation is governed by an interaction of cognitive and acoustical factors for both age groups.

ACKNOWLEDGMENTS

The authors would like to thank all the participants for their dedicated listening efforts. Funding for this research was provided in part by the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council, and the Canada Foundation for Innovation. The authors would like to acknowledge Robert Bolia for providing them with the CRM materials, and Rabia Murad and Marco Coletta who assisted with data collection and analysis.

[1]One possibility is that auditory spatial attention may facilitate the inhibitory mechanisms for suppressing the activation of goal-irrelevant information (e.g., Hasher and Zacks, 1988)

[2]Kidd *et al.* (2005) suggested that, presumably because of the difficulty associated with this listening task when the callsign cue was provided after stimulus presentation (i.e., simultaneously attending to and keeping in memory three spatially distributed speech streams), the strategy adopted by most listeners was to attend to the most "likely" location. See their article for a detailed discussion of performance patterns for the callsign cue after conditions.

Arbogast, T. L., and Kidd, G., Jr. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.

Best, V., Gallun, F. J., Ihlefeld, A., and Shinn-Cunningham, B. G. (**2006**). "The influence of spatial separation on divided listening," J. Acoust. Soc. Am. **120**, 1506–1516.

Bilger, R. C., Nuetzel, M. J., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT Press, Cambridge, MA).

Broadbent, D. E. (**1958**) *Perception and Communication* (Pergamon Press, London).

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. **86**, 117–128.

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.

Bronkorst, A. W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–1383.

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

CHABA (Committee on Hearing, Bioacoustics, and Biomechanics) (**1988**), "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859–895.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Colburn, H. S., Shinn-Cunningham, B., Kidd, G., Jr., and Durlach, N. (**2006**). "The perceptual consequences of binaural hearing," Int. J. Audiol. **45**, S34–S44.

Cranford, J. L., Andres, M. A., Piatz, K. K., and Reissig, K. L. (**1993**). "Influences of age and hearing loss on the precedence effect in sound localization," J. Speech Hear. Res. **36**, 437–441.

Cranford, J. L., Boose, M., and Moore, C. A. (**1990**). "Effects of aging on the precedence effect in sound localization," J. Speech Hear. Res. **33**, 654–659.
Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.
Divenyi, P., and Simon, H. (**1999**). "Hearing in aging: Issues old and young," Current Opinion in Otolaryngology and Head Neck Surgery **7**, 282–289.
Dubno, J. R., Ahlstrom, J. B., and HorwitzA. R. (**2002**). "Spectral contributions to the benefit from spatial separation of speech in noise," J. Speech Lang. Hear. Res. **45**, 1297–1310.
Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am. **76**, 87–96.
Duquesnoy, A. J. (**1983**). "The intelligibility of sentences in quiet and in noise in aged listeners," J. Acoust. Soc. Am. **74**, 1136–1144.
Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.
Ebata, M. (**2003**). "Spatial unmasking and attention related to the cocktail party problem," Acoust. Sci. & Tech. **24**, 208–219.
Folk, C. L. and Hoyer W. J. (**1992**). "Aging and shifts of visual spatial attention," Psychol. Aging **7**, 453–465.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Frisina, D. R., and Frisina, R. D. (**1997**). "Speech recognition in noise and presbycusis: Relations to possible neural mechanisms," Hear. Res. **106**, 95–104.
Gelfand, S. A., Ross, L., and Miller, S. (**1988**). "Sentence recognition in noise from one versus two sources: Effects of aging and hearing loss," J. Acoust. Soc. Am. **83**, 248–257.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1995**). "Recognition of multiply degraded speech by young and elderly listeners," J. Speech Hear. Res. **38**, 1150–1156.
Greenwood, P. M. and Parasuraman R. (**1994**). "Attentional disengagement deficit in nondemented elderly over 75 years of age," Aging and Cognition **1**, 188–202.
Grose, J. H. (**1996**). "Binaural performance and aging," J. Am. Acad. Audiol **7**, 168–174.
Hasher, L., and Zacks, R. T. (**1988**). "Working memory comprehension, and aging: A review and new view," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, edited by G. H. Bower (Academic, San Diego, CA), Vol. **22**, pp. 193–225.
Haykin, S., and Chen, Z. (**2005**). "The cocktail party problem," Neural Comput. **17**, 1875–1902.
Helfer, K. S. (**1992**). "Aging and the binaural advantage in reverberation and noise," J. Speech Hear. Res. **35**, 1394–1401.
Herman, G. E., Warren, L. R., and Wagener, J. W. (**1977**). "Auditory lateralization: Age differences in sensitivity to dichotic time and amplitude cues," J. Gerontol. **32**, 187–191.
Hornsby, B. W., Ricketts, T. A., and Johnson, E. E. (**2006**). "The effects of speech and speechlike maskers on unaided and aided speech recognition in persons with hearing loss," J. Am. Acad. Audiol **17**, 432–447.
Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.
Koehnke, J., and Besing, J. (**2001**). "The effects of aging on binaural and spatial hearing," Semin. Hear. **22**, 2415–2453.
Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.
Li, L., and Yue, Q. (**2002**). "Auditory gating processes and binaural inhibition in the inferior colliculus," Hear. Res. **168**, 98–109.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.
Madden, D. J., Connelly, S. L., and Pierce T. W. (**1994**). "Adult age differences in shifting focused attention," Psychol. Aging **9**, 528–538.
McDowd, J. M., and Shaw, R. J. (**2000**). "Aging and attention: A functional perspective," in *The Handbook of Aging and Cognition*, 2nd ed., edited by T. A. Salthouse, and F. I. M. Craik (Lawrence Erlbaum Associates, Mahwah, NJ).
Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting and focusing auditory spatial attention,"J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.
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.
Pashler, H. (**1993**). "Dual-task interference and elementary mental mechanisms," in *Attention and Performance*, edited by D. Meyer and S. Kornblum (MIT Press, Cambridge, MA), Vol. **14**, pp. 245–264.
Pichora-Fuller, M. K. (**1997**). "Language comprehension in older listeners," J. Speech Lang. Path. Audiol. **21**, 125–142.
Pichora-Fuller, M. K. (**2003**). "Cognitive aging and auditory information processing," Int. J. Audiol. **42**, 2S26–2S32.
Pichora-Fuller, M. K. (**2007**). "Audition and cognition: What audiologists need to know about listening," in *Hearing care for adults*, edited by C. Palmer and R. Seewald (Phonak, Stafa, Switzerland), pp. 71–85.
Pichora-Fuller, M. K., and Schneider, B. A. (**1992**). "The effect of interaural delay of the masker on masking-level differences in young and old adults," J. Acoust. Soc. Am. **91**, 2129–2135.
Pichora-Fuller, M. K., and Schneider, B. A. (**1998**). "Masking-level differences in older adults: The effect of the level of the masking noise," Percept. Psychophys. **60**, 1197–1205.
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. (**2003**). "Effects of aging on auditory processing of speech," Int. J. Audiol. **42**, S11–S16.
Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).
Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.
Roberts, R., and Lister, J. (**2004**). "Effects of age and hearing loss on gap detection and the precedence effect: Broad-band stimuli," J. Speech Lang. Hear. Res. **47**, 965–978.
Sach, A. J., and Bailey, P. J. (**2004**). "Some characteristics of auditory spatial attention revealed using rhythmic masking release," Percept. Psychophys. **66**, 1379–1387.
Sach, A. J., Hill, N. I., and Bailey, P. J. (**2000**). "Auditory spatial attention using interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **26**, 717–729.
Schneider, B. A., Daneman, M., and Pichora-Fuller, M. K. (**2002**). "Listening in aging adults: From discourse comprehension to psychoacoustics," Can. J. Exp. Psychol. **56**, 139–152.
Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and adults," J. Acoust. Soc. Am. **95**, 980–991.
Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.
Tun, P. A., and Wingfield, A. (**1999**). "One voice too many: Adult age differences in language processing with different types of distracting sounds," J. Gerontol. B Psychol. Sci. Soc. Sci. **54**, P317–P27.
Verhaeghen, P., and Cerella, J. (**2002**). "Aging, executive control, and attention: A review of meta-analyses," Neurosci. Biobehav Rev. **26**, 849–857.
Vulkan, N. (**2000**). "An economist's perspective on probability matching," J. Econ. Surv. **14**, 101–118.
Wallach H., Newman E. B., Rosenzweig M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336.
Watson, C. S. (**1987**). "Uncertainty, informational masking and the capacity of immediate auditory meaning," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.
Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.
Wu, X., Wang, C., Chen, J., Qu, H., Li, W., Wu, Y., Schneider, B. A., and Li, L. (**2005**). "The effect of perceived spatial separation on informational masking of Chinese speech." Hear. Res. **199**, 1–10.
Yost, W. (**1997**). "The cocktail party problem: Forty years later." in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. R Auderson (Lawrence Earlbaum Associates, Hillsdale, NJ) pp. 329–348.
Zurek, P. M. (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 953–964.

Segregation of unvoiced speech from nonspeech interference

Guoning Hu[a)]
Biophysics Program, The Ohio State University, Columbus, Ohio 43210

DeLiang Wang[b)]
Department of Computer Science and Engineering and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

(Received 5 September 2007; revised 7 May 2008; accepted 8 May 2008)

Monaural speech segregation has proven to be extremely challenging. While efforts in computational auditory scene analysis have led to considerable progress in voiced speech segregation, little attention has been given to unvoiced speech, which lacks harmonic structure and has weaker energy, hence more susceptible to interference. This study proposes a new approach to the problem of segregating unvoiced speech from nonspeech interference. The study first addresses the question of how much speech is unvoiced. The segregation process occurs in two stages: Segmentation and grouping. In segmentation, the proposed model decomposes an input mixture into contiguous time-frequency segments by a multiscale analysis of event onsets and offsets. Grouping of unvoiced segments is based on Bayesian classification of acoustic-phonetic features. The proposed model for unvoiced speech segregation joins an existing model for voiced speech segregation to produce an overall system that can deal with both voiced and unvoiced speech. Systematic evaluation shows that the proposed system extracts a majority of unvoiced speech without including much interference, and it performs substantially better than spectral subtraction. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2939132]

PACS number(s): 43.72.Dv [DOS] Pages: 1306–1319

I. INTRODUCTION

In a daily environment, target speech is often corrupted by various types of acoustic interference, such as crowd noise, music, and other voices. Acoustic interference poses a serious problem for many applications including hearing aid design, automatic speech recognition (ASR), telecommunication, and audio information retrieval. In the hearing aid application, for example, it is well known that listeners with hearing loss have substantially greater difficulty in understanding speech in a noisy background (Moore, 2007). Hearing aids improve the audibility of noisy speech by means of amplification. However, their ability to improve the intelligibility of noisy speech is very limited, and how to remove or attenuate background noise is considered one of the biggest challenges facing hearing aid design (Dillon, 2001). Applications like this often require speech segregation. In addition, in many practical situations, monaural segregation is either necessary or desirable. Monaural speech segregation is especially difficult because one cannot utilize spatial filtering afforded by a microphone array to separate sounds from different directions. For monaural segregation, one has to consider the intrinsic properties of target speech and interference in order to disentangle them. Various methods have been proposed for monaural speech enhancement (Benesty *et al.*, 2005), and they usually assume stationary and quasistationary interference and achieve speech enhancement based on certain assumptions or models of speech and interference. These methods tend to lack the capacity to deal with general interference as the variety of interference makes it very difficult to model and predict.

While monaural speech segregation by machines remains a great challenge, the human auditory system shows a remarkable ability for this task. The perceptual segregation process is called auditory scene analysis (ASA) by Bregman (1990), who considers ASA to take place in two conceptual stages. The first stage, called segmentation (Wang and Brown, 1999), decomposes the auditory scene into sensory elements (or segments), each of which should primarily originate from a single sound source. The second stage, called grouping, aggregates the segments that likely arise from the same source. Segmentation and grouping are governed by perceptual principles, or ASA cues, which reflect intrinsic sound properties, including harmonicity, onset and offset, location, and prior knowledge of specific sounds (Bregman, 1990; Darwin, 1997).

Research in ASA has inspired considerable work in computational ASA (CASA) [for a recent, extensive review see Wang and Brown (2006)]. Many CASA studies have focused on monaural segregation and have performed the task without making strong assumptions about interference. Mirroring the two-stage model of ASA, a typical CASA system includes separate stages of segmentation and grouping that operate on a two-dimensional time-frequency (T-F) representation of the auditory scene (see Wang and Brown, 2006, Chap. 1). The T-F representation is typically created by an auditory peripheral model that analyzes an acoustic input by an auditory filterbank and decomposes each filter output

a)Electronic mail: hu.117@osu.edu
b)Electronic mail: dwang@cse.ohio-state.edu

 0001-4966/2008/124(2)/1306/14/$23.00

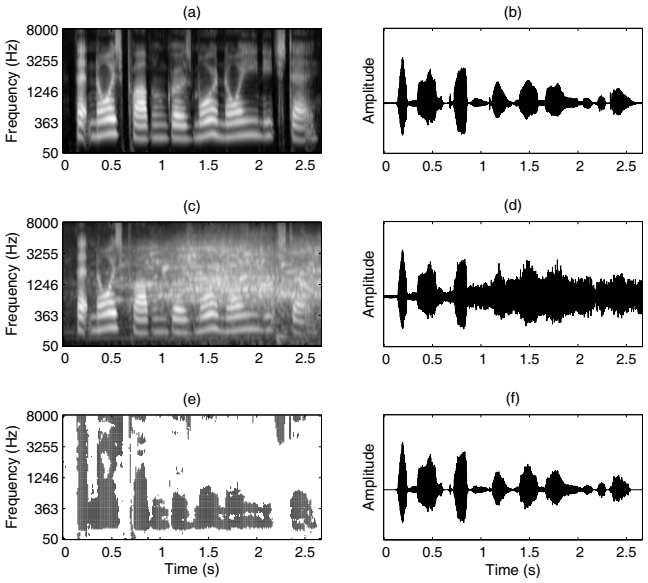


FIG. 1. CASA illustration. (a) T-F decomposion of a female utterance, "That noise problem grows more annoying each day." (b) Waveform of the utterance. (c) T-F decomposition of the utterance mixed with a crowd noise. (d) Waveform of the mixture. (e) Target stream composed of all the T-F units (black regions) dominated by the target (ideal binary mask). (f) Waveform resynthesized from the target stream.

into time frames. The basic element of the representation is called a T-F unit, corresponding to a filter channel and a time frame.

We have suggested that a reasonable goal of CASA is to retain the mixture signals within the T-F units where target speech is more intense than interference and to remove others (Hu and Wang, 2001; 2004). In other words, the goal is to compute a binary T-F mask, referred to as an ideal binary mask, where 1 indicates that the target is stronger than interference in the corresponding T-F unit and 0 otherwise. See Wang (2005) and Brungart *et al.* (2006) for more discussions on the notion of the ideal binary mask and its psychoacoustical support.

As an illustration, Fig. 1(a) shows a T-F representation of the waveform signal in Fig. 1(b). The signal is a female utterance, "That noise problem grows more annoying each day," from the TIMIT database (Garofolo *et al.*, 1993). The peripheral processing is carried out by a 128-channel gammatone filterbank with 20-ms time frames and a 10-ms frame shift (see Sec. III A for details). Figures 1(c) and 1(d) show the corresponding representations of a mixture of this utterance and crowd noise, where the signal-to-noise ratio (SNR) is 0 dB. In Figs. 1(a) and 1(c) a brighter unit indicates stronger energy. Figure 1(e) illustrates the ideal binary mask for the mixture in Fig. 1(d). With this mask, target speech can then be synthesized by retaining the filter responses of the T-F units having the value of 1 and eliminating the filter responses of the units of the value of 0. Figure 1(f) shows the synthesized waveform signal, which is close to the clean utterance in Fig. 1(b).

Natural speech contains both voiced and unvoiced portions (Stevens, 1998; Ladefoged, 2001). Voiced speech consists of portions that are mainly periodic (harmonic) or quasiperiodic. Previous CASA and related separation studies have focused on segregating voiced speech based on harmonicity (Parsons, 1976; Weintraub, 1985; Brown and Cooke, 1994; Hu and Wang, 2004). Although substantial advances have been made on voiced speech segregation, unvoiced speech segregation has not been seriously addressed and remains a major challenge. A recent system by Radfar *et al.* (2007) exploits vocal-tract filter characteristics (spectral envelopes) to separate two voices, which have the potential to deal with unvoiced speech. However, it is not clear how well their system performs when both speakers utter unvoiced speech and the assumption of two-speaker mixtures limits the scope of application.

Compared with voiced speech segregation, unvoiced speech segregation is clearly more difficult for two reasons. First, unvoiced speech lacks harmonic structure and is often acoustically noiselike. Second, the energy of unvoiced speech is usually much weaker than that of voiced speech; as a result, unvoiced speech is more susceptible to interference. Nevertheless, both voiced and unvoiced speech carry crucial information for speech understanding, and both need to be segregated.

In this paper, we propose a CASA system to segregate unvoiced speech from nonspeech interference. For auditory segmentation, we apply a multiscale analysis of event onsets and offsets (Hu and Wang, 2007), which has the important property that segments thus formed correspond to both voiced and unvoiced speech. By limiting interference to nonspeech signals, we propose to identify and group segments corresponding to unvoiced speech by a Bayesian classifier that decides whether segments are dominated by unvoiced speech on the basis of acoustic-phonetic features derived from these segments. The proposed algorithm, together with our previous system for voiced speech segregation (Hu and Wang, 2004; 2006), leads to a CASA system that segregates both unvoiced and voiced speech from nonspeech interference.

Before tackling unvoiced speech segregation, we first address the question of how much speech is unvoiced. This is the topic of the next section. Section III describes early stages of the proposed system, and Sec. IV details the grouping of unvoiced speech. Section V presents systematic evaluation results. Further discussions are given in Sec. VI.

II. HOW MUCH SPEECH IS UNVOICED?

Voiced speech refers to the part of speech signal that is periodic (harmonic) or quasiperiodic. In English, voiced speech includes all vowels, approximants, nasals, and certain stops, fricatives, and affricates (Stevens, 1998; Ladefoged, 2001). It comprises a majority of spoken English. Unvoiced speech refers to the part that is mainly aperiodic. In English, unvoiced speech comprises a subset of stops, fricatives, and affricates. These three consonant categories contain the following phonemes:

(1) Stops: /t/, /d/, /p/, /b/, /k/, and /g/.
(2) Fricatives: /s/, /z/, /f/, /v/, /ʃ/, /ʒ/, /θ/, /ð/, and /h/.
(3) Affricates: /tʃ/ and /dʒ/.

TABLE I. Occurrence percentages of six consonant categories.

Phoneme type	Conversational	Written	TIMIT
Voiced stop	6.7	6.9	7.9
Unvoiced stop	15.1	11.9	12.8
Voiced fricative	7.5	9.5	7.7
Unvoiced fricative	8.6	8.6	9.8
Voiced affricate	0.3	0.4	0.6
Unvoiced affricate	0.3	0.5	0.5
Total	38.5	37.8	39.3

TABLE II. Duration percentages of six consonant categories.

Phoneme type	Conversational	TIMIT
Voiced stop	5.6	5.2
Unvoiced stop	16.2	12.9
Voiced fricative	5.3	5.8
Unvoiced fricative	9.6	12.0
Voiced affricate	0.3	0.6
Unvoiced affricate	0.4	0.7
Total	37.4	37.2

In phonetics, all these phonemes, except /h/, are called obstruents. To simplify notations, we refer to the above phonemes as expanded obstruents. Eight of the expanded obstruents, /t/, /p/, /k/, /s/, /f/, /ʃ/, /θ/, and /tʃ/, are categorically unvoiced. In addition, /h/ may be pronounced either in the voiced or the unvoiced manner. The other phonemes are categorized as voiced, although in articulation they often contain unvoiced portions. Note that an affricate can be treated as a composite phoneme, with a stop followed by a fricative.

Dewey (1923) conducted an extensive analysis of the relative frequencies of individual phonemes in written English, and this analysis concludes that unvoiced phonemes account for 21.0% of the total phoneme usage. For spoken English, French *et al.* (1930) [see also Fletcher (1953)] conducted a similar analysis on 500 telephone conversations containing a total of about 80 000 words and concluded that unvoiced phonemes account for about 24.0%. Another extensive phonetically labeled corpus is the TIMIT database, which contains 6300 sentences read by 630 different speakers from various dialect regions in America (Garofolo *et al.*, 1993). Note that the TIMIT database is constructed to be phonetically balanced. Many of the same sentences are read by multiple speakers, and there are a total of 2342 different sentences. We have performed an analysis of relative phoneme frequencies for distinct sentences in the TIMIT corpus, and found that unvoiced phonemes account for 23.1% of the total phonemes.

Table I shows the occurrence percentages of six phoneme categories from these studies. Several observations may be made from the table. First, unvoiced stops occur much more frequently than voiced stops, particularly in conversations where they occur more than twice as often as their voiced counterparts. Second, affricates are used only occasionally. It is remarkable that the percentages of the six consonant categories are comparable despite the fact that written, read, and conversational speech are different in many ways. In particular, the total percentages of these consonants are almost the same for the three different kinds of speech.

What about the relative durations of unvoiced speech in spoken English? Unfortunately, the data reported on the telephone conversations (French *et al.*, 1930) do not contain durational information. To get an estimate, we use the durations obtained from a phonetically transcribed subset of the switchboard corpus (Greenberg *et al.*, 1996), which also consists of conversations over the telephone. The amount of labeled data in the Switchboard corpus, i.e., 72 min of conversation, is much smaller than that in the telephone conservations analyzed by French *et al.* (1930). Hence we do not use the labeled Switchboard corpus to obtain phoneme frequencies; instead we assign the median durations from the transcription to the occurrence frequencies in the telephone conservations in order to estimate the relative durations of unvoiced sounds. Table II lists the resulting duration percentages of six phoneme categories. Also listed in the table are the corresponding data from the TIMIT corpus. The table shows that for stops and fricatives, unvoiced sounds last much longer than their voiced counterparts. In addition, affricates have a minor contribution in terms of duration, similar to that in terms of occurrence frequency. Once again, the percentages from conversational speech are comparable to those from read speech. In terms of overall time duration, unvoiced speech accounts for 26.2% in telephone conversations and 25.6% in the read speech of the TIMIT corpus. These duration percentages are a little higher than the corresponding frequency percentages.

Tables I and II show that unvoiced sounds account for more than 20% of spoken English in terms of both occurrence frequency and time duration. In addition, since voiced obstruents are often not entirely voiced, unvoiced speech may occur more than suggested by the above estimates.

III. EARLY PROCESSING STAGES

Our proposed system for unvoiced speech segregation has the following stages of computation: Peripheral analysis, feature extraction, auditory segmentation, and grouping. In this section, we describe the first three stages. The stage of grouping is described in the next section. A list of all the symbols used in system description is given in the Nomenclature.

A. Auditory peripheral analysis

This stage derives a T-F representation of an input scene by performing a frequency analysis using a gammatone filterbank (Patterson *et al.*, 1988), which models human cochlear filtering. Specifically, we employ a bank of 128 gammatone filters, whose center frequencies range from 50 to 8000 Hz; this frequency range is adequate for speech understanding (Fletcher, 1953; Pavlovic, 1987). The impulse response of a gammatone filter centered at frequency f is

$$g(f,t) = \begin{cases} b^a t^{a-1} e^{-2\pi bt} \cos(2\pi ft), & t \geqslant 0 \\ 0 & \text{otherwise}, \end{cases} \quad (1)$$

where $a=4$ is the order of the filter and b is the equivalent rectangular bandwidth (Glasberg and Moore, 1990), which increases as the center frequency f increases.

Let $x(t)$ be the input signal. The response from a filter channel c, $x(c,t)$, is given by

$$x(c,t) = x(t) * g(f_c,t), \quad (2)$$

where "*" denotes convolution, and f_c the center frequency of filter channel c. In each filter channel, the output is further divided into 20-ms time frames with a 10-ms shift between consecutive frames.

B. Feature extraction

Previous studies suggest that in a T-F region dominated by a periodic signal, T-F units in adjacent channels tend to have highly correlated filter responses (Wang and Brown, 1999) or response envelopes (Hu and Wang, 2004). In this stage, we calculate such cross-channel correlations. These correlations will be used to determine T-F units dominated by unvoiced speech in the grouping stage.

Cross-channel correlation of filter responses measures the similarity between the responses of two adjacent filter channels. Since these responses have channel-dependent phases, we perform phase alignment before measuring their correlation. Specifically, we first compute their autocorrelation functions (Licklider, 1951; Lyon, 1984; Slaney and Lyons, 1990) and then use their autocorrelation responses to calculate cross-channel correlation.

Let u_{cm} denote a T-F unit for frequency channel c and time frame m, the corresponding autocorrelation of the filter response is given by

$$A(c,m,\tau) = \sum_n x(c,mT_m - nT_n)x(c,mT_m - nT_n - \tau T_n). \quad (3)$$

Here, τ is the delay and n denotes discrete time. $T_m=10$ ms is the frame shift and T_n is the sampling time. The above summation is over 20 ms, the length of a time frame. The cross-channel correlation between u_{cm} and $u_{c+1,m}$ is given by

$$C(c,m) = \frac{\Sigma_\tau [A(c,m,\tau) - \overline{A(c,m)}][A(c+1,m,\tau) - \overline{A(c+1,m)}]}{\sqrt{\Sigma_\tau [A(c,m,\tau) - \overline{A(c,m)}]^2 \Sigma_\tau [A(c+1,m,\tau) - \overline{A(c+1,m)}]^2}}, \quad (4)$$

where $\bar{A}$ denotes the average value of A.

When the input contains a periodic signal, auditory filters with high center frequencies respond to multiple harmonics. Such a filter response is amplitude modulated, and the response envelope fluctuates at the F0 of the periodic signal (Helmholtz, 1863). As a result, adjacent channels in the high-frequency range tend to have highly correlated response envelopes. To extract these correlations, we calculate response envelope through half-wave rectification and bandpass filtering, where the passband corresponds to the plausible F0 range of target speech, i.e., [70 Hz, 400 Hz], the typical pitch range for adults (Nooteboom, 1997). The resulting bandpassed envelope in channel c is denoted by $x_E(c,t)$.

Similar to Eqs. (3) and (4), we compute envelope autocorrelation as

$$A_E(c,m,\tau) = \sum_n x_E(c,mT_m - nT_n)x_E(c,mT_m - nT_n - \tau T_n) \quad (5)$$

and then obtain cross-channel correlation of response envelopes as

$$C_E(c,m) = \frac{\Sigma_\tau [A_E(c,m,\tau) - \overline{A_E(c,m)}][A_E(c+1,m,\tau) - \overline{A_E(c+1,m)}]}{\sqrt{\Sigma_\tau [A_E(c,m,\tau) - \overline{A_E(c,m)}]^2 \Sigma_\tau [A_E(c+1,m,\tau) - \overline{A_E(c+1,m)}]^2}}. \quad (6)$$

C. Auditory segmentation

Previous CASA systems perform auditory segmentation by analyzing common periodicity (Brown and Cooke, 1994; Wang and Brown, 1999; Hu and Wang, 2004), and thus cannot handle unvoiced speech. In this study, we apply a segmentation algorithm based on a multiscale analysis of event onsets and offsets (Hu and Wang, 2007). Onsets and offsets are important ASA cues (Bregman, 1990) because different sound sources in an acoustic environment seldom start and end at the same time. In the time domain, boundaries between different sound sources tend to produce onsets and offsets. Common onsets and offsets also provide natural cues

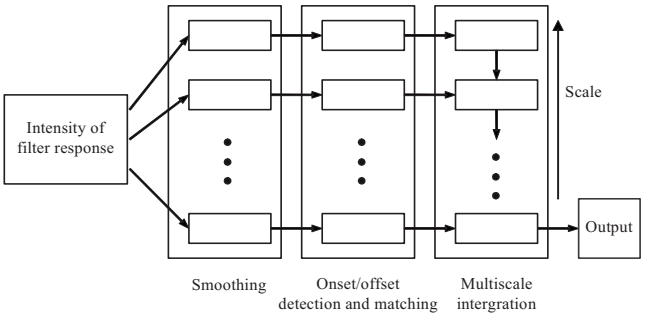


FIG. 2. Diagram of the segmentation stage. In each processing step, a rectangle represents a particular scale, which increases from bottom to top.

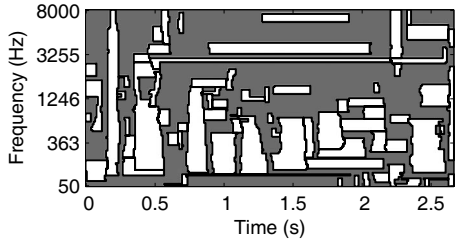


FIG. 3. Bounding contours of estimated segments. The input is the mixture shown in Fig. 1(d). The background is represented by gray.

to integrate sounds from the same source across frequency. Because onset and offset are cues common to all the sounds, this algorithm is applicable to both voiced and unvoiced speech. Figure 2 shows the diagram of the segmentation stage. It has three steps: Smoothing, onset/offset detection, and multiscale integration.

Onsets and offsets correspond to sudden intensity increases and decreases, respectively. A standard way to identify such intensity changes is to find the peaks and valleys of the time derivative of signal intensity (Wang and Brown, 2006, Chap. 3). We calculate the intensity of a filter response as the square of the response envelope, which is extracted using half-wave rectification and low-pass filtering. Because of the intensity fluctuation within individual events, many peaks and valleys of the derivative do not correspond to real onsets and offsets. Therefore, in the first step of segmentation, we smooth the intensity over time to reduce such fluctuations. Since an acoustic event tends to have synchronized onset and offset across frequency, we additionally perform smoothing over frequency, which helps to enhance such coincidences in neighboring frequency channels. This procedure is similar to the standard Canny edge detector in image processing (Canny, 1986). The degree of smoothing over time and frequency is referred to as the two-dimensional scale. The larger the scale is, the smoother the intensity is. The smoothed intensities at different scales form the so-called scale space (Romeny *et al.*, 1997).

In the second step of segmentation, our system detects onsets and offsets in each filter channel. Onset and offset candidates are detected by marking peaks and valleys of the time derivative of the smoothed intensity. The system then merges simultaneous onsets and offsets in adjacent channels into onset and offset fronts, which are contours connecting onset and offset candidates across frequency. Segments are obtained by matching individual onset and offset fronts.

As a result of smoothing, event onsets and offsets of small T-F regions may be blurred at a larger (coarser) scale. Consequently, we may miss some true onsets and offsets. On the other hand, at a smaller (finer) scale, the detection may be sensitive to insignificant intensity fluctuations within individual events. Consequently, false onsets and offsets may be generated and some true segments may be oversegmented. We find it generally difficult to obtain satisfactory segmentation with a single scale. In the last step of segmentation, we deal with this issue by performing multiscale integration from the largest scale to the smallest scale in an orderly manner. More specifically, at each scale, our system first locates more accurate boundaries for the segments obtained at a larger scale. Then it creates new segments outside the existing ones. The details of the segmentation stage are given in Hu and Wang (2007); see also Hu (2006).

As an illustration, Fig. 3 shows the bounding contours of obtained segments for the mixture in Fig. 1(d). The background is represented in gray. Compared with the ideal binary mask in Fig. 1(e), the obtained segments capture a majority of the target speech. Some segments for the interference are also formed. Note that the system does not, in this stage, distinguish between target and interference for each segment, which is the task of grouping described below.

IV. GROUPING

Our general strategy for grouping is to first segregate voiced speech and then deal with unvoiced speech. This strategy is motivated by the consideration that voiced speech segregation has been well studied and can be applied separately, and segregated voiced speech can be useful in subsequent unvoiced speech segregation.

To segregate the voiced portions of a target utterance, we apply our previous system for voiced speech segregation (Hu and Wang, 2006), which is slightly extended from an earlier version (Hu and Wang, 2004) and produces good segregation results. Target pitch contours needed for segregation are obtained from a clean target by PRAAT, a standard pitch determination algorithm for clean speech (Boersma and Weenink, 2004). This way, we avoid pitch tracking errors which could adversely influence the performance of unvoiced speech segregation—the focus of this study. We refer to the resulting stream of voiced target as S_T^1.

The task of grouping unvoiced target amounts to labeling segments already obtained in the segmentation stage. A segment may be dominated by voiced target, unvoiced target, or interference, and we want to group segments dominated by unvoiced target while rejecting segments dominated by interference. Since an unvoiced phoneme is often strongly coarticulated with a neighboring voiced phoneme, some unvoiced target is included in segments dominated by voiced target (Hu, 2006; Hu and Wang, 2007). So we need to group segments dominated by voiced target to recover this part of unvoiced speech.

Our system first groups segments dominated by voiced target. Then among the remaining segments, we label those dominated by unvoiced target in two steps: Segment removal and segment classification.

A. Grouping segments dominated by voiced target

A segment dominated by voiced target should have a significant overlap with the segregated voiced target, S_T^1. Hence we label a segment as dominated by voiced target if

(1) more than half of its total energy is included in the voiced time frames of target, and
(2) more than half of its energy in the voiced frames is included in the T-F units belonging to S_T^1.

All the segments labeled as dominated by voiced target are grouped into the segregated target stream.

By grouping segments dominated by voiced target, we recover more target-dominant T-F units than S_T^1. However, some interference-dominant T-F units are also included due to the mismatch error in segmentation, i.e., the error of putting both target- and interference-dominant units into one segment (Hu and Wang, 2007). We found that a significant amount of the mismatch error in segmentation stems from merging T-F areas in adjacent channels into one segment (Hu, 2006). To minimize the amount of interference-dominant T-F units being wrongly grouped into the target stream, we consider estimated segments in individual channels, referred to as T-segments, instead of whole T-F segments. Specifically, if a T-segment is dominated by a voiced target based on the above two criteria, all the T-F units within the T-segment are grouped into the voiced target. The resulting stream is referred to as S_T^2.

B. Acoustic-phonetic features for segment classification

The next task is to label or classify segments dominated by unvoiced speech. Since the signal within a segment is mainly from one source, it is expected to have similar acoustic-phonetic properties to that source. Therefore, we identify segments dominated by unvoiced speech using acoustic-phonetic features.

A basic speech sound is characterized by the following acoustic-phonetic properties: Short-term spectrum, formant transition, voicing, and phoneme duration (Stevens, 1998; Ladefoged, 2001). These features have proven to be useful in speech recognition, e.g., to distinguish different phonemes or words (Rabiner and Juang, 1993; Ali and Van der Spiegel, 2001b, 2001a). These properties may also be useful in distinguishing speech from nonspeech interference. However, it is important to treat these properties, appropriately considering that we are dealing with noisy speech. In particular, we give the following considerations.

(1) *Spectrum*. The short-term spectrum of an acoustic mixture at a particular time may be quite different from that of the target utterance or that of the interference in the mixture. Therefore, features representing the overall shape of a short-term spectrum may not be appropriate for our task. On the other hand, the short-term spectra in the T-F regions dominated by speech are expected to be similar to those of clean utterances, while the short-term spectra of other T-F regions tend to be different. Therefore, we use the short-term spectrum within a T-F region as a feature to decide whether this region is dominated by speech or interference. More specifically, we use the energy within individual T-F units as the feature to represent the short-term spectrum.
(2) *Formant transition*. It is difficult to estimate the formant frequency of a target utterance in the presence of strong interference. In addition, formant transition is embodied in the corresponding short-term spectrum. Therefore, we do not explicitly use formant transition in this study.
(3) *Voicing*. Voicing information of a target utterance is not utilized since we are handling unvoiced speech.
(4) *Duration*. While the duration of an interfering sound is unpredictable, for speech each phoneme lasts for a range of durations. However, we may not be able to detect the boundaries of phonemes that are strongly coarticulated. Therefore it is difficult to find the accurate durations of individual phonemes from an acoustic mixture, and the durations of individual phonemes are not utilized in this study.

In summary, we use the signal energy within individual T-F segments to derive the acoustic-phonetic features for distinguishing speech and nonspeech interference.

C. Segment removal

Since our task is to group segments for unvoiced speech, segments that mainly contain periodic or quasiperiodic signals unlikely originate from unvoiced speech and should be removed. A segment is removed if more than half of its total energy is included in the T-F units dominated by a periodic signal. We consider unit u_{cm} dominated by a periodic signal if it is included in the segregated voiced stream or has a high cross-channel correlation, the latter indicating that two neighboring channels respond to the same harmonic or formant (Wang and Brown, 1999). Specifically, a cross-channel correlation is considered high if $C(c,m)>0.985$ or $C_E(c,m)>0.985$.

Among the remaining segments, a segment dominated by unvoiced target is unlikely located at time frames corresponding to voiced phonemes other than expanded obstruents. This property is, however, not shared by some interference-dominant segments that can have significant energy in such voiced frames. We remove these segments as follows.

We first label the voiced frames of a target utterance that unlikely contain an expanded obstruent, according to the segregated voiced target. Let $H_0(m_1,m_2)$ be the hypothesis that a T-F region between frame m_1 and frame m_2 is dominated by speech and $H_1(m_1,m_2)$ the hypothesis that the region is dominated by interference. In addition, let $H_{0,a}(m_1,m_2)$ be the hypothesis that this region is dominated by an expanded obstruent and $H_{0,b}(m_1,m_2)$ by any other phoneme.

Let $X(c,m)$ be the energy in u_{cm} and $X(m)=\{X(c,m),\forall c\}$ the vector of the energy in all the T-F units at

time frame m. $X(m)$ is referred to as the *cochleagram* at frame m (Wang and Brown, 2006). Let $X_T(m) = \{X_T(c,m), \forall c\}$ be the cochleagram of the segregated target at frame m, that is,

$$X_T(c,m) = \begin{cases} X(c,m) & \text{if } u_{cm} \in S_T^2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

A voiced frame m is labeled as obstruent dominant if

$$P(H_{0,a}(m)|X_T(m)) > P(H_{0,b}(m)|X_T(m)). \tag{8}$$

We assume that, given $X_T(m)$, these posterior probabilities do not depend on a particular frame index. In other words, for any two frames m_1 and m_2,

$$P(H(m_1)|X_T(m_1)) = P(H(m_2)|X_T(m_2)) \quad \text{if } X_T(c,m_1) = X_T(c,m_2), \forall c. \tag{9}$$

To simplify calculations, we further assume that the prior probabilities of $H_{0,a}(m)$, $H_{0,b}(m)$, and $H_1(m)$ are constant for individual frames within a given T-F region. A frame index can then be dropped from these frame-level hypotheses. In the following, we use a hypothesis without a frame index to refer to that hypothesis for a single frame of a T-F segment. Then Eq. (8) becomes

$$P(H_{0,a}|X_T(m)) > P(H_{0,b}|X_T(m)). \tag{10}$$

Given that $X_T(m)$ corresponds to the voiced target, we have $P(H_{0,b}|X_T(m)) = 1 - P(H_{0,a}|X_T(m))$. Therefore, we have

$$P(H_{0,a}|X_T(m)) > 0.5. \tag{11}$$

We construct a multilayer perceptron (MLP) to compute $P(H_{0,a}|X_T(m))$. The MLP uses sigmoidal activation functions and has one hidden player. The input to the MLP is $X_T(m)$, the cochleagram of a target utterance at a voiced frame. Consequently, this MLP has 128 units in the input layer. It has one unit in the output layer. The desired output of this unit is 1 if the corresponding frame is dominated by an expanded obstruent and 0 otherwise. Note that when there are sufficient training samples, the trained MLP yields a good estimate of the probability (Bridle, 1989). The MLP is trained with a corpus that includes all the utterances from the training part of the TIMIT database and 100 intrusions. These intrusions include crowd noise and environmental sounds, such as wind, bird chirp, and ambulance alarm.[1] Utterances and intrusions are mixed at 0-dB SNR to generate training samples. We use PRAAT to label voiced frames. The cochleagram of the target at voiced frames is determined using the ideal binary mask of each mixture. The number of units in the hidden layer of the MLP is determined using cross validation. Specifically, we divide the training samples into two equal sets, one for training and the other for validation. The resulting MLP has 20 units in the hidden layer.

We label every voiced frame based on Eq. (11). A segment is removed if more than 50% of its energy is included in the voiced frames that are not dominated by an expanded obstruent. As a result of segment removal, many segments dominated by interference are removed. We find that this step increases the robustness of the system and greatly reduces the computational burden for the following segment classification.

D. Segment classification

In this step, we classify the remaining segments as dominated by either unvoiced speech or interference. Let s be a remaining segment lasting from frame m_1 to m_2, and $X_s(m) = \{X_s(c,m), \forall c\}$ be the corresponding cochleagram at frame m. That is,

$$X_s(c,m) = \begin{cases} X(c,m) & \text{if } u_{cm} \in s \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Let $\mathbf{X}_s = [X_s(m_1), X_s(m_1+1), \ldots, X_s(m_2)]$. s is classified as dominated by unvoiced speech if

$$P(H_{0,a}(m_1,m_2)|\mathbf{X}_s) > P(H_1(m_1,m_2)|\mathbf{X}_s). \tag{13}$$

Because segments have varied durations, directly evaluating $P(H_{0,a}(m_1,m_2)|\mathbf{X}_s)$ and $P(H_1(m_1,m_2)|\mathbf{X}_s)$ for each possible duration is not computationally feasible. Therefore, we consider a simplifying approximation that each time frame is statistically independent (more discussion on this approximation will be given later in this section). Since

$$P(H_{0,a}(m_1,m_2)|\mathbf{X}_s) = P(H_{0,a}(m_1), H_{0,a}(m_1+1), \ldots, H_{0,a}(m_2)|\mathbf{X}_s) \tag{14}$$

Applying the chain rule,

$$\begin{aligned} P(H_{0,a}(m_1,m_2)|\mathbf{X}_s) &= P(H_{0,a}(m_1)|\mathbf{X}_s) \\ &\times P(H_{0,a}(m_1+1)|H_{0,a}(m_1),\mathbf{X}_s)\cdots \\ &\times P(H_{0,a}(m_2)|H_{0,a}(m_1), H_{0,a}(m_1+1), \ldots, H_{0,a}(m_2 - 1),\mathbf{X}_s). \end{aligned} \tag{15}$$

From the independence assumption, we have

$$\begin{aligned} &P(H_{0,a}(m_1+k)|H_{0,a}(m_1), H_{0,a}(m_1+1), \\ &\quad \ldots, H_{0,a}(m_2+k-1), \mathbf{X}_s) \\ &= P(H_{0,a}(m_1+k)|\mathbf{X}_s) = P(H_{0,a}(m_1+k)|X_s(m_1+k)). \end{aligned} \tag{16}$$

Therefore,

$$P(H_{0,a}(m_1,m_2)|\mathbf{X}_s) = \prod_{m=m_1}^{m_2} P(H_{0,a}(m)|X_s(m)), \tag{17}$$

and the same calculation can be done for $P(H_1(m_1,m_2)|\mathbf{X}_s)$. Now Eq. (13) becomes

$$\prod_{m=m_1}^{m_2} P(H_{0,a}(m)|X_s(m)) > \prod_{m=m_1}^{m_2} P(H_1(m)|X_s(m)). \tag{18}$$

By applying the Bayesian rule and the assumption made in Sec. IV C that the prior and the posterior probabilities do not depend on a frame index within a given segment, the above inequality becomes

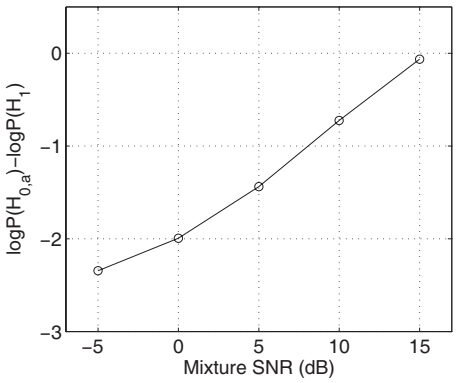


FIG. 4. Ratio of the prior probability of the target to that of interference as a function of mixture SNR.

$$\left[\frac{P(H_{0,a})}{P(H_1)}\right]^{m_2-m_1+1}\prod_{m=m_1}^{m_2}\frac{p(X_s(m)|H_{0,a})}{p(X_s(m)|H_1)}>1. \tag{19}$$

The prior probabilities $P(H_{0,a})$ and $P(H_1)$ depend on the SNR of acoustic mixtures. Figure 4 shows the observed logarithmic ratios between $P(H_{0,a})$ and $P(H_1)$ from the training data at different mixture SNR levels. We approximate the relationship shown in the figure by a linear function,

$$\log\frac{P(H_{0,a})}{P(H_1)}=0.1166\ \text{SNR}-1.8962. \tag{20}$$

If we can estimate the mixture SNR, we will be able to estimate the log ratio of $P(H_{0,a})$ and $P(H_1)$ and use it in Eq. (19). This allows us to be more stringent in labeling a segment as speech dominant when the mixture SNR is low.

We propose to estimate the SNR of an acoustic mixture by capitalizing on the voiced target that has already been segregated from the mixture. Let E_1 be the total energy included in the T-F units labeled 1 at the voiced frames of the target. One may use E_1 to approximate the target energy at voiced frames and estimate the total target energy as αE_1 that includes unvoiced target speech. By analyzing the training part of the TIMIT database, we find that parameter α—the ratio between the total energy of a speech utterance and the total energy at the voiced frames of the utterance—varies substantially across individual utterances. In this study, we set α to 1.09, the average value of all the utterances in the training part of the TIMIT database. Similarly, let E_2 be the total energy included in the T-F units labeled 0 at the voiced frames of the target, N_1 the total number of these voiced frames, and N_2 the total number of other frames. Hence, E_2 approximates the interference energy at voiced frames, and the average interference energy per voiced frame is then E_2/N_1. Assuming that interference is relatively steady, we can use E_2/N_1 to approximate the interference energy per frame and estimate the total interference energy as $E_2(N_1+N_2)/N_1$. Consequently, the estimated mixture SNR is

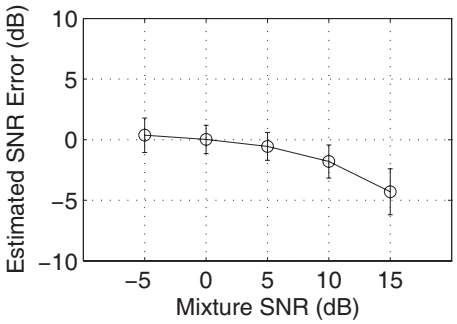


FIG. 5. Mean and standard deviation of estimated mixture SNRs in the test corpus.

$$\overline{\text{SNR}}=10\log_{10}\frac{\alpha N_1E_1}{(N_1+N_2)E_2}=10\log_{10}\frac{E_1}{E_2}+10\log_{10}\alpha+10\log_{10}\frac{N_1}{N_1+N_2}. \tag{21}$$

With $\alpha=1.09$, $10\log_{10}\alpha=0.37$ dB. We have applied this SNR estimation to the test corpus. Figure 5 shows the mean and the standard deviation of the estimation error at each SNR level of the original mixtures; the estimation error equals the estimated SNR subtracted by the true SNR. As shown in the figure, the system yields a reasonable estimate when the mixture SNR is lower than 10 dB. When the mixture SNR is greater than or equal to 10 dB, Eq. (21) tends to underestimate the true SNR. As discussed in Sec. II, some voiced frames of the target, such as those corresponding to expanded obstruents, may contain unvoiced target energy that fails to be included in E_1 but ends up in E_2. When the mixture SNR is low, this part of unvoiced energy is much lower than the interference energy. Therefore, it is negligible and Eq. (21) provides a good estimate. When the mixture SNR is high, this unvoiced target energy can be comparable to interference energy, and as a result the estimated SNR tends to be systematically lower than the true SNR.

Alternatively, one can also estimate the mixture SNR at the unvoiced frames of the target or estimate the target energy at the unvoiced frames based on the average frame-level energy ratio of unvoiced speech to voiced speech. These alternatives have been evaluated in Hu (2006), and they do not yield more accurate estimates. Of course, for the TIMIT corpus we can simply correct the systematic bias shown in Fig. 5. We choose not to do so for the sake of generality.

To label a segment as either expanded obstruent or interference according to Eq. (19), we need to estimate the likelihood ratio between $p(X_s(m)|H_{0,a})$ and $p(X_s(m)|H_1)$. When $P(H_{0,a})$ and $P(H_1)$ are equal, we have by the Bayesian rule

$$\frac{p(X_s(m)|H_{0,a})}{p(X_s(m)|H_1)}=\frac{P(H_{0,a}|X_s(m))}{P(H_1|X_s(m))}. \tag{22}$$

We train an MLP to estimate $P(H_{0,a}|X_s(m))$ when $P(H_{0,a})$ and $P(H_1)$ are equal. The MLP has the same structure as the one described in Sec. IV C. The desired output of this MLP is 1 if a frame of a segment is dominated by an expanded obstruent and 0 if it is dominated by nonspeech interference.

The MLP is trained with the cochleagrams of target utterances at time frames corresponding to expanded obstruents and those of nonspeech intrusions from the same training set described in Sec. IV C. Since $P(H_1|X_s(m))=1-P(H_{0,a}|X_s(m))$, given that frame m corresponds to an expanded obstruent, we are able to calculate the likelihood ratio of $p(X_s(m)|H_{0,a})$ and $p(X_s(m)|H_1)$ using the output from the trained MLP.

Using the above estimate of the likelihood ratio and the estimated mixture SNR to calculate the prior probability ratio of $P(H_{0,a})$ and $P(H_1)$, we label a segment as either expanded obstruent or interference according to Eq. (19). All the segments labeled as unvoiced speech are added to the segregated voiced stream, S_T^2, yielding the final segregated stream, referred to as S_T^3.

This method for segregating unvoiced speech is very similar to a previous version (Wang and Hu, 2006) where we used fixed prior probabilities for all SNR levels. We find that using SNR-dependent prior probabilities gives better performance, especially when the mixture SNR is high. In an earlier study (Hu and Wang, 2005), we used Gaussian mixture models (GMM) to model both speech and interference and then classify a segment using the obtained models. The performance in that study is not as good as the present method. The main reason, we believe, is that although GMM is trained to represent the distributions of speech and interference accurately, MLP is trained to distinguish speech and interference and therefore has more discriminative power. We have also considered the dependence between consecutive frames, instead of treating individual frames as independent. The obtained result is comparable to that obtained with the independence assumption, probably due to the fact that the signal within a segment is usually quite stable across time so that considering the dynamics within a segment does not provide much additional information for classification.

As an example, Figs. 6(e) and 6(f) show the final segregated target and the corresponding synthesized waveform for the mixture in Fig. 1(d). Compared with the ideal mask in Fig. 1(e) and the corresponding synthesized wave form in Fig. 1(f), our system segregates most of target energy and rejects most of interfering energy. In addition, Figs. 6(a) and 6(b) show the mask and the waveform of the segregated voiced target, i.e., S_T^1. Figures 6(c) and 6(d) show the mask and the waveform of the resulting stream after grouping T-segments dominated by voiced speech, i.e., S_T^2. The target utterance, "That noise problem grows more annoying each day," includes five stops (/t/ in "that," /p/ and /b/ in "problem," /g/ in "grows," and /d/ in "day"), three fricatives (/ð/ in "that," /z/ in "noise," and /z/ in "grows"), and one affricate (/tʃ/ in "each"). The unvoiced parts of some consonants with strong coarticulation with the voiced speech, such as /ð/ in "that" and /d/ in "day," are segregated by using T-segments. The unvoiced part of /z/ in "noise" and /tʃ/ in "each" are segregated by grouping the corresponding segments. Except for a significant loss of energy for /p/ in "problem" and some energy loss for /t/ in "that," our system segregates most of the energy of the above consonants.

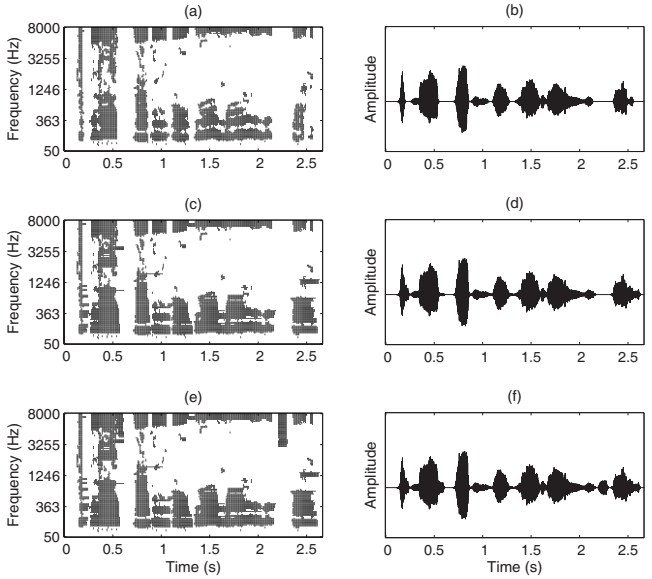


FIG. 6. Segregated target of the mixture shown in Fig. 1(d). (a) Mask of segregated voiced target (black regions). (b) Waveform resynthesized from the mask in (a). (c) Mask of the resulting target stream after grouping estimated T-segments (black regions). (d) Waveform resynthesized from the mask in (c). (e) Mask of the final segregated target (black regions). (f) Waveform resynthesized from the mask in (e).

V. EVALUATION

We now systematically evaluate the performance of our system. Here, we use a test corpus containing 20 target utterances randomly selected from the test part of the TIMIT database mixed with 15 nonspeech intrusions including five with crowd noise. Table III lists the 20 target utterances. The intrusions are as follows: N1—white noise, N2—rock music,

TABLE III. Target utterances in the test corpus.

Target	Content
S1	Put the butcher block table in the garage.
S2	Alice's ability to work without supervision is noteworthy.
S3	Barb burned paper and leaves in a big bonfire.
S4	Swing your arm as high as you can.
S5	Shaving cream is a popular item on Halloween.
S6	He then offered his own estimate of the weather, which was unenthusiastic.
S7	The morning dew on the spider web glistened in the sun.
S8	Her right hand aches whenever the barometric pressure changes.
S9	Why yell or worry over silly items.
S10	Aluminum silverware can often be flimsy.
S11	Guess the question from the answer.
S12	Medieval society was based on hierarchies.
S13	That noise problem grows more annoying each day.
S14	Don't ask me to carry an oily rag like that.
S15	Each untimely income loss coincided with the breakdown of a heating system part.
S16	Combine all the ingredients in a large bowl.
S17	Fuss, fuss, old man.
S18	Don't ask me to carry an oily rag like that.
S19	The fish began to leap frantically on the surface of the small lake.
S20	The redcoats ran like rabbits.

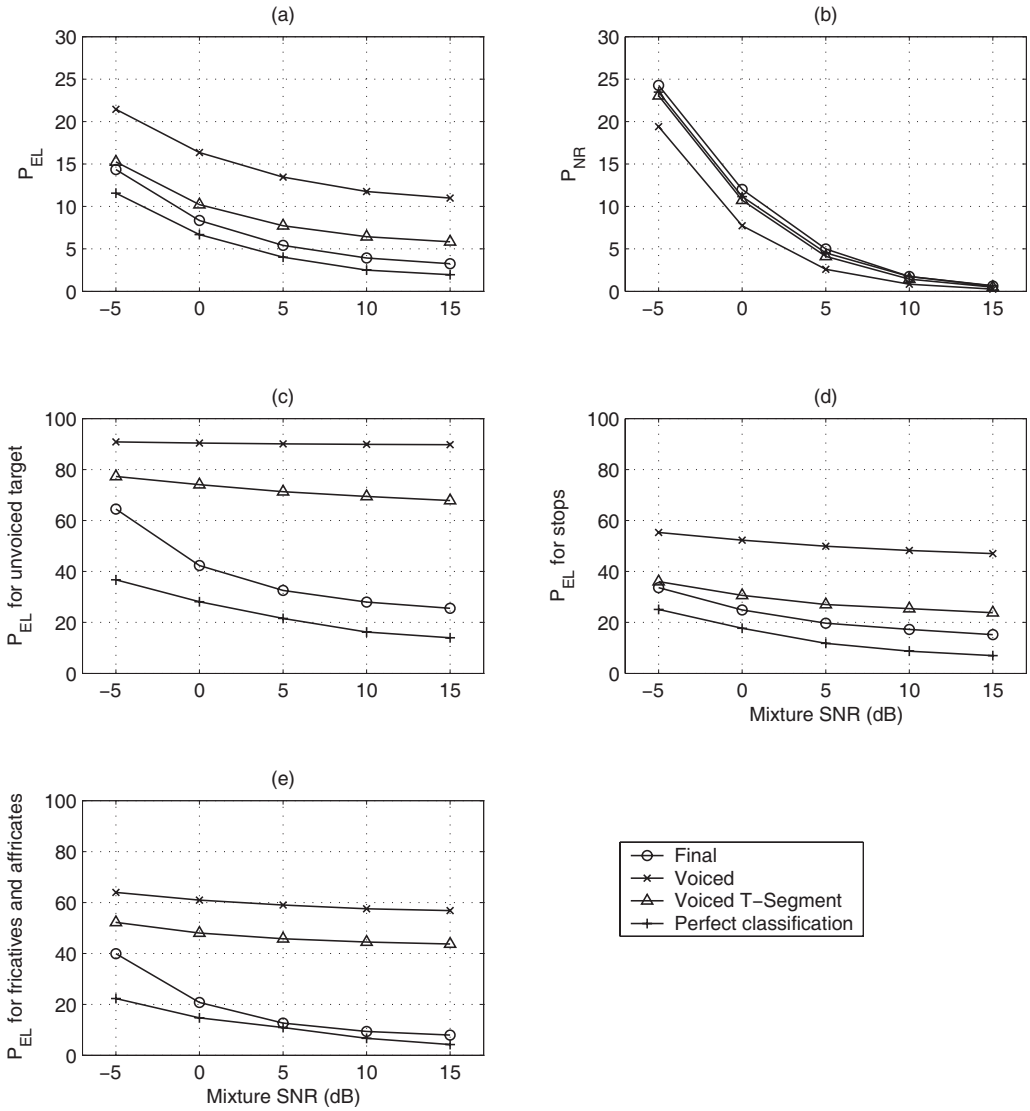


FIG. 7. System performance. In this figure, "Final" refers to the final segregated target, "Voiced" the segregated voiced target, "Voiced T-segment" the segregated target after grouping T-segments dominated by voiced target, and "Perfect classification" segregated target with perfect segment classification. (a) Average percentage of energy loss. (b) Average percentage of noise residue. (c) Average percentage of energy loss for unvoiced speech. (d) Average percentage of energy loss for stop consonants. (e) Average percentage of energy loss for fricatives and affricates.

N3—siren, N4—telephone ring, N5—electric fan, N6—clock alarm, N7—traffic noise, N8—bird chirp with water flowing, N9—wind, and N10—rain, N11—cocktail party noise, N12—crowd noise at a playground, N13—crowd noise with music, N14—crowd noise with clap, and N15—babble noise (16 speakers). This set of intrusions is not used during training, and represents a broad range of nonspeech sounds encountered in typical acoustic environments. Each target utterance is mixed with individual intrusions at −5-, 0-, 5-, 10-, and 15-dB SNR levels. The test corpus has 300 mixtures at each SNR level and 1500 mixtures altogether.

We evaluate our system by comparing the segregated target with the ideal binary mask—the stated computational goal. The performance of segregation is given by comparing the estimated mask and the ideal binary mask with two measures (Hu and Wang, 2004):

(1) the percentage of energy loss, P_{EL}, which measures the amount of energy in the target-dominant T-F units that are labeled as interference (hence removed) relative to the total energy in target-dominant T-F units and
(2) the percentage of noise residue, P_{NR}, which measures the amount of energy in the interference-dominant T-F units that are labeled as target (hence retained) relative to the total energy in T-F units estimated as target dominant.

P_{EL} and P_{NR} provide complementary error measures of a segregation system, and a successful system needs to achieve low errors in both measures.

The P_{EL} and P_{NR} values for S_T^3 at different input SNR levels are shown in Figs. 7(a) and 7(b). Each value in the figure is the average over the 300 mixtures of individual targets and intrusions N1–N15. As shown in the figure, for the final segregation, our system captures an average of 85.7% of target energy at −5-dB SNR. This value increases to 96.7% when the mixture SNR increases to 15 dB. On average 24.3% of the segregated target belongs to interfer-

ence at −5 dB. This value decreases to 0.6% when the mixture SNR increases to 15 dB. In summary, our system captures a majority of target without including much interference.

To see the performance of our system on unvoiced speech in detail, we measure P_{EL} for target speech in the unvoiced frames. The average of these P_{EL} values at different SNR levels are shown in Fig. 7(c). Note that since some voiced frames contain unvoiced target, these are not exactly the P_{EL} values of unvoiced speech. Nevertheless, they are close to the real values. As shown in the figure, our system captures 35.5% of the target energy at the unvoiced frames when the mixture SNR is −5 dB and 74.4% when the mixture SNR is 15 dB. Overall, our system is able to capture more than 50% of target energy at the unvoiced frames when the mixture SNR is 0 dB or higher.

As discussed in Sec. II, expanded obstruents often contain voiced and unvoiced signals at the same time. Therefore, we measure P_{EL} for these phonemes separately in order to gain more insight into system performance. Because affricates do not occur often and they are similar to fricatives, we measure P_{EL} for fricatives and affricates together. The averages of these P_{EL} values at different SNR levels are shown in Figs. 7(d) and 7(e). As shown in the figure, our system performs somewhat better for fricatives and affricates when the mixture SNR is 0 dB or higher. On average, the system captures about 65% of these phonemes when the mixture SNR is −5 dB and about 90% when the mixture SNR is 15 dB.

For comparison, Fig. 7 also shows the P_{EL} and P_{NR} values for segregated voiced target, i.e., S_T^1 (labeled as "Voiced"), and the resulting stream after grouping T-segments dominated by voiced target, S_T^2 (labeled as "Voiced T-segments"). As shown in the figure, S_T^1 only includes about 10% of target energy in unvoiced frames, while S_T^2 includes about 17% more on average [Fig. 7(c)]. This additional 17% mainly corresponds to unvoiced phonemes that have strong coarticulation with neighboring voiced phonemes. By comparing these P_{EL} and P_{NR} values with those of the final segregated target, we can see that grouping segments dominated by unvoiced speech helps to recover a large amount of unvoiced speech. It also includes a small amount of additional interference energy, especially when the mixture SNR is low [Fig. 7(b)].

In addition, Fig. 7 shows the P_{EL} and P_{NR} values for segregated target obtained with perfect segment classification. As shown in the figure, there is a performance gap that can be narrowed with better classification, especially when the mixture SNR is low.

We also measure the system performance in terms of SNR by treating the target synthesized from the corresponding ideal binary mask as signal (Hu and Wang, 2004, 2006). Figures 8(a) and 8(b) show the overall average SNR values of segregated targets at different levels of mixture SNR and the corresponding SNR gain. Figures 8(c) and 8(d) show the corresponding values at unvoiced frames. Our system improves SNR in all input conditions.

To put our performance in perspective, we have compared with spectral subtraction, a standard method for speech enhancement (Huang *et al.*, 2001), with the above SNR measures. The spectral subtraction method is applied as follows. For each acoustic mixture, we assume that the silent portions of a target utterance are known and use the short-term spectra of interference in these portions as the estimates of interference. Interference is attenuated by subtracting the most recent interference estimate from the mixture spectrum at every time frame. The resulting SNR measures of the spectral subtraction method are also shown in Fig. 8. As shown in the figure, our system performs substantially better for both voiced and unvoiced speech than the spectral subtraction method even when it is applied with perfect speech pause detection; the only exception occurs for unvoiced speech at the input SNR of 15 dB. The improvement is more pronounced when the mixture SNR is low. At 15-dB SNR, the error in SNR estimation becomes significant (see Fig. 5), and the unvoiced speech energy that fails to be grouped becomes relatively large in comparison with interference energy. The spectral subtraction method is based on the estimation of interference and is less sensitive to input SNR.

We should point out that our system has significantly higher computational complexity than the spectral subtraction method. Note, however, that the spectral subtraction method implemented in our comparison assumes the prior knowledge of silent intervals of target speech, which greatly simplifies noise estimation—a nontrivial task that can itself be computationally intense. The major computational load of our system stems from calculating autocorrelations in the stage of feature extraction (Sec. III B) and extracting response envelopes in the stages of feature extraction and segmentation (Sec. III C). Also, our system needs to perform these computations in 128 frequency channels. As the calculations of the autocorrelations and response envelopes can be carried out in different channels independently, one can substantially improve computational efficiency by utilizing parallel computing hardware.

VI. DISCUSSION

Several insights have emerged from this study. The first is that the temporal properties of acoustic signals are crucial for speech segregation. Our system makes extensive use of temporal properties. In particular, we group target sound in consecutive frames based on the temporal continuity of speech signal. Furthermore, our system generates segments by analyzing sound intensity across time, i.e., onset and offset detection. The importance of temporal properties of speech for human speech recognition has been convincingly demonstrated by Shannon *et al.* (1995). In addition, studies in ASR suggest that long-term temporal information helps to improve recognition rate (see, e.g., Hermansky and Sharma, 1999). All these observations show that temporal information plays a critical role in sound organization and recognition.

Second, we find it advantageous to segregate voiced speech first and then use the segregated voiced speech to aid the segregation of unvoiced speech. As discussed before, unvoiced speech is more vulnerable to interference and more difficult to segregate. Segregation of voiced speech is more reliable and can be used to assist in the segregation of unvoiced speech. Our study shows that the unvoiced speech

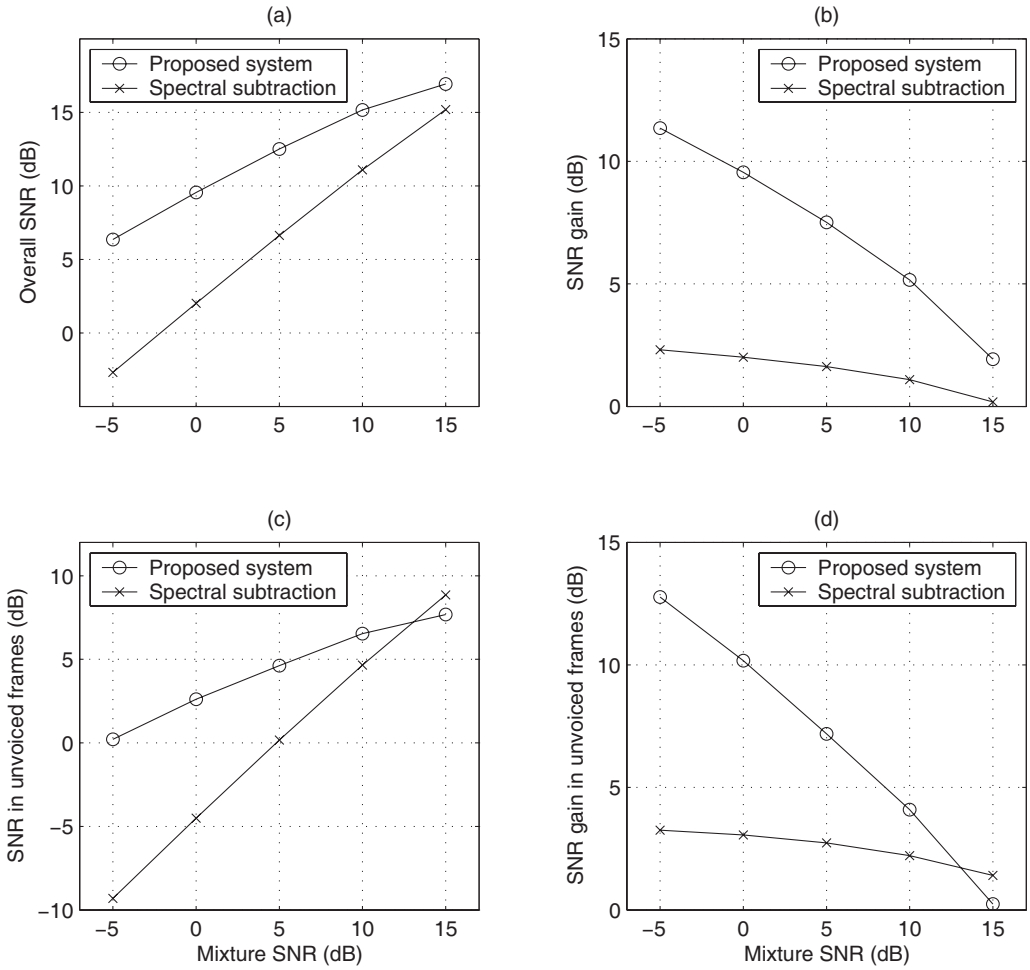


FIG. 8. SNR performances of the proposed system and spectral subtraction. (a) SNR of segregated target. (b) SNR gain of segregated target. (c) SNR of segregated target at unvoiced frames. (d) SNR gain of segregated target at unvoiced frames.

with strong coarticulation with voiced speech can be segregated using segregated voiced speech and estimated T-segments. Segregated voiced speech is also used to delineate the possible T-F locations of unvoiced speech. As a result, our system need not search the entire T-F domain for segments dominated by unvoiced speech and less likely identifies an interference-dominant segment as target. In addition, we have proposed an estimate of the mixture SNR from segregated voiced speech, which helps the system to adapt the prior probabilities in segment classification.

In addition, auditory segmentation is important for unvoiced speech segregation. In our system, the segmentation stage provides T-segments that help to segregate unvoiced speech that has strong coarticulation with voiced speech. As shown by Cole *et al.* (1996), such portions of speech are important for speech intelligibility. More importantly, segments are the basic units for classification, which enables the grouping of unvoiced speech.

A natural speech utterance contains silent gaps and other sections masked by interference. In practice, one needs to group the utterance across such time intervals. This is the problem of sequential grouping (Bregman, 1990; Wang and Brown, 2006). In this study, we handle this problem in a limited way by applying feature-based classification, assuming nonspeech interference. Systematic evaluation shows that, although our system yields good performance, it can be further improved with better sequential grouping. The assumption of nonspeech interference is obviously not applicable to mixtures of multiple speakers. Alternatively, grouping T-F segments sequentially may be achieved by using speech recognition (Barker *et al.*, 2005) or speaker recognition (Shao and Wang, 2006) in a top-down manner. Although these model-based studies on sequential grouping show promising results, the need for training with a specific lexicon or speaker set limits their scope of application. Substantial effort is needed to develop a general approach to sequential grouping.

VII. CONCLUSION

We have proposed a monaural CASA system that segregates unvoiced speech by performing onset-offset-based segmentation and feature-based classification. This system, together with our previous model for voiced speech segregation, yields a complete system that segregates both voiced and unvoiced speech from nonspeech interference. To our knowledge, this is the first systematic study on unvoiced speech segregation. Quantitative evaluation shows that our system captures most of unvoiced speech without including much interference.

ACKNOWLEDGMENT

This research was supported in part by an AFOSR grant (No. FA9550-04-01-0117), an AFRL grant (No. FA8750-04-1-0093), and an NSF grant (No. IIS-0534707).

NOMENCLATURE

a = Order of a gammatone filter
$A(c,m,\tau)$ = Autocorrelation function of filter response at delay τ in channel c and frame m
$\overline{A(c,m)}$ = Average of $A(c,m,\tau)$ over τ
$A_E(c,m,\tau)$ = Autocorrelation function of response envelope at delay τ in channel c and frame m
$\overline{A_E(c,m)}$ = Average of $A_E(c,m,\tau)$ over τ
α = Ratio of total speech energy to total voiced speech energy
b = Equivalent rectangular bandwidth of a gammatone filter
c = Filter channel index
$C(c,m)$ = Cross-channel correlation of filter responses between channels c and $c+1$ at frame m
$C_E(c,m)$ = Cross-channel correlation of response envelopes between channels c and $c+1$ at frame m
E_1 = Total target energy in voiced speech frames
E_2 = Total interference energy in voiced speech frames
f = Center frequency of a gammatone filter
f_c = Center frequency of filter channel c
$g(f,t)$ = Impulse response of a gammatone filter centered at frequency f
H = Hypothesis
H_0 = Hypothesis that a T-F region is target dominant
$H_{0,a}$ = Hypothesis that a T-F region is dominated by an expanded obstruent
$H_{0,b}$ = Hypothesis that a T-F region is dominated by any phoneme other than an expanded obstruent
H_1 = Hypothesis that a T-F region is interference dominant
m = Time frame index
n = Discrete time
N_1 = Total number of voiced speech frames
N_2 = Total number of frames other than voiced speech frames
p = Probability density
P = Probability
P_{EL} = Percentage of energy loss
P_{NR} = Percentage of noise residue
s = Segment
S_T^1 = Segregated stream for voiced speech
S_T^2 = Segregated stream for voiced speech and voiced T-segments
S_T^3 = Segregated stream for both voiced and unvoiced speech
t = Continuous time
T_m = Frame shift
T_n = Discrete sampling time
τ = Time delay
u_{cm} = Time-frequency unit of channel c and frame m
$x(t)$ = Input signal
$x(c,t)$ = Response of filter channel c to input signal
$x_E(c,t)$ = Response envelope of filter channel c
$X(c,m)$ = Cochleagram value in channel c and frame m
$X(m)$ = Cochleagram at frame m
$\mathbf{X}_s$ = Cochleagram of segment s
$X_s(c,m)$ = Cochleagram value of segment s in channel c and frame m
$X_s(m)$ = Cochleagram of segment s at frame m
$X_T(c,m)$ = Cochleagram value of segregated voiced target in channel c and frame m
$X_T(m)$ = Cochleagram of segregated voiced target at frame m

[1]Nonspeech sounds are posted at http://www.cse.ohio-state.edu/pnl/corpus/HuCorpus.html.

Ali, A. M. A., and Van der Spiegel, J. (**2001a**). "Acoustic-phonetic features for the automatic classification of fricatives," J. Acoust. Soc. Am. **109**, 2217–2235.
Ali, A. M. A., and Van der Spiegel, J. (**2001b**). "Acoustic-phonetic features for the automatic classification of stop consonants," IEEE Trans. Audio, Speech, Lang. Process. **9**, 833–841.
Barker, J., Cooke, M., and Ellis, D. (**2005**). "Decoding speech in the presence of other sources," Speech Commun. **45**, 5–25.
Benesty, J., Makino, S., and Chen, J., Ed. (**2005**). *Speech Enhancement* (Springer, New York).
Boersma, P., and Weenink, D. (**2004**). PRAAT, doing phonetics by computer, Version 4.2.31, (http://www.fon.hum.uva.nl/praat/) Last viewed 6/20/08.
Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
Bridle, J. (**1989**). "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in *Neurocomputing: Algorithms, Architectures, and Applications*, edited by F. Fogelman-Soulie and J. Herault, (Springer, New York), pp. 227–236.
Brown, G. J., and Cooke, M. (**1994**). "Computational Auditory Scene Analysis," Comput. Speech Lang. **8**, 297–336.
Brungart, D., Chang, P. S., Simpson, B. D., and Wang, D. L. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.
Canny, J. (**1986**). "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell. **8**, 679–698.
Cole, R. A., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (**1996**). "The contribution of consonants versus vowels to word recognition in fluent speech," in Proceedings of IEEE ICASSP, pp. 853–856.
Darwin, C. J. (**1997**). "Auditory grouping," Trends Cogn. Sci. **1**, 327–333.
Dewey, G. (**1923**). *Relative Frequency of English Speech Sounds* (Harvard University Press, Cambridge, MA).
Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York).
Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).
French, N. R., Carter, C. W., and Koenig, W. (**1930**). "The words and sounds of telephone conversations," Bell Syst. Tech. J. **9**, 290–324.
Garofolo, J. *et al.* (**1993**). "DARPA TIMIT acoustic-phonetic continuous speech corpus," Technical Report No. NISTIR 4930, National Institute of Standards and Technology.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Greenberg, S., Hollenback, J., and Ellis, D. (**1996**). "Insights into spoken language gleaned from phonetic transcription of the switchboard corpus," in Proceedings of ICSLP, pp. 24–27.
Helmholtz, H. (**1863**). *On the Sensation of Tone*, 2nd English ed., translated by A. J. Ellis (Dover, New York).
Hermansky, H., and Sharma, S. (**1999**). "Temporal patterns (TRAPs) in ASR of noisy speech," in Proceedings of IEEE ICASSP, pp. 289–292.

Hu, G. (**2006**). "Monaural speech organization and segregation," Ph.D. thesis, The Ohio State University Biophysics Program.
Hu, G., and Wang, D. L. (**2001**). "Speech segregation based on pitch tracking and amplitude modulation," in Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 79–82.
Hu, G., and Wang, D. L. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.
Hu, G., and Wang, D. L. (**2005**). "Separation of fricatives and affricates," in Proceedings of IEEE ICASSP, pp. 749–752.
Hu, G., and Wang, D. L. (**2006**). "An auditory scene analysis approach to monaural speech segregation," in *Topics in Acoustic Echo and Noise Control*, edited by E. Hansler and G. Schmidt (Springer, Heidelberg, Germany), pp. 485–515.
Hu, G., and Wang, D. L. (**2007**). "Auditory segmentation based on onset and offset analysis," IEEE Trans. Audio, Speech, Lang. Process. **15**, 396–405.
Huang, X., Acero, A., and Hon, H.-W. (**2001**). *Spoken Language Processing: A Guide to Theory, Algorithms, and System Development* (Prentice-Hall, Upper Saddle River, NJ).
Ladefoged, P. (**2001**). *Vowels and Consonants* (Blackwell, Oxford, UK).
Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.
Lyon, R. F. (**1984**). "Computational models of neural auditory processing," in Proceedings of IEEE ICASSP, pp. 41–44.
Moore, B. C. J. (**2007**). *Cochlear Hearing Loss*, 2nd ed. (Wiley, Chichester, UK).
Nooteboom, S. G. (**1997**). "The prosody of speech: Melody and rhythm," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle, and J. Laver (Blackwell, Oxford, UK), pp. 640–673.
Parsons, T. W. (**1976**). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. **60**, 911–918.
Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (**1988**). "SVOS final report, Part B: Implementing a gammatone filterbank," Report No. 2341, MRC Applied Psychology Unit.
Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.
Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
Radfar, M. H., Dansereau, R. M., and Sayadiyan, A. (**2007**). "A maximum likelihood estimation of vocal-tract-related filter characteristics for single channel speech separation," EURASIP J. Audio Speech Music Process., Paper No. 84186.
Romeny, B. H., Florack, L., Koenderink, J., and Viergever, M., Ed. (**1997**). *Scale-Space Theory in Computer Vision* (Springer, New York).
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shao, Y., and Wang, D. L. (**2006**). "Model-based sequential organization in cochannel speech," IEEE Trans. Audio, Speech, Lang. Process. **14**, 289–298.
Slaney, M., and Lyons, R. F. (**1990**). "A perceptual pitch detector," in Proceedings of IEEE ICASSP, pp. 357–360.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Wang, D. L. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.
Wang, D. L., and Brown, G. J. (**1999**). "Separation of speech from interfering sounds based on oscillatory correlation," IEEE Trans. Neural Netw. **10**, 684–697.
Wang, D. L., and Brown, G. J., Ed. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley, Hoboken, NJ).
Wang, D. L., and Hu, G. (**2006**). "Unvoiced speech segregation," in Proceedings of IEEE ICASSP, pp. 953–956.
Weintraub, M. (**1985**). "A theory and computational model of auditory monaural sound separation," Ph.D. thesis, Stanford University Department of Electrical Engineering.

Influence of wall vibrations on the behavior of a simplified wind instrument

Guillaume Nief,[a] François Gautier, Jean-Pierre Dalmont, and Joël Gilbert
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Av. O. Messiaen, 72085 Le Mans Cedex 9, France

(Received 31 October 2007; revised 19 May 2008; accepted 20 May 2008)

The issue of the influence of wall vibrations on the behavior of wind instruments is still under debate. The mechanisms of vibroacoustic couplings involved in these vibrations are difficult to investigate, as fluid-structure interactions are weak. Among these vibroacoustic interactions, the present study is focused on the coupling between the internal acoustic field and the mechanical behavior of the duct. For this purpose, a simplified single reed instrument consisting of a brass tube connected to a clarinet mouthpiece has been studied. A theoretical model of coupling between the plane inner acoustic wave and mechanical modes is developed and suggests that in order to obtain measurable effects of wall vibrations, the geometrical parameters of the studied tube have to be unusual compared to that of real instruments. For a slightly oval-shaped and very thin brass tube, it is shown theoretically and experimentally that a coupling between the inner plane acoustic wave and ovalling mechanical modes occurs and results in disturbances of the input impedance, which can slightly affect the tone color of the sound produced. It is concluded that the reported effects are unlikely to occur in real instruments except for some organ pipes.
 [DOI: 10.1121/1.2945157]

PACS number(s): 43.75.Pq [NHF] Pages: 1320–1331

I. INTRODUCTION

In the past few decades, much progress has been made toward a better understanding of wind instruments.[1–3] Their main physical principles of operation are nowadays quite well understood. Nevertheless, a few topics are still debated. The issue of the effects of wall vibrations on the behavior of wind instruments is one of them. Walls are usually considered as infinitely stiff compared to the enclosed air and thus perfectly rigid. However, wall vibrations can easily be experienced by any player since they can be felt during playing. Optical holography techniques have also been used to observe such vibrations.[4] Moreover, they are claimed to be of great influence by many musicians and instrument makers. As a consequence, this subject, which is often associated with the influence of the constitutive material of the instrument, has been studied since the quite early days in the history of musical acoustics of wind instruments.[5,6] However, past studies dealing with this topic are not very numerous and have produced mixed results. These studies can be classified into groups considering the origin of the vibrations and their effects.

The various vibroacoustic couplings involved in a musical instrument have been investigated theoretically.[7] Two different sources may generate wall vibrations, an acoustical one and a mechanical one. Firstly, the acoustical source involves the excitation of the walls by the inner pressure field. Secondly, the impacts of the reed on the mouthpiece for single reed instruments or direct transmission from vibrations of the lips for brass instruments are mechanical sources that can be responsible for wall vibrations.

Whitehouse[8] investigated the source of the wall vibrations on a simplified brass instrument: a trombone mouthpiece connected to a simple tube. By coupling or uncoupling the lips to the tube or the air column to the tube, he concluded that the mechanical source of vibration was stronger than the acoustical one. For the excitation of the walls by the inner pressure field, Watkinson and Bowsher[9] studied the *in vacuo* modes of trombone bells with varying material and geometrical properties using a finite element technique. They calculated the mechanical response to the internal acoustic field. They found that the modes usually came by pairs with slightly different associated eigenfrequencies and normal modal shapes due to the asymmetry of the system. They also observed that the inner acoustic plane wave couples only with "non-perfectly-symmetrical" mode shapes.

Although the fluid-structure interactions in wind instruments are weak, the vibrations of the walls, no matter their origin, may then induce sound radiation in the surrounding external or internal fluid. Direct radiation to the external fluid may contribute to the overall radiated field. The radiation to the internal fluid may induce a perturbation of the oscillations of the air column and thus disturb indirectly the radiated field. For the coupling with the inner acoustic field, Yousri and Fahy[10] described theoretically, with a nonmusical aim, the coupling between the modes of a cylindrical shell and the inner plane acoustical wave below the cutoff frequency of the first helical mode. The nonaxisymmetrical mechanical modes can be coupled only if the cylindrical shell is geometrically slightly distorted or presents inhomogeneous material properties. Backus[11] investigated the effects of the vibration of walls considering both the external radiation from the instrument's body and the possible air column alteration. He concluded that the vibrations are extremely

[a]Electronic mail: guillaume.nief.etu@univ-lemans.fr

weak and do not affect the steady tones for both of the considered coupling mechanisms. On the contrary, Nederveen and Dalmont[12] observed a spectacular effect of wall vibrations in the case of an organ pipe, the bifurcation toward a pseudoperiodic regime. Nief *et al.*[13] reported similar effects in the case of a very thin plastic tube connected to a clarinet mouthpiece. Moreover, Picó Vila and Gautier[14] developed a multimodal coupling model and showed theoretically that the input impedance of acoustic waves in the pipe could be slightly disturbed by wall vibrations.

Among the possible couplings, the present article studies the coupling between the inner fluid and wall vibrations. It is supposed that the walls are excited by the inner pressure field, and it is investigated how the resulting vibrations can have an influence on the input impedance and on the produced sounds. This study is carried out on a simplified instrument composed of a single reed and a simple straight tube. In Sec. II, theoretical considerations are given for the input impedance of a vibrating tube and for the instrument in playing conditions, and show that to obtain a measurable effect, the tested tube has to be very thin (0.2 mm) and slightly oval shaped, which is not realistic for a wind instrument. In Sec. III, experimental mechanical modal analysis and acoustical input impedance measurements are presented in order to study in detail the vibroacoustic coupling. The system is also tested in playing conditions to investigate the effects on sound produced in terms of spectral content and playing frequency. The measured effects are then interpreted using the physical model in a playing situation, which allows the musical sounds to be simulated.

II. THEORETICAL CONSIDERATIONS

A. Acoustic input impedance of a tube and influence of wall vibrations

1. Input impedance of a rigid tube

A musical wind instrument can be described as a coupled exciter-resonator system. The exciter can be characterized by a nonlinear relation between acoustical pressure and acoustical volume velocity.[15] The resonator can be characterized by a simple linear proportionality in the frequency domain between the acoustical pressure $P(\omega)$ and the acoustic volume velocity $U(\omega)$, where ω is the angular frequency,

$$Z(\omega) = \frac{P(\omega)}{U(\omega)}. \tag{1}$$

Nonlinear phenomena such as shock waves or nonlinear losses at open ends sometimes have to be taken into consideration when describing the acoustic resonator, but they are not considered here. The quantity $Z(\omega)$ expressed at the input of the resonator is the acoustic input impedance. It depends mainly on the bore shape of the tube and is often used for the study of the resonator part of a wind instrument. Indeed, it can be linked to important physical properties of the instrument under musical performance. The positions of the peaks are strongly linked with the playing frequency and thus with the intonation of the instrument.[16] Input impedance characteristics are also related to the timbre and facility of sound production of different notes.[17]

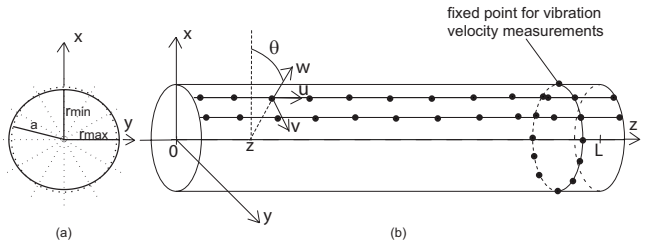


FIG. 1. (a) Polar graph of the oval cross section of the tube. (b) Notations and coordinates systems used in the Appendix for the pseudocylindrical shell. Position of the measured points on the tested tube for modal analysis (dots:tested points).

For a cylindrical and perfectly rigid tube open at the end, considering only the plane acoustic wave and taking into account the length correction due to radiation, the acoustic input impedance can be written as

$$Z_r(\omega) = \frac{\rho_0 c_0}{S} j \tan[k(\omega)L], \tag{2}$$

where c_0 is the speed of sound, ρ_0 the air density, L the equivalent length of the tube, taking into account the end correction due to radiation, and S its cross-sectional area. $k(\omega)$ is the complex wave number, taking into account the viscous and thermal losses at the walls,[18]

$$k(\omega) = k_0\left(1 + \frac{\sqrt{2}(1-j)}{a\sqrt{k_0}}(\sqrt{l_v} + (\gamma - 1)\sqrt{l_h})\right)^{1/2}, \tag{3}$$

where $k_0 = \omega/c_0$, $l_v \approx 4 \times 10^{-8}$ m is the characteristic length for viscous effects in air, $l_h \approx 5.6 \times 10^{-8}$ m is the characteristic length for the thermal effects in air, and $\gamma = 1.4$ is the ratio of specific heats for air. The index r associated with the input impedance indicates that Eq. (2) is the expression for a rigid tube. The modulus of this input impedance Z_r shows peaks corresponding to the acoustical resonances of the closed-open air column.

2. Correction factor to input impedance due to wall vibration

The previous description does not take into account possible wall vibrations. In order to describe those vibrations, it is necessary to consider that the tube behaves as an elastic shell. The instrument is modeled in a simple way by a cylindrical shell clamped on one end and free on the other. The shell is excited by the internal acoustic pressure field and set into vibration. The vibrations induce a disturbance of the initial pressure field. This modification of the air column oscillation yields a slight modification of the input impedance of the perfectly rigid tube.[14] In this approach, structural vibrations and inner acoustic pressure are coupled. The central point for modeling the acoustic input impedance of the vibrating tube is to take into account a slight ellipticity of the tube, which models the effective asymmetry. In a first approximation, its radius r can be written using the polar equation, which is plotted in Fig. 1(a),

$$r(\theta) = a[1 + \epsilon \cos(2\theta)], \tag{4}$$

where a denotes the mean radius and ϵ an ellipticity parameter, which is small compared to unity. With notations of Fig. 1(a), ϵ is equal to $(r_{\max} - r_{\min})/2a$.

The consequence of this asymmetry is the existence of a vibroacoustic coupling between the acoustic plane wave and the ovaling modes of the structure. Ovaling modes are characterized by a circumferential modal shape in a $\sin(2\theta)$ form and are often the lowest ones for geometries similar to wind instruments. Physically, a section of the tube is subjected to a uniform pressure distribution due to the plane wave. If the tube is perfectly circular, this tends to dilate and contract it only in a cylindrically symmetric manner, and the section of the tube is only subjected to tension forces. If the tube is oval shaped, then the isotropic pressure distribution implies also bending forces that tend to round the tube by enlarging the small diameter and shortening the bigger one. This movement is directly linked to the ovaling deformation of the pipe. A detailed description of the vibroacoustic coupling is given in the Appendix. The conclusion of this Appendix is the analytical expression of the acoustic input impedance of the vibrating tube, which can be written as

$$Z(\omega) = Z_r(\omega)[1 + C(\omega)], \tag{5}$$

where C is a correction factor describing the wall vibration effect. Considering only the interaction between the internal acoustic pressure and a single ovaling mode, it is shown that

$$C(\omega) \propto \frac{\epsilon^2}{(1 - e^{-2jkL})\cos(kL) m_\mu [\omega_\mu^2 (1 - j\eta_\mu) - \omega^2]}, \tag{6}$$

where m_μ is the modal mass, which is defined in the Appendix in Eq. (A5), ω_μ is the modal natural angular frequency, and η_μ is the modal damping, which is the natural decrease rate of the mode. These parameters are obtained experimentally from the modal testing of the tube used in the experiments (see Sec. III B 1). Equation (6) allows a direct interpretation of the correction factor. Firstly, the bigger the ellipticity ϵ, the more the input impedance is disturbed. This is due to the increase in the coupling between the ovaling mode and the inner pressure field. Secondly, C is inversely proportional to the modal mass m_μ, which means that disturbance will be more important when the cylinder is thin and light. Thirdly, when the driving angular frequency ω approaches the mechanical angular eigenfrequency, the disturbance increases. Finally, at the acoustical resonances of the tube, when $\cos(kL)$ is minimum, the disturbance is maximum. The correction factor C may then take significant values when a frequency coincidence is realized between an acoustical frequency and a mechanical ovaling eigenfrequency.

The value of C and Z can be computed. In Fig. 2, the modulus of the correction factor C is plotted versus frequency. The regularly spaced peaks correspond to the effect of acoustical resonance, and the other peaks correspond to the effects of the resonances of the ovaling modes. Figure 3 represents the modulus of the computed input impedance of the vibrating tube. It shows additional peaks due to the mechanical resonances of ovaling modes.

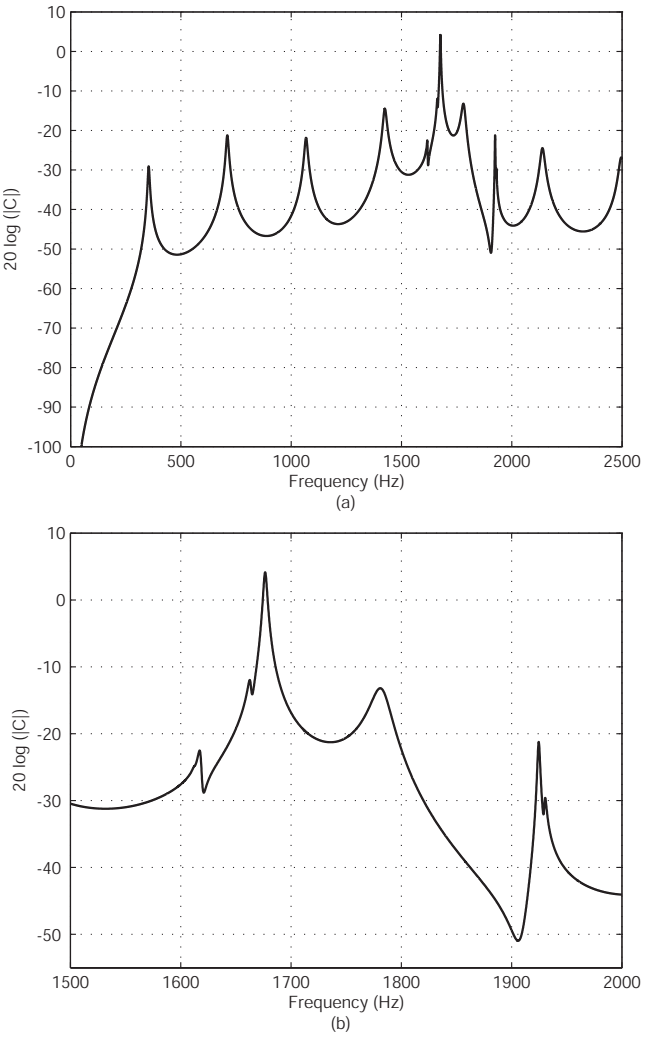


FIG. 2. Calculated correction factor C of a brass tube, 24 cm long, 0.2 mm thick, 7.5 mm in radius, with an 8% ellipticity. (a) Broad frequency band; (b) zoom (modulus in the dB scale).

B. Model of a clarinetlike instrument in playing conditions

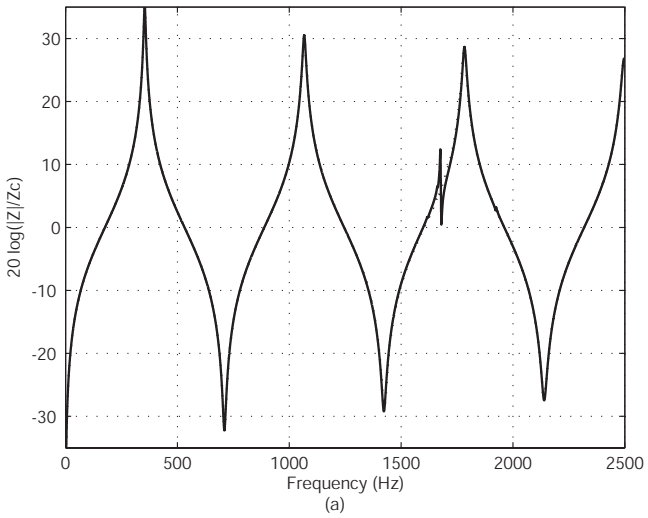
In the playing situation, the air column oscillation described by its input impedance is coupled to the mouthpiece and the player. In order to take into account this coupling, the instrument-player system is assimilated to a dynamic system whose behavior can be described by three equations.[19]

(a) The linear behavior of the tube resonator is described by the input impedance Z and the associated equation [Eq. (1)]

(b) The reed is assumed to behave like a single degree of freedom oscillator whose equation of motion is

$$\frac{d^2h(t)}{dt} + g_r \frac{dh(t)}{dt} + \omega_r^2 h(t) = \frac{P_m - p(t)}{\mu_r}, \tag{7}$$

where $h(t)$ is the dynamic reed displacement, $p(t)$ the acoustic pressure inside the mouthpiece, g_r the reed damping factor, ω_r its natural frequency, μ_r its mass over area ratio density, and P_m the static pressure in the player's mouth. (c) The last equation establishes the link between the acoustic volume velocity $u(t)$ through the slit of height H in the absence



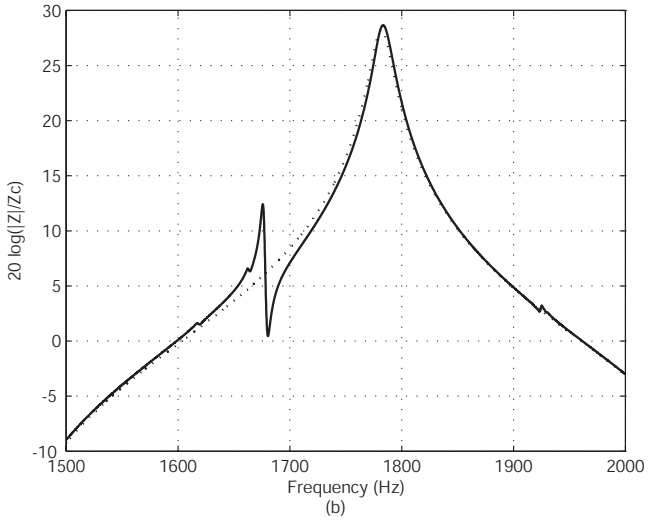


FIG. 3. Calculated acoustic input impedance of a brass tube, 24 cm long, 0.2 mm thick, 7.5 mm in radius, with an 8% ellipticity. (a) Broad frequency band; (b) zoom (modulus in the dB scale). Perfectly rigid tube (dotted line): peaks correspond to acoustical resonances. Vibrating tube (solid line): additional peaks correspond to mechanical ovaling mode resonances.

of flow and of effective width w and the pressure difference $P_m - p(t)$,

$$u(t) = [H + h(t)]w\sqrt{\frac{2|P_m - p(t)|}{\rho}}\,\mathrm{sgn}[P_m - p(t)] \quad \text{if } |P_m - p(t)| \leqslant P_M,$$

$$u(t) = 0 \quad \text{if } |P_m - p(t)| \geqslant P_M, \tag{8}$$

where P_M is the pressure at which the reed closes the slit. For various sets of input parameters, the nonlinear system of equation has various types of solutions. Trivial solutions correspond to situations where the equilibrium position of the reed is stable. Acoustic pressure is then null, and no sound is produced. Periodic solutions correspond to self-sustained regimes of the system, which are of musical interest. Other types of unstable solutions, like unstable periodic regimes, can occur and lead, for example, to multiphonics or to unwanted regimes in a musical context.

This type of model[19] is classically used for the simulation of single reed instruments, allows the description of the dynamical behavior of reed instruments, and favors the interpretation of the experimental observations. At the end of the next section, the experimental measurements in playing situations are simulated using this model.

III. EXPERIMENTS

A. Dimensioning of experimental system

Influence of wall vibrations is described by a correction factor C to input impedance, which is usually very small. As a consequence, no significant effect of wall vibration is generally observed experimentally. This effect can be observed only for particular tubes with specific dimensions. In order to emphasize the possible effects of wall vibrations, the geometrical characteristics and material of a tube whose acoustic input impedance is significantly modified by these vibrations have to be determined. As discussed in the theory explained in Sec. II and developed in the Appendix, a significant perturbation of the acoustic input impedance is expected if mechanical eigenfrequencies occur in the vicinity of one of the first acoustic resonances or antiresonances. In order to provoke this frequency coincidence, the eigenfrequencies of cylindrical shells are computed. Eigenfrequencies depend on geometrical parameters (thickness h, radius a, and length L) and constitutive material parameters (Young's modulus E, Poisson's ratio ν, and density ρ_s) of the tube. For tubes whose length is much larger than the radius, the first modes of the shell are usually the ovaling modes. Our study is thus focused on these ovaling modes. The computation of the modal basis of a cylindrical shell is analytically tractable only for simply supported boundary conditions.[20,21] For other boundary conditions, a semianalytical and exact procedure can be implemented to compute eigenfrequencies and modal shapes.[22] Some approximate formulas can also be used.[23] For slender tubes ($L/a \gg 1$), the first eigenfrequencies of shells of finite length are close to cut on frequencies f_m^c of the flexural wave associated with the circumferential index $m \geqslant 2$ ($m=2$ for ovaling modes) of infinite cylinders. This frequency is given[23] by

$$f_m^c = \frac{m(m^2-1)}{4\pi\sqrt{3}\sqrt{m^2+1}}\frac{h}{a^2}\sqrt{\frac{E}{\rho_s(1-\nu^2)}}, \tag{9}$$

which can be used with $m=2$ to estimate the eigenfrequency of the first ovaling mode.

From the acoustical point of view, the computation of acoustical eigenfrequencies of a closed-open cylindrical tube neglecting the viscous and thermal losses and considering a null pressure at the open end can be performed using

$$f_n = (2n-1)\frac{c}{4L}, \tag{10}$$

where c is the speed of sound in air, L the length of the tube, and n the acoustical longitudinal index.

As experiments are carried out on a clarinetlike instrument, the total length of the system is about 50 cm and the internal radius is $a=7.5$ mm. This gives the following series

of acoustical eigenfrequencies for the first five acoustical modes: 170, 510, 850, 1190, and 1530 Hz. In order to satisfy the frequency coincidence, geometrical and material parameters of a tube have to be determined so that the ovaling mode eigenfrequency is close to one of these values. The material chosen is brass, as it is often used for wind instruments. The material parameters relative to brass are set to $E=110$ GPa, $\rho_s=8700$ kg/m^3, and $\nu=0.3$. According to Eq. (9), the only free parameter is the thickness h. A thickness of $h=0.2$ mm gives a frequency of 1630 Hz, which is in the vicinity of the fifth acoustical resonance. The tube is then connected to a rigid slide, which makes it possible to vary continuously the acoustical resonance frequencies, thanks to a variable total length of the tube, without changing the fixed mechanical resonance frequencies of the vibrating tube. A vibrating tube and a rigid slide satisfying the dimensioning explained in this section has been machined in order to exhibit the wall vibration effect. This device, which makes it possible to satisfy exactly the frequency coincidence or, on the contrary, to avoid it, is connected to a clarinet mouthpiece in an artificial mouth. The vibrating tube is made from brass tubing used in musical instrument making, carefully machined to a thickness of about 0.2 mm. Its length is about 24 cm and its radius is 7.5 mm. The added lengths of the rigid slide in open position and of the mouthpiece that is inside the artificial mouth make an overall length of about 50 cm. The vibrating tube is then used for the mechanical and acoustical measurements.

B. Mechanical and acoustical measurements

1. Experimental modal analysis

Structural modes of the vibrating tube are determined using a standard experimental modal analysis. Eigenfrequencies, mode shapes, and damping parameters can be extracted from measured frequency response functions (FRFs). The FRF, vibration velocity to force ratios, are computed, recording the applied force with an impact hammer and the vibration velocity signal with a laser vibrometer. The laser vibrometer is fixed on a single point of the structure, and the impacts were carried out on different points positioned on a circumference and on two generatrices of the studied tube, as shown in Fig. 1(b). An example of FRF is plotted in Fig. 4. These measurements give a set of FRFs, which is used for modal identification. The method used for this identification is the least squares complex exponential (LSCE) method,[24,25] which is implemented in the modal analysis software from LMS. The results of this modal analysis in terms of modal frequencies, modal damping coefficients, and mode shapes are presented in Fig. 5.

On the plotted frequency band of Fig. 4, six modes have been identified. As shown by their circumferential modal shapes, each mode corresponds to an ovaling mode, with a circumferential index $m=2$. They differ by their respective orientations indicated by the angle φ. Actually, the modes can be classified into three pairs, the modes from each pair having very close eigenfrequencies and the same longitudinal index ($p=1$ for one node, $p=2$ for two nodes, and $p=3$ for three nodes). The duplication of these modes is due to the asymmetry of the cylinder, which implies the phenomenon of breaking of modal degeneracy: a mode with a circumferential index m ($m=2$ for ovaling) and a longitudinal index p splits into two modes with different modal frequencies and circumferential mode shapes differing in rotational angle by $\pi/(2m)$ theoretically, which is $\pi/4$ for ovaling modes. Modes are subsequently referred to as triplets $\mu=(m,p,s)$ for an easier identification, where s is the symmetry index ($s=0$ or $s=1$). For example, the second mode of Fig. 5 is referred to as the triplet (2, 1, 1). In order to validate this modal model obtained from the LSCE method, simulated FRF have been computed. In Fig. 4, a synthesized FRF is plotted against a measured one. The relative error between the two FRF is only a few percent, which supports the validity of the modal basis. It is also noticeable that the measured value of the first eigenfrequency agrees rather well with the calculated one in Sec. II A, which means that the dimensioning of the tube was accurate.

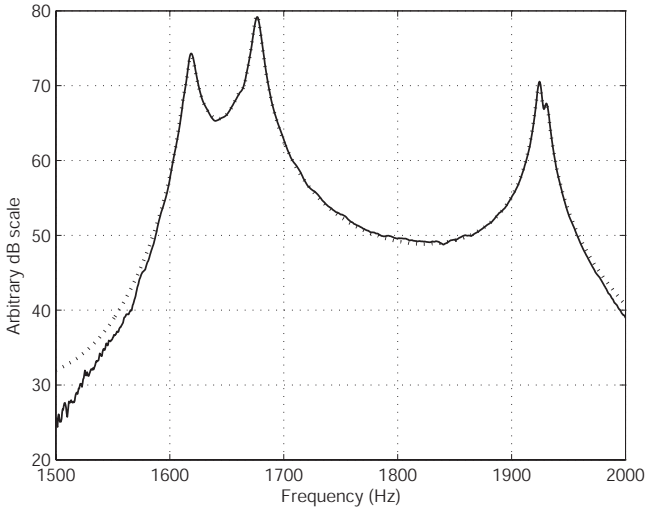


FIG. 4. Measured FRF (solid line) and synthesized FRF using the parameters obtained from a modal analysis (dotted line) for the thin brass tube.

It is shown theoretically that these ovaling modes may couple to the inner air column oscillations and alter acoustic input impedance in the vicinity of their eigenfrequencies. Measurements of input impedances of the vibrating tube are presented in the next section.

2. Acoustic input impedance measurements

The device[26] used for measuring the acoustic input impedances is specially designed in order to impose a carefully calibrated acoustic velocity at the entrance of the tube and to measure the resulting pressure using a microphone. The excitation is managed using a microphone cartridge used as the acoustic velocity calibrated source. The brass tube is measured in two configurations. For the first one, the cross section is quasicircular, corresponding to a quasinull ellipticity parameter $\epsilon=0\%$. For the second one, the tube is made oval by flattening between to small planks so that $\epsilon\approx 8\%$. The tube is actually perfectly circular at its entrance so that it can be connected to the circular slide; the ellipticity increases to be maximum at the end of the tube. In Fig. 6, measurements corresponding to these two configurations are shown.

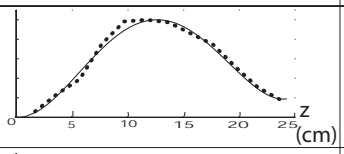
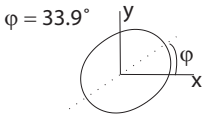
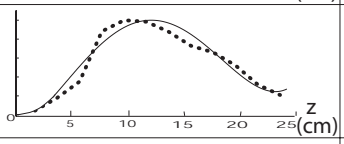
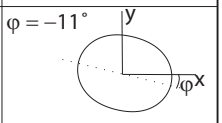
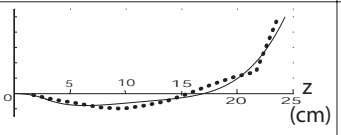
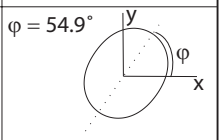
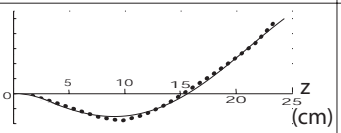
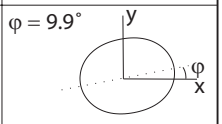
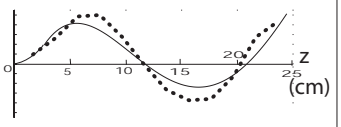
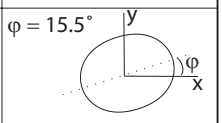
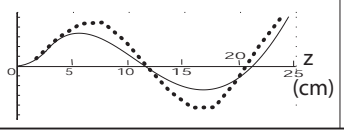
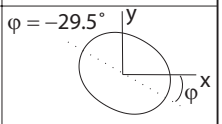
Modal indices (m,p,s)	Frequency (Hz)	Modal damping coefficient (%)	Longitudinal mode shape	Circumferential mode shape
(2,1,0)	1612.6	0.18		
(2,1,1)	1618.1	0.22		
(2,2,0)	1663.3	0.20		
(2,2,1)	1676.6	0.21		
(2,3,0)	1924.6	0.12		
(2,3,1)	1930.1	0.12		

FIG. 5. Modal parameters (eigenfrequencies and dampings) of the brass tube. Longitudinal modal shape: measurements (dotted line) and fit (solid line), circumferential modal shapes.

It can be clearly seen in Fig. 6 that mechanical modes have a strong influence on the input impedance of the tube only if it is slightly flattened. Otherwise the coupling between the plane acoustic wave and the ovaling modes does not occur. The good agreement between the measured disturbances of input impedance and the calculated ones in Sec. II A 2 is also noticeable (see Figs. 3 and 6). These disturbance of input impedance may then influence the sound produced by the system in playing conditions. This is investigated in the next sections using an artificial blowing machine.

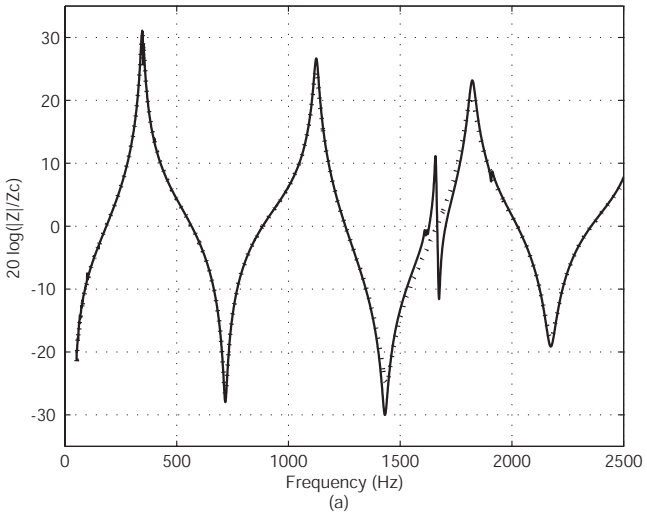
C. Study of the self-sustained oscillations of the system in playing conditions

1. Experimental setup

In order to blow and play the resonator as a musical instrument, the slide-tube device mentioned in Sec. III A is connected to a clarinet mouthpiece, which is put into an artificial mouth. This artificial mouth can produce stable musical sounds for long times, a few minutes or a few hours, allowing us to keep the embouchure parameters, for example, the mouth pressure P_m, constant. The setup, composed of the artificial mouth containing the clarinet mouthpiece, the slide, and the vibrating tube, can be described as an "artificially blown slide clarinet."

In order to investigate possible changes of sound quality due to wall vibration, some sound pressure measurements are carried out. The acoustic pressure inside the mouthpiece is recorded using an Endevco microphone located in a little hole in the mouthpiece. The external radiated pressure is recorded in the near field in order to avoid parasite reflection from surroundings using a Sennheiser KE-4 microphone. This microphone is kept at a fixed position, 5 cm from the end of the tube and 1 cm from the axis, in order to avoid problems due to the spatial dependence of the external acoustic field. This microphone was used to record sounds and to perform informal listening tests (see Sec. III C 2). Pressures are recorded for various playing frequencies obtained by pulling or pushing the slide. As the vibrations of the walls modify the input impedance in the vicinity of the eigenfrequencies of ovaling modes, effects on the acoustic signals are expected when a harmonic of the playing frequency matches one of these mechanical eigenfrequencies. As the resonator is cylindrical, this coincidence can occur on an even harmonic with an acoustical antiresonance and on an odd harmonic with an acoustical resonance. These two configurations are studied in Secs. III C 2 and III C 3, respectively.

The amplitudes and phases of the harmonics of the two pressure signals are recorded using a Stanford SR850 lock-in amplifier. This is performed using the built-in lock-in detec-



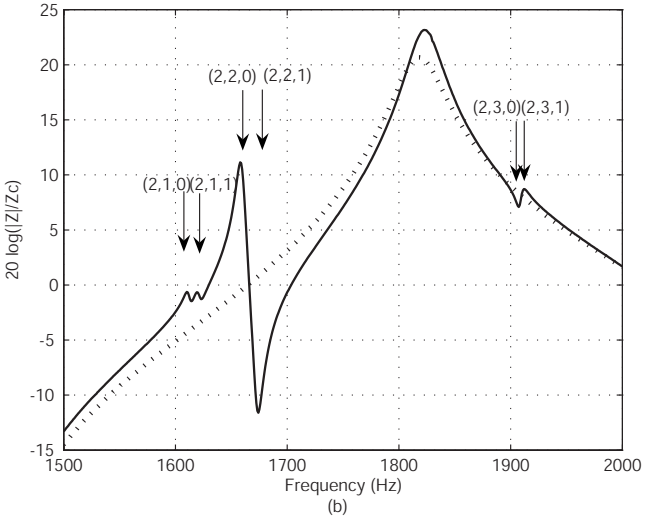


FIG. 6. Measured input impedance of the thin brass tube: (a) broad frequency band; (b) zoom on the frequency band containing mechanical modal frequencies. (modulus in the dB scale). Quasicylindrical tube (dotted line): no coupling with mechanical modes. Slightly flattened tube (solid line): coupling with mechanical modes.

tion algorithm. The lock-in detection was chosen because it has a much better signal to noise ratio for the values of amplitude and phase of each harmonic of the signal than the fast Fourier transform of recorded sounds.

2. Acoustic resonance perturbation

For a particular playing configuration depending on the slide length, the perturbation of input impedance due to wall vibration can be positioned on an acoustical resonance. A perturbation of the self-sustained oscillations occurs when the system plays on the second periodic regime of oscillation, a musical 12th above the fundamental regime. This is illustrated in Fig. 7, where the input impedance is plotted with the position of the harmonics (fundamental, $H2$, $H3$, and $H4$). The third harmonic $H3$ is expected to be perturbed when the system is played because the mechanical additional peak strongly disturbs the input impedance around this harmonic. The amplitudes of the harmonics of the inner pressure signal versus playing frequency are plotted on Fig. 8. Each measurement point corresponds to a position of the slide.

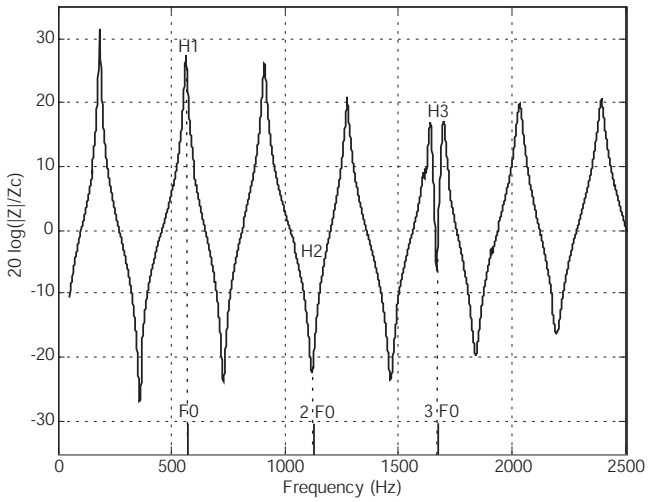


FIG. 7. Input impedance of the vibrating tube connected to the slide and to the mouthpiece when the frequency coincidence between the third harmonic $H3$ and the mechanical eigenfrequency of the (2, 2, 0) ovaling mode is satisfied. Frequencies of harmonics $H1$, $H2$, and $H3$ are multiples of the playing frequency F_0.

It appears that the third harmonic is strongly perturbed for two particular frequency ranges. This corresponds to the influence of two couples of mechanical ovaling modes (2, 1, 0) at 1612 Hz and (2, 1, 1) at 1618 Hz for the playing frequencies around 538 Hz ($3\times538=1614$ Hz, see Fig. 5) and (2, 2, 0) at 1663 Hz and (2, 2, 1) at 1676 Hz for the playing frequencies around 555 Hz ($3\times555=1665$ Hz, see Fig. 5). The other harmonics are also perturbed around these two frequencies. This can be attributed to a nonlinear coupling between harmonics. Using the complex amplitude of each harmonic, an additive sound synthesis is performed. The fundamental frequency is kept constant for each step in order to focus attention on varying tone color and not on varying pitch. The informal listening of the results of this synthesis shows that these perturbations are audible. Another way to illustrate this tone color modification is the direct listening of the sounds produced by the system. In the case of frequency coincidence, a slight timbral change is heard when comparing the sound of the pipe when it is grasped by the hand and

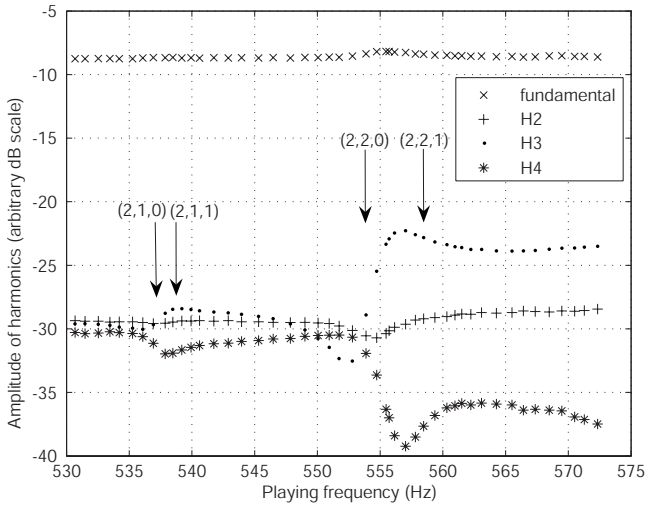


FIG. 8. Amplitude of harmonics of the inner pressure signal as a function of playing frequency (measurements).

the sound of the free pipe. The sounds could be easily recognized in *a-b* type comparison tests. In the other case, when the coincidence is not satisfied, no difference is heard whether the tube is grasped or not.

For this coincidence configuration between mechanical and acoustical resonance, the disturbance is rather important and is reflected not only on the third harmonic, as expected, but also on other harmonics. This could involve an effect not only on the spectral content but also on the playing frequency. In order to investigate this effect, the added length of the pipe due to the pulling of the slide was measured precisely for each position of the slide using a calliper rule. These lengths are used to calculate theoretical playing frequencies using Eq. (10), adding the length of the tube itself, the equivalent length of the mouthpiece, and the length correction due to radiation impedance. A comparison between the computed and measured frequencies is performed for two configurations: the free-vibrating tube and the tube with vibration damped. A small effect is pointed out, showing a maximum difference of 1 Hz when the playing frequency is about 555 Hz. This corresponds to the coincidence of harmonic $H3$ with modes (2, 2, 0) and (2, 2, 1).

Using the model of a single reed instrument in playing conditions described in Sec. II B, a simulation based on the experimental setup characteristics can be performed. The set of three equations is solved using the harmonic balance technique,[27] which is a method developed to determine the periodic response of nonlinear dynamic systems. The application of this method to the three equation system provides the playing frequency and the complex values of the amplitudes of the harmonics of the pressure signal.

In order to simulate the variable positions of the slide, the original input impedance is transformed to the new input of the tube and is computed using the following formula:

$$Z_2 = \frac{Z_1 + j\tan(k\delta L)}{1 + jZ_1\tan(k\delta L)}. \tag{11}$$

In Eq. (11), Z_1 is the measured input impedance when the slide is at its minimum length, k is the acoustic wave number, and δL is the added length due to the pulling of the slide. In this case, Z_2 represents the input impedance of the system when the slide is pulled by a length δL. Small variations of δL are used to build a set of input impedances. For each input impedance of this set, the harmonic balance is performed considering that the values of the other input parameters are constant and set to realistic values ($g_r = 2900\ \mathrm{s}^{-1}$, $\omega_r = 2\pi \times 3000$ rad/s, $\mu_r = 0.02\ \mathrm{kg/m}^2$, $H = 1$ mm, $\omega = 1$ cm, and $P_m = 8000$ Pa). The results in terms of amplitudes of harmonics and playing frequencies are displayed in Fig. 9. This figure shows that the model can describe the perturbation of the third harmonic as expected and already measured. The perturbation of other harmonics due to nonlinear coupling is also observed. Some important differences between measurements and simulations can be observed, showing that the used simplified model is not perfectly accurate. Nevertheless, the disturbance due to wall vibrations can be simulated. The observed differences may have various origins. There are a few slight changes in diameter due to the slide and mechanical parts used to connect the tube, which means that the use

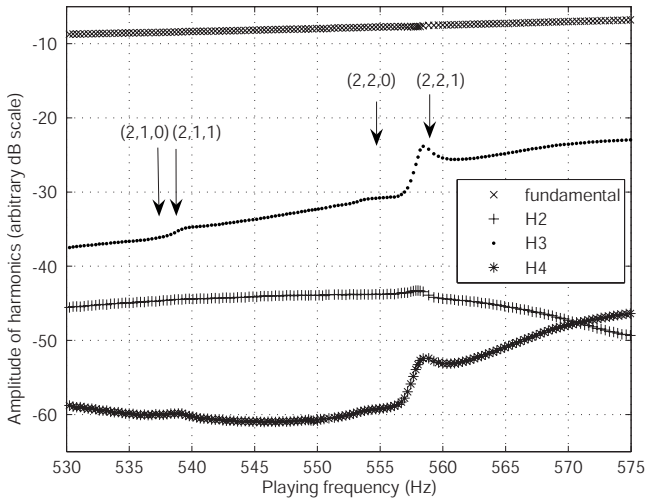


FIG. 9. Amplitude of harmonics of the inner pressure signal as a function of playing frequency (harmonic balance simulation).

of Eq. (11) used for the computation of the input impedance for each simulation may be an error source as the bore is not a pure cylinder. Particularly, the position of antiresonances may not be exactly correct, and thus the even harmonics may be simulated with errors. Moreover, the input parameters of the model like the mouth pressure or the height of the slit may be slightly inaccurate and also imply differences between measurements and simulations.

3. Acoustic antiresonance perturbation by wall vibrations

Using the slide described in Sec. III C 2, a playing configuration can be found where the perturbation of input impedance due to wall vibrations matches an antiresonance. This can be realized when the system plays on the third periodic regime of oscillation, two octaves and a third above the fundamental regime, for a particular position of the slide. This is illustrated in Fig. 10, where the input impedance is plotted with the positions of the harmonics. In such a con-

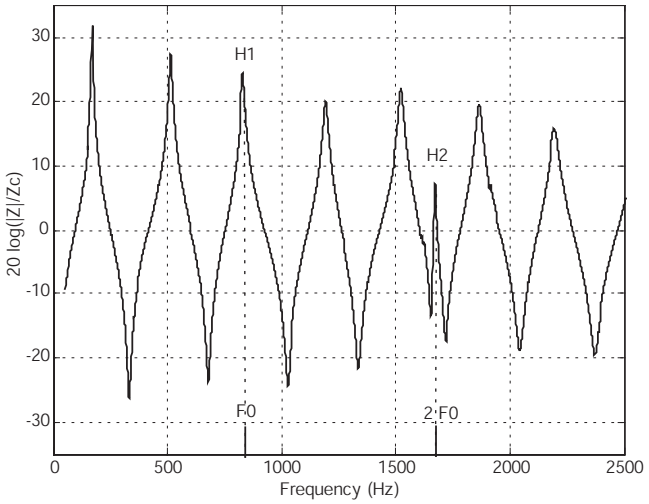
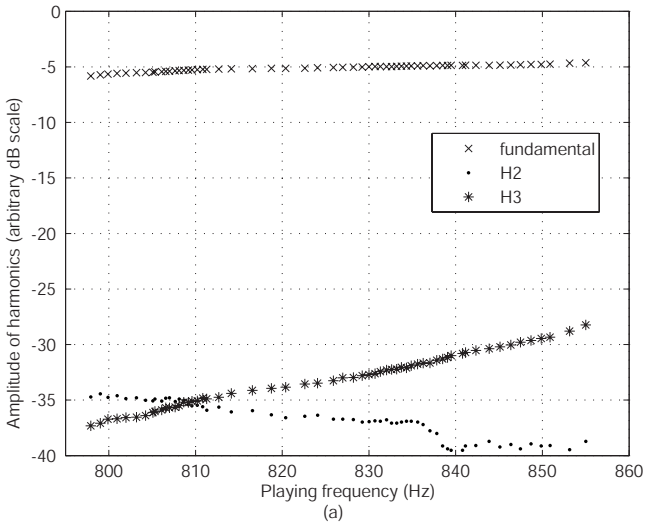


FIG. 10. Input impedance of the vibrating tube connected to the slide and to the mouthpiece when the frequency coincidence between the second harmonic $H2$ and the mechanical eigenfrequency of the (2, 2, 0) ovaling mode is satisfied. Frequencies of harmonics $H1$ and $H2$ are multiples of the playing frequency F_0.



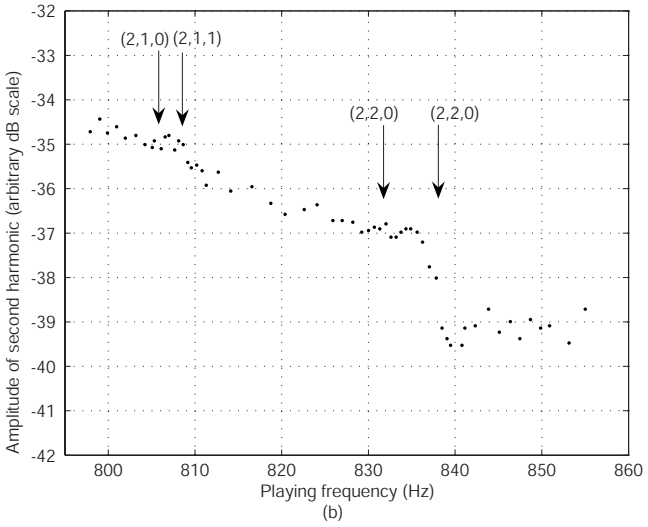


FIG. 11. Amplitude of harmonics of the inner pressure signal as a function of playing frequency (measurements).

figuration a perturbation of the second harmonic $H2$ is expected when the system is played. Small variations of the slide's length allows this coincidence condition to be either satisfied or not.

For each position of the slide, the amplitudes of the harmonics of the measured inner pressure signal versus playing frequency are recorded. Results are given in Fig. 11. It is noticeable that the second harmonic ($H2$) is perturbed at two particular frequencies corresponding to the influence of two couples of mechanical ovaling modes (2, 1, 0), (2,1,1) for the playing frequencies around 807 Hz ($2\times807=1614$ Hz, see Fig. 5) and (2, 2, 0), (2, 2, 0) for the playing frequencies around 834 Hz ($2\times834=1668$ Hz, see Fig. 5). In this case of a coincidence between a mechanical ovaling resonance and an acoustic antiresonance, spectral changes in the musical sound are also induced.

In the same way as for the case of an acoustic resonance perturbation (Sec. III C 2), the experiment has been simulated numerically using the three equation model. A set of acoustic input impedances corresponding to various positions of the slide is calculated. For each of them, the model is solved using the harmonic balance method. Results are displayed in Fig. 12. Significant differences from the measurements are again observed, which show the limit of the simplified model, for the same reasons as in Sec. III C 2. Spectral changes are however fairly well simulated using this model.

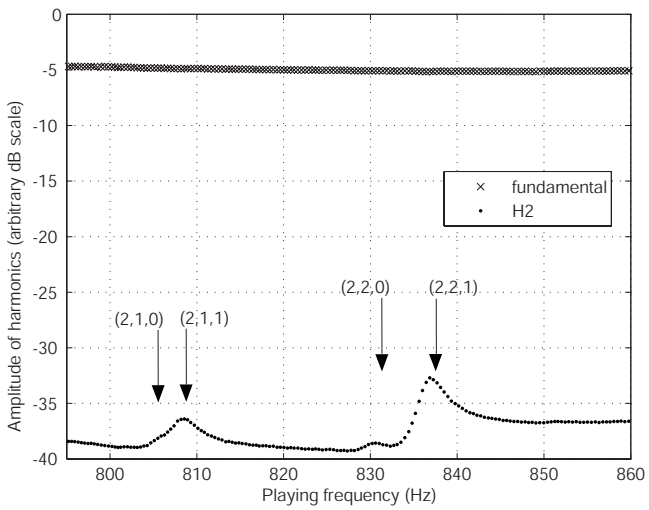


FIG. 12. Amplitude of harmonics of the inner pressure signal as a function of playing frequency (harmonic balance simulation).

IV. DISCUSSION

In Sec. III C, for this unusual instrument, it has been shown that some slight tone color variations can occur when an acoustic resonance or antiresonance coincides with the resonance frequency of an ovaling mode of the resonator. An approximate value of the first ovaling frequency has been computed for real instruments using realistic geometrical and material parameters. The values, given in Table I, differ a lot between various wind instruments. As the eigenfrequency of the first ovaling mode for the flute is 1900 Hz, effects similar to the ones described in the present paper, such as slight tone color modification, might occur but may however be even weaker due to weak ellipticity and because of additional mechanical damping due to stoppers or finger tips. Acoustic impedance measurements of a real flute resonator has been carried out, but no clear evidence of additional peaks appeared. Brass instrument bells, which have large diameters and low mechanical frequencies[28] and which are free to vibrate, might show small acoustic impedance disturbance. For metal clarinet or trombone slide, coincidence may occur on a high harmonic, and effects are unlikely to be expected. For the wooden clarinet, such effects, even if present, are likely to be inaudible as the ovaling frequency lies in the ultrasonic range. On the contrary, because of their softer construction material (lead tin alloy) and a larger inner radius, greater effects can occur for organ pipes, as the ovaling frequency is only a few hundred hertz, as shown in Table I. This mechanical frequency may match the fundamental playing frequency on particular pipes in a stop. In that case, wall vibrations may even induce transition to pseudoperiodic regimes of oscillation and wolf notes, as has been already observed experimentally.[12] To confirm this assumption, input impedance measurements of a real organ pipe resonator have been

TABLE I. Approximate value of the frequency of the first ovaling mode for wind instruments.

	Ebony Bb clarinet	Metal Bb clarinet	Organ pipe	Flute	Trombone slide
Thickness (mm)	5	0.5	0.5	0.35	0.5
Internal radius (mm)	7	7	25	9.5	7.5
Length (m)	0.5	0.5	0.6	0.6	2.5
Young's modulus (GPa)	3	110	35	125	110
Poisson's ratio	0.35	0.33	0.33	0.33	0.33
Density (kg/m^3)	1000	8500	8700	8700	8500
First ovaling mode frequency (Hz)	23 000	4800	210	1900	4200

carried out and have shown additional peaks due to mechanical ovaling modes at frequencies of about 300 Hz.

V. CONCLUSION

In this paper, effects of wall vibrations on a simplified wind instrument have been investigated, considering a coupling between the duct and the internal pressure field. For a very thin and oval shaped tube, with geometrical parameters unusual in comparison to that of real instruments, such a coupling has been demonstrated.

Firstly, the theoretical prediction of a coupling between the mechanical ovaling mode of a duct with an oval cross section and the plane propagating acoustic mode has been presented and confirmed experimentally. Due to this coupling, the acoustic input impedance is perturbed. The perturbation can be significant near the eigenfrequencies of the ovaling modes of the duct. The importance of asymmetry already noticed[1,9] and studied[14] has also been shown experimentally and theoretically; it has been demonstrated that for a perfectly symmetrical tube such a coupling cannot occur. In practice, for real instruments, perfect symmetry is not achievable due to the presence of side holes or keys, for example, but would not be sufficient to achieve measurable coupling with mechanical ovaling modes of the structure.

Secondly, the influence of the modification of input impedance on the sound produced by the tested vibrating tube connected to a clarinet mouthpiece has been investigated experimentally and simulated numerically. The results show that when the eigenfrequency of a mechanical mode matches an acoustic resonance or antiresonance, particular behaviors, different from the perfectly rigid case, are found. In this case of frequency coincidence, the spectral content of sounds is slightly different from that of the perfectly rigid case. These spectral changes involve an audible tone color modification. An artificial oval shaped and thin walled system has been constructed for these effects to be measurable. For real orchestral instruments, although the previously described mechanism is not to be excluded completely for a particular instrument, similar effects are thus unlikely to be measurable and, even if they were, would only occur for one or a few notes. Higher mechanical eigenfrequencies (except for the flute), additional vibration damping due to finger tips, stoppers and stronger walls, and low ellipticity weaken the effects shown in this paper. The regular lead-tin organ pipes are exceptions because they have low ovaling frequencies. Such effects can only occur for pipes whose mechanical eigenfrequency coincides with the fundamental playing frequency or with one of its harmonics. However, other physical processes could possibly induce wall vibration effects as all vibroacoustic couplings have not been investigated in this paper.

ACKNOWLEDGMENTS

The authors are grateful to Ruben Picó Vila for fruitful discussions. The authors would also like to thank Serge Collin, Patrick Colas, and Emmanuel Brasseur for their technical contributions to this study.

APPENDIX A: COMPUTATION OF THE CORRECTION FACTOR *C* TO INPUT IMPEDANCE

The principle of the computation of the correction factor C is based on a perturbation method. Firstly, the vibration velocity field of the walls of the resonator resulting from the inner pressure field, as it would be with no vibration of walls, is computed. The vibration field is then used in order to calculate the new inner pressure field, taking into account the inner radiation from the vibrating walls. The input impedance of the vibrating resonator is then computed. This method is a similar approach to the one of Picó Vila and Gautier,[14] but the present one takes into account more realistic mechanical boundary conditions and a more precise description of the mechanical modal basis.

The resonator of the instrument is a pseudocylindrical shell, clamped at the entrance and free at its end, presenting a slight ellipticity, which can be described using the polar equation $r(\theta)=a[1+\epsilon\cos(2\theta)]$ [see Fig. 1(a)]. The acoustic boundary conditions are modeled by a uniform acoustic velocity V_0 at the entrance of the tube and a null pressure at the end. Figure 1(b) displays the notations of the following description.

The motion of this shell is described by the displacement field $\boldsymbol{X}(M)=[u,v,w]^t$, which results from the inner acoustic field $p(M)$. In the harmonic regime, $e^{j\omega t}$ is implicit and the motion equation can be written in a compact form,

$$-p_s h[\omega^2+\omega_a^2\mathcal{L}]\mathbf{X}(M)=p(M)\cdot\mathbf{n}, \tag{A1}$$

where $\omega_a=1/r(\theta)\sqrt{E/\rho_s(1-\nu)}$ is a parameter equaling the ring frequency if $r(\theta)=a$, ρ_s the density of the material, E its

Young modulus, ν its Poisson's ratio, and h the thickness of the shell.[20] $\mathcal{L}$ is the Donnell stiffness operator of the nondistorted shell. The eigenmodes of the nondistorted shell $\boldsymbol{\phi}_\mu$ in vacuum, which can be referenced using the triplet of indices $\mu=(p,m,s)$, can be written as

$$\boldsymbol{\phi}_\mu = \begin{bmatrix} U_p(z)\sin[m(\theta-\varphi)+s\pi/2] \\ V_p(z)\sin[m(\theta-\varphi)+s\pi/2] \\ W_p(z)\sin[m(\theta-\varphi)+s\pi/2] \end{bmatrix}, \tag{A2}$$

where U_p, V_p, and W_p are the longitudinal modal shapes of the mode of axial index $p=1,2,\ldots$ and $\sin[m(\theta-\varphi)+s\pi/2]$ is the circumferential modal shape with circumferential index $m=0,1,2,\ldots$ and symmetry index $s=0$ or 1. The angle φ represents the direction of the principal axis of the considered mode, which is not necessarily null, as explained by Soedel.[21] Using the expansion over the *in vacuo* modes $\boldsymbol{X}(M)=\Sigma_\mu A_\mu \boldsymbol{\phi}_\mu$, reporting it in Eq. (A1), and projecting it on mode $\boldsymbol{\phi}_{\mu'}$, one can write

$$\rho_s h \sum_\mu A_\mu \int_S \boldsymbol{\phi}^t_{\mu'} \cdot \boldsymbol{\phi}_\mu \frac{\omega_\mu^2-\omega^2}{\omega_a^2} dS = \int_S \frac{\boldsymbol{\phi}^t_{\mu'} \cdot \mathbf{n} p(M)}{\omega_a^2} dS. \tag{A3}$$

The θ dependency of ω_a in the left-hand term of Eq. (A3) is neglected, which means that mechanical modes are not coupled through the ellipticity of the cylinder. It is also considered that the exciting pressure $p(z)$ is the acoustic field in the case of perfectly rigid walls assuming the light fluid approximation. Equation (A4) can then be written as

$$A_\mu m_\mu(\omega_\mu^2-\omega^2) = \int_0^L \int_0^{2\pi} \boldsymbol{\phi}^t_\mu \cdot \mathbf{n} p(z)[1+\epsilon\cos(2\theta)]^2 \times r(\theta) dz d\theta, \tag{A4}$$

where $\mathbf{n}$ is the radial vector so that $\boldsymbol{\phi}^t_\mu \cdot \mathbf{n} = W_p(z)\sin[m(\theta-\varphi)+s\pi/2]$, and μ the modal mass is defined by

$$m_\mu = \int_S \rho_S h \boldsymbol{\phi}^t_\mu \cdot \boldsymbol{\phi}_\mu dS. \tag{A5}$$

The pressure field inside the rigid tube is expressed as

$$p(z) = j\rho_0 c_0 V_0 \frac{\sin[k(L-z)]}{\cos(kL)}. \tag{A6}$$

Using these expressions, considering that $\epsilon \ll 1$ and introducing the parameter η_μ modeling the structural modal damping, the unknown modal amplitudes A_μ are given by

$$A_\mu m_\mu[\omega_\mu^2(1-j\eta_\mu)-\omega^2] = \frac{j\rho_0 c_0 V_0 a}{\cos(kL)} \int_0^L W_p(z)\sin[k(L-z)]dz \times \int_0^{2\pi} \sin[m(\theta-\varphi)+s\pi/2][1+3\epsilon\cos(2\theta)]d\theta, \tag{A7}$$

where ρ_0 is the air density, c_0 the speed of sound, and V_0 the uniform acoustic velocity at the entrance of the cylinder. This expression shows that if $m=2$ (ovaling mode) and $\epsilon=0$ (perfectly circular shell), the modal amplitudes are null and thus the coupling cannot occur. The two integrals of Eq. (A7) are noted as follows:

$$I_p = \int_0^L W_p(z)\sin[k(L-z)]dz,$$

$$I_m = \int_0^{2\pi} \sin[m(\theta-\varphi)+s\pi/2][1+3\epsilon\cos(2\theta)]d\theta. \tag{A8}$$

For example, for a simply supported shell $W_p(z)$ is $W_p(z)=\sin(p\pi/L)$. For other boundary conditions the integral can be calculated using an experimental axial modal profile $W_p(z)$, which has been done in this paper. Knowing $W_p(z)$, the z-axis mode shape that is fitted using a polynomial, A_μ are known and thus the vibrations of the shell are known.

Using this velocity vibration field, the new inner pressure is then calculated using the integromodal approach. The Green's function of the infinite cylinder, considering only the plane wave is,

$$G(z,z_0) = \frac{e^{-jk|z-z_0|}}{2j\pi a^2 k}. \tag{A9}$$

The integral representation of the acoustic field in the tube with vibrating walls is

$$p(z) = \int_S G(z,z_0)\partial_n p(z_0) - p(z_0)\partial_n G(z,z_0) dS_0. \tag{A10}$$

On the entrance surface of the cylinder, we have

$$\partial_n G = \frac{-e^{-jkz}}{2a^2\pi}, \quad p(0)=P_0 \text{ and } \partial_n p = j\omega V_0 \rho_0.$$

At the end surface of the cylinder,

$$\partial_n G = \frac{-e^{jk(L-z)}}{2a^2\pi}, \quad p(L)=0 \text{ and } \partial_n p = -j\omega V_L \rho_0.$$

On the lateral surface,

$$\partial_n G = 0, \quad \partial_n p = -j\rho_0 \omega V \text{ and } V = \dot{w} = j\omega w.$$

Developing Eq. (A10) it is possible to write

$$p(z) = [B^+ + D^+(z)]e^{-jkz} + [B^- + D^-(z)]e^{-jk(L-z)}, \tag{A11}$$

where functions D^+ and D^- take into account the vibrations of the walls. B^+, B^-, D^+, and D^- are then written as

$$B^+ = \frac{1}{1+e^{-2jkL}}[\rho_0 c_0 V_0 + D^-(0)e^{-jkL} - D^+(L)e^{-2jkL}],$$

$$B^- = \frac{1}{1+e^{-2jkL}}[-\rho_0 c_0 V_0 e^{-jkL} - D^-(0)e^{-2jkL} - D^+(L)e^{-jkL}],$$

$$D^+(z) = \frac{-j\rho_0 c_0 \omega}{2a\pi} \sum_\mu \left\{ A_\mu \int_0^{2\pi} \sin[m(\theta_0-\varphi)+s\pi/2] \times [1+\epsilon\cos(2\theta_0)]d\theta_0 \int_0^z e^{jkz_0} W_p(z_0) dz_0 \right\}.$$

$$D^-(z) = \frac{-j\rho_0 c_0 \omega}{2a\pi} \sum_\mu \left\{ A_\mu \int_0^{2\pi} \sin[m(\theta_0 - \varphi) + s\pi/2] \times [1 + \epsilon \cos(2\theta_0)] d\theta_0 \int_z^L e^{-jk(z_0 - L)} W_p(z_0) dz_0 \right\}. \tag{A12}$$

The modal amplitudes are known using Eq. (A7), and the integrals can be calculated knowing $W_p(z)$. The integrals of Eq. (A12) are noted as follows:

$$J_p = \int_0^L e^{jkz_0} W_p(z_0) dz_0,$$

$$K_p = \int_0^L e^{-jk(z_0 - L)} W_p(z_0) dz_0,$$

$$I'_m = \int_0^{2\pi} \sin[m(\theta_0 - \varphi) + s\pi/2][1 + \epsilon \cos(2\theta_0) d\theta_0]. \tag{A13}$$

Using Eq. (A11), the input impedance is

$$Z = \frac{P_0}{V_0} = \frac{B^+ + [B^- + D^-(0)]e^{jkL}}{V_0}. \tag{A14}$$

After some algebra using Eqs. (A7) and (A12), it is possible to write the input impedance as

$$Z = Z_r(1 + C), \tag{A15}$$

where $Z_r = j\rho_0 c_0 \tan(kL)$ is the input impedance of the rigid cylinder and C a correction factor to this input impedance, which is equal to

$$C = \frac{2e^{-jkL}}{\rho_0 c_0 (1 - e^{-2jkL})} \frac{D^-(0)}{V_0} - \frac{2e^{-2jkL}}{\rho_0 c_0 (1 - e^{-2jkL})} \frac{D^+(L)}{V_0}, \tag{A16}$$

where

$$D^+(L) = \sum_\mu \frac{\rho_0^2 c_0^2 V_0 \omega I_m I'_m I_p J_p}{2\pi \cos(kL) m_\mu [\omega_\mu^2 (1 - j\eta_\mu) - \omega^2]},$$

$$D^-(0) = \sum_\mu \frac{\rho_0^2 c_0^2 V_0 \omega I_m I'_m I_p K_p}{2\pi \cos(kL) m_\mu [\omega_\mu^2 (1 - j\eta_\mu) - \omega^2]}. \tag{A17}$$

For a single ovaling mode of index $\mu = (2, p, s)$, the coefficient C is proportional to

$$\frac{\epsilon^2}{(1 - e^{-2jkL}) \cos(kL) m_\mu [\omega_\mu^2 (1 - j\eta_\mu) - \omega^2]}. \tag{A18}$$

[1] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 2005).
[2] C. J. Nederveen, *Acoustical Aspects of Woodwind Instruments* (Northern Illinois University Press, DeKalb, Illinois, 1998).
[3] A. H. Benade, *Fundamentals of Musical Acoustics* (Dover, New York, 1990).
[4] R. Smith, "The effect of material in brass instruments: A review," Proceedings of the Institute of Acoustics **8**, 91–96 (1986).
[5] D. C. Miller, "The influence of the material of wind-instruments on the tone quality," Science **29**, 161–171 (1909).
[6] V. C. Mahillon, *Eléments d'acoustique musicale et instrumentale ("Elements of musical and instrumental acoustics")* (Manufacture générale d'instruments de musique, Bruxelles, 1874).
[7] F. Gautier and N. Tahani, "Vibroacoustic behavior of a simplified musical wind instrument," J. Sound Vib. **213**, 107–125 (1998).
[8] J. Whitehouse, "A study of the wall vibrations excited during the playing of lip-reed instruments," Ph.D. thesis, Technology Faculty, The Open University, 2003.
[9] P. S. Watkinson and J. M. Bowsher, "Vibration characteristics of brass instrument bells," J. Sound Vib. **85**, 1–17 (1982).
[10] S. N. Yousri and F. J. Fahy, "Distorted cylindrical shell response to internal acoustic excitation below the cut-off frequency," J. Sound Vib. **52**, 441–452 (1977).
[11] J. Backus, "Effect of wall material on the steady-state tone quality of woodwind instruments," J. Acoust. Soc. Am. **36**, 1881–1887 (1964).
[12] C. J. Nederveen and J. P. Dalmont, "Pitch and level changes in organ pipes due to wall resonances," J. Sound Vib. **271**, 227–239 (2004).
[13] G. Nief, J.-P. Dalmont, F. Gautier, and J. Gilbert, "Influence des vibrations de parois d'un tuyau sur son impédance d'entrée: Application aux instruments à vent" ("Influence of wall vibrations of a pipe on its input impedance: Application to wind instruments"), Actes du 8ème Congrès Français d'Acoustique, Tours, 2006.
[14] R. Picó Vila and F. Gautier, "The vibroacoustics of slightly distorted cylindrical shells: A model for acoustic input impedance," J. Sound Vib. **302**, 18–38 (2007).
[15] J.-P. Dalmont, J. Gilbert, and S. Ollivier, "Nonlinear characteristics of single reed instruments: Quasistatic volume flow and reed opening measurements," J. Acoust. Soc. Am. **114**, 2253–2262 (2003).
[16] J. Backus, "Input impedance curves for the reed woodwind instruments," J. Acoust. Soc. Am. **56**, 1266–1279 (1974).
[17] J. P. Dalmont, B. Gazengel, J. Gilbert, and J. Kergomard, "Some aspects of tuning and clean intonation in reed instruments," Appl. Acoust. **46**, 19–60 (1995).
[18] M. Bruneau, *Manuel d'acoustique fondamentale ("Fundamentals of acoustics")* (Hermes, Paris, 1998), p. 141–147.
[19] J. Kergomard, "Elementary considerations on reed-instruments oscillations," in *Mechanics of Musical Instruments*, Lecture Notes on CISM, edited by A. Hirschberg A. Hirschberg, J. Kergomard, and G. Weinreich (Springer, Verlag, 1995), pp. 229–290.
[20] A. W. Leissa, *Vibration of Shells* (Acoustical Society of America, Woodbury, NY, 1993).
[21] W. Soedel, *Vibration of Shells and Plates* (Marcel Dekker, New York, 1981).
[22] D. F. Vronay and B. L. Smith, "Free vibration of circular cylindrical shells of finite length," AIAA J. **8**, 601–603 (1970).
[23] R. D. Blevins, *Formulas for Natural Frequency and Mode Shape* (Krieger, Malabar, Florida, 1979).
[24] D. J. Ewins, *Modal Testing: Theory, Practice and Application*, 2nd ed. (Taylor & Francis, London, 2001).
[25] J. Piranda, "Analyse modale expérimentale" ("Experimental modal analysis"), Techniques de l'ingénieur **R6**, 1–29 (2001).
[26] J. P. Dalmont and A. M. Bruneau, "Acoustic impedance measurements: Plane-wave mode and first helical mode contributions," J. Acoust. Soc. Am. **91**, 3026–3033 (1992).
[27] J. Gilbert, J. Kergomard, and E. Ngoya, "Calculation of the steady state oscillations of a clarinet using the harmonic balance technique," J. Acoust. Soc. Am. **86**, 35–41 (1989).
[28] T. R. Moore, J. D. Kaplon, G. D. McDowall, and K. A. Martin, "Vibrational modes of trumpet bells," J. Sound Vib. **254**, 777–786 (2002).

Courtship and agonistic sounds by the cichlid fish *Pseudotropheus zebra*

J. Miguel Simões,[a] Inês G. Duarte, and Paulo J. Fonseca
Departamento de Biologia Animal e Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Bloco C2 Campo Grande, 1749-016 Lisboa, Portugal

George F. Turner
Department of Biological Sciences, University of Hull, HU6 7RX, United Kingdom

M. Clara Amorim
Unidade de Investigação em Eco-Etologia, ISPA, Rua Jardim do Tabaco 34, 1149-041 Lisboa, Portugal

(Received 22 March 2007; revised 6 May 2008; accepted 28 May 2008)

Courtship and agonistic interactions in an African cichlid species present a richer diversity of acoustic stimuli than previously reported. Male cichlids, including those from the genus *Pseudotropheus (P.)*, produce low frequency short pulsed sounds during courtship. Sounds emitted by *P. zebra* males in the early stages of courtship (during quiver) were found to be significantly longer and with a higher number of pulses than sounds produced in later stages. During agonistic intrasexual quiver displays, males produced significantly longer sounds with more pulses than females. Also, male sounds had a shorter duration and pulse period in courtship than in male–male interactions. Taken together, these results show that the acoustic repertoire of this species is larger than what was previously known and emphasize the importance of further research exploiting the role of acoustic stimuli in intra- and interspecific communication in African cichlids.
 [DOI: 10.1121/1.2945712]

PACS number(s): 43.80.Ka [MCH] Pages: 1332–1338

I. INTRODUCTION

In recent years much attention has focused on the role of interspecific mate choice on the impressive rate of speciation of cichlid fishes from the Great African Lakes that have undergone some of the fastest and most extensive adaptive radiations among vertebrates (e.g., Turner, 1999; Albertson *et al.*, 2003). Many authors have proposed that sexual selection driven by female choice acting on male courtship colors may have been a significant factor on the rapid speciation of these fishes (e.g., Couldridge and Alexander, 2001; Genner and Turner, 2005). Males of several African cichlid fishes are known to produce sounds during courtship [reviewed in Lobel (1998) and Amorim (2006)] and recently acoustic signaling has also been pointed out as a possible mechanism involved in reproductive isolation (Lobel, 1998; Amorim *et al.*, 2004) among African Great Lake cichlids.

Less attention has been given to the role of acoustic communication in intraspecific mate choice in these fishes. In the early stages of courtship, male *Pseudotropheus (P.)* quiver to females, producing low-frequency short-pulsed sounds (Amorim *et al.*, 2004), but there are no published records of sound production associated with behavioral elements characteristic of the later stages of courtship (Baerends and Baerends van Roon, 1950), or during agonistic displays. If there is sufficient intraspecific variability in sound production then acoustic communication may play a role in intra- and intersexual selection and influence the outcome of fights and mating decisions in *Pseudotropheus*, as observed in other animals (e.g., Ladich *et al.*, 1992).

The present study was aimed at investigating the full acoustic repertoire of *P. zebra* males and females associated with both courtship and agonistic contexts.

II. METHODS

A. Fish stocks and maintenance

Twenty adult male and twelve adult female first-generation offspring bred from a stock of wild-caught adult *Pseudotropheus zebra* from Nkhata Bay, Malawi (11°36′ N; 34°17′ E) were used in this study. After each trial, the fish were returned to stock tanks. Each tank was fitted with an external power filter and maintained at 25–27 °C by an internal 250-W heater, on a 12:12 h light:dark cycle provided by room lights. A third of the tank's water (pH 7.5–8.5) was changed weekly. Fish were fed twice daily with a mixture of commercial cichlid sticks and koi pellets.

B. Experimental protocol

Experiments were conducted between January and September 2005. Trials were conducted in two aquaria (120 × 60 × 45 cm high) placed on top of a concrete plate supported by two rockwool blocks (100 × 50 × 30 cm). This setup proved to be effective to minimize external noise transmitted through the building improving considerably noise to signal ratio at the low frequencies considered in this study (Fig. 1). Each experimental tank was divided transversally

[a]Electronic mail: jsimoes@ispa.pt

 0001-4966/2008/124(2)/1332/7/$23.00

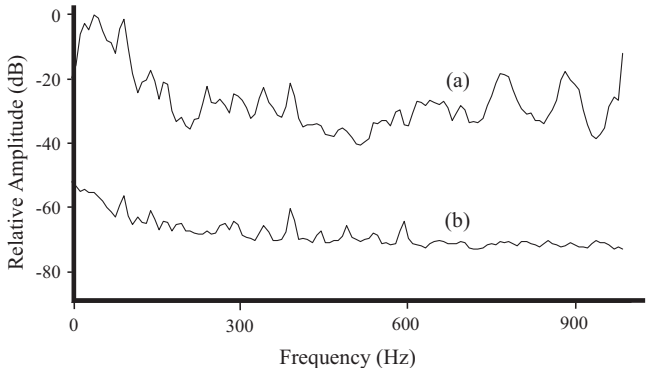


FIG. 1. Comparison between spectra of laboratory background noise recorded in (**a**) a stock tank, placed on top of a thin layer (2 cm) of expanded polystyrene and in (**b**) an experimental tank, placed on top of a thick layer (50 cm) of rockwool. Amplitude levels (dB) are relative to the maximum value of the spectra. Sampling frequency 48 kHz, 2048 point FFT, filter bandwidth 15 Hz, Hamming window, and 50% overlap.

by two opaque removable partitions into three compartments: one of 50 cm in the middle and two of 35 cm.

During courtship experiments, a single male was introduced into each of the smaller lateral compartments. These compartments were provided with terracotta pots that served as refuges and prospective spawning sites. In the central compartment, six or seven females were kept permanently. Males were left visually isolated to acclimatize for a minimum of 36 h prior to the beginning of the recordings. This period was required for males to become territorial, as shown by "digging" behavior around the refuge. Before the recording period, all electrical devices were switched off, apart from the room lights. Then, one of the opaque partitions was removed, and one male had free access to the females in the central compartment. During courtship behavior, male *P. zebra* perform a number of distinct types of displays to the females, which are not always shown in a fixed order. These include the behavioral patterns dart, quiver, lead swim, and circling with the female (Baerends and Baerends van Roon, 1950; Amorim *et al.*, 2004). During the recordings, we noted which visual displays were accompanied by sound production. However, sometimes males would produce sounds without performing any behavioral display, such as during swimming or when standing still in the water column. Once recording was complete, the tested subject was weighed (wet mass, M), measured (standard length, SL) and returned to a stock tank. Only 12 males and 5 females emitted sounds suitable for analysis. Male size averaged 107.1 mm SL [±SD (range)= ± 11.8 (88.0–122.0) mm, where SD is standard deviation] and 40.5 g M [±9.7 (22.0–57.7) g], whereas females averaged 103.6 mm SL [±0.02 (100.0–106.5) mm] and 30.4 g M [±2.5 (27.4–33.5) g].

Sounds from females were recorded from female–female interactions that naturally occurred when they were in the middle compartment isolated from the males. Sounds from male–male interactions were recorded by placing another male in the middle compartment (instead of the females), and following a similar procedure to the courtship sound recordings. Agonistic encounters consisted of frontal and lateral displays and chasing, occasionally escalating to physical contact, including biting (Baerends and Baerends van Roon, 1950). During lateral displays, animals often quiver, a behavior that is similar to the courtship quiver. To avoid physical injuries, fish were separated before or at the first sign of escalation to physical contact.

Recordings lasted 10 min for female–female, 15 min for male–male, and 20 min for male–female interactions. The duration of the recording sessions was derived from preliminary observations. All individuals were identified by natural features, such as number and position of eggspots, fin length, and marks on the body and fins.

C. Sound recording and analysis

Sounds were recorded using two High Tech 94 SSQ hydrophones (sensitivity of −165 dB re 1 V μPa^{-1}, flat frequency response ±1 dB up to 6 kHz) and a Pioneer DVD Recorder DVR-3100 (±1.5 dB from 40 Hz to 2 kHz, sampled at 48 kHz, 16 bit). One hydrophone was placed above the terracotta pot, where the territorial individuals would most likely exhibit courtship or agonistic behaviors. A second hydrophone was placed in the middle of the main compartment or in the location where individuals would more actively display at each other. The use of two hydrophones improved the probability of recording sounds close to the sound producer and also provided information on the degradation of the acoustic signals with distance. Recorded sounds could be attributed to the subject males because their intensity varied with distance from the hydrophones and were consistently associated with particular courtship displays.

Sounds were analyzed with Adobe Audition 2.0. (Adobe Systems Inc., 2005) and Raven 1.2.1 for Windows (Cornell Lab of Ornithology, 2003). Only sounds that showed a clear structure and a high signal-to-noise ratio were considered. These were typically recorded at a distance of 1-2 body lengths of the focal fish. The acoustic parameters analyzed (Fig. 2) were sound duration; number of pulses in a sound; mean pulse period of the entire sound (Mean PP); and sound-peak frequency (for a description of the acoustic parameters see Amorim *et al.*, 2004). In addition, other parameters also considered included the mean pulse period of the first five pulses in a sound (Initial PP), and a second previously undetected sound-peak frequency (PF1) typically around 150 Hz, which is of higher energy than the sound peak in the 450 Hz region (PF2) described by Amorim *et al.* (2004). PF1 is easily confounded with background noise if the recording aquarium is insufficiently acoustically insulated. When comparing an uninsulated stock tank with the experimental tanks, background noises differed by approximately 50 dB at 100 Hz, i.e., around the frequency region of PF1, and by 30 dB at 450 Hz, i.e., around the frequency of PF2 (Fig. 1). Temporal features were measured from oscillograms and sound peak frequencies from power spectra based on 2048 point FFT with a Hamming window applied. Data are presented in relative units as it was not possible to measure absolute sound levels.

Statistical analyses were performed using Statistica 7.0 (StatSoft Inc., 2005). Nonparametric statistics were used whenever the assumptions for parametric tests were not met

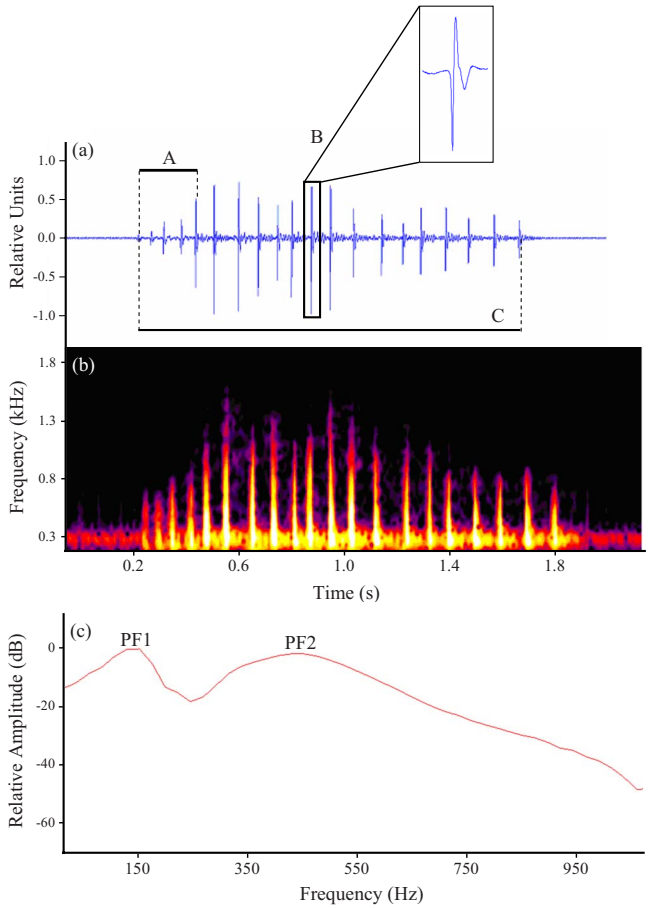


FIG. 2. (Color online) (**a**) Oscillogram, (**b**) sonogram, and (**c**) power spectrum of a *Pseudotropheus zebra* male courtship sound, representing some of the acoustic parameters measured: mean pulse period of the (**A**) first five pulses, (**B** is an example of a pulse) number of pulses, and (**C**) sound duration in (**a**) and peak frequency 1 (PF1) and 2 (PF2) in (**c**). Sampling frequency 48 kHz, 2048 point FFT, filter bandwidth 15 Hz, Hamming window, and 50% overlap.

after applying data transformations. One-way analysis of variance (ANOVA) was used to compare differences among means of the acoustic parameters of male courtship quiver sounds. The square root transformation was applied to the number of pulses to meet the ANOVA assumptions. Spearman rank correlation was used to estimate whether courtship quiver sound parameters were related to male morphological features (standard length, weight, and number of eggspots). Twelve males with an average of 17 sounds per male ($\pm$12.1 SD) were considered for these analyses.

The Kruskal–Wallis nonparametric test was used to compare the acoustic characteristics of sounds produced during different stages of courtship (lead swim, quiver, no visual display, and circle). Because few interactions proceeded to late stages of courtship, relatively few sounds were recorded during activities characteristic of such phases. The following sample sizes were considered: 36 quiver sounds from 9 males; 12 lead-swim sounds from 5 males; 10 circle sounds from 1 male; and 8 no display sounds from 4 males. Note that in this analysis, data concerning quiver sounds were restricted to 36 randomly selected sounds from 9 males from the whole data set (i.e., 4 sounds per each male), to avoid large imbalances between factor levels sample sizes. Circling sounds were extremely hard to record, not only due to the fact that this species rarely got to the ending stages of courtship during trials, but also because circling did not always occur near the hydrophone. Thus, even though a few other circling interactions were observed, it was possible in only one case to analyze their uttered sounds. Nevertheless, the comparison between circling and other sounds seemed necessary to ascertain the variability in the acoustic repertoire of this species and was included for analyses. *Post-hoc* pairwise comparisons were made with Dunn tests to determine differences between groups of courtship behaviors (Zar, 1984).

The hypothesis that sounds produced by males when courting females could be different from those produced by both sexes during agonistic interactions was also tested with one-way ANOVA. For these analyses, 198 courtship quiver sounds from the 12 analyzed males (i.e., all quiver sounds recorded during courtship interactions), as well as 124 agonistic sounds emitted by 9 males and 27 sounds produced in agonistic contexts by 5 females were considered. An average of 14 sounds ($\pm$4.4) and 5 sounds ($\pm$2.1) were considered per male and per female, respectively, in agonistic contexts. The square root transformation was applied to the number of pulses, whereas logarithmic transformations were carried out for sound duration and the mean pulse period of the first five pulses to meet the requirements of normality and homoscedasticity. Because PF2 is correlated with male SL (see Sec. III), an analysis of covariance (ANCOVA) was used to compare this frequency parameter among sexes and social context, having fish SL as a covariate to control for the effect of male size. *Post-hoc* pairwise comparisons were made with Tukey tests for unequal sample sizes.

III. RESULTS

A. Male courtship sounds

During intersexual courtship trials, female sounds were not detected. Focal male *Pseudotropheus zebra* varied in their tendency to court females. Sounds were more frequently produced when individuals from both sexes showed a greater courtship activity. Eight of the twenty males tested neither attempted courtship nor produced any sound suitable for analysis. Only four males displayed late-stage courtship behavior. Most recorded sounds (86.4%) were produced by males during quivering, the main early stage courtship behavior. Because few encounters proceeded to the late stages of courtship, such as lead swim and circle, there was a relatively scarce sample (13.2%) of sounds produced during final courtship.

Male quiver sounds had two main sound-peak frequencies at approximately 150 Hz (PF1) and 450 Hz (PF2). The mean quiver sound duration was around 700 ms with approximately 9 pulses per sound. The mean pulse period was approximately 90 ms and the initial pulse period circa 80 ms. There were significant differences between males in all sonic characteristics measured, except for sound duration (Table I). Intraindividual variation was generally high, especially for sound duration and number of pulses and lowest for PF1 and PF2, as shown by their coefficients of variation (Table I).

Larger males produced quiver courtship sounds with lower frequencies at PF2 (mass: r_S=−0.62, N=12, P

TABLE I. Characteristics of sounds produced by *P. zebra* males and females during quiver in inter- and intrasexual interactions (male–female—courtship interactions; male–male and female–female—agonistic interactions). Means, SD, and range are based on fish means. Coefficients of variation (COV = SD/mean × 100) represent intraindividual variability of the acoustic parameters. Results for one-way ANOVA testing differences between males for courtship quiver acoustic parameters, and testing differences between sounds made during different contexts and gender are presented.

	Male–female		Male–Male		Female–female		Differences between males (courtship quiver)		Differences between contexts/gender	
Sound parameters	Mean ± SD (range)	COV (%)	Mean ± SD (range)	COV (%)	Mean ± SD (range)	COV (%)	$F_{11,186}$	P	$F_{2,23}$	P
Duration (ms)	671.7 ± 135.59 (421.4–856.8)	60.37	960.5 ± 295.29 (549.1–1429.5)	69.13	524.2 ± 152.95 (358.3–732.6)	72.00	0.87	ns	8.56[c]	0.002
Number of pulses	8.6 ± 1.67 (6.6–12.4)	51.23	8.7 ± 3.48 (4.7–13.8)	52.59	4.9 ± 0.99 (3.8–6.3)	52.46	1.86[a]	0.047	6.16[a]	0.007
Mean pulse period (ms)	86.8 ± 14.37 (67.5–113.3)	22.32	125.7 ± 23.91 (90.1–160.9)	36.39	123.8 ± 27.06 (92.9–165.3)	39.83	3.69	<0.001	11.10	<0.001
Initial pulse period (ms)	76.7 ± 15.31 (52.1–103.6)	26.43	110.8 ± 27.97 (79.2–149.8)	34.64	116.7 ± 34.52 (91.7–176.4)	41.15	3.77	<0.001	8.45[c]	0.002
PF1 (Hz)	155.6 ± 26.20 (129.4–220.7)	15.17	138.0 ± 14.97 (117.2–164.1)	5.99	143.1 ± 6.72 (133.9–152.3)	8.33	10.80	<0.001	2.04	ns
PF 2 (Hz)	488.8 ± 40.84 (423.9–557.8)	8.77	462.9 ± 35.40 (433.6–550.8)	6.06	480.2 ± 29.14 (445.3–525.0)	8.05	8.13[b]	<0.001	1.80[b]	ns

Squared root transformation is applied.
Results from ANCOVA using fish SL as a covariate.
Logarithmic transformation is applied.

=0.028; length: r_s=−0.81, N=12, P=0.001). Eggspot number was not significantly related to male size (M and SL: r_s=0.50, P>0.05). Males with larger number of eggspots tended to make calls with lower PF1 frequencies (r_s=−0.82, N=9, P=0.001) and higher pulse repetition rates, i.e., shorter pulse periods (mean pulse period: r_s=−0.68, N=9, P=0.04).

The duration of sounds and their number of pulses differed according to the courtship behavior performed by the males with longer sounds containing more pulses emitted during quivering bouts (Table II, Fig. 3). The mean pulse period of the first five pulses was shorter in sounds associated with quivering than in sounds registered when males were not displaying, with lead-swim and circle sounds being intermediate (Table II, Fig. 3). The PF1 also differed significantly according to which behavior the sound was associated with (Table II), but Dunn tests were unable to distinguish any pair of behavioral categories.

B. Agonistic sounds

Sound production by males during agonistic interactions frequently occurred after a brief fight, where males would silently display frontally or laterally. Following such a contest, the dominant male (normally the resident or the larger fish) displayed laterally and quivered to the submissive male. Submissive males rapidly lost their bright colors, becoming pale. They sometimes bit at the dominant male's anal fin eggspots, in a similar manner to a female during courtship.

TABLE II. Characteristics of courtship sounds made by *P. zebra* during lead swim, quiver, with no associated display and circle. Data are pooled for all recorded individuals due to the small sample size (for quiver sounds only a subsample of 4 sounds per male was considered in the analyses—see Sec. II). Coefficients of variation are also given: COV = SD/mean × 100. Results for Kruskal–Wallis statistics testing differences between sounds associated with different courtship behaviors are presented.

	Lead swim		Quiver		No display		Circle		Kruskal–Wallis	
Sound parameters	Mean ± SD (range)	COV (%)	Mean ± SD (range)	COV (%)	Mean ± SD (range)	COV (%)	Mean ± SD (range)	COV (%)	H	P
Duration (ms)	567.3 ± 247.14 (214.0–1210.0)	43.6	1198.6 ± 647.32 (298.0–2622.0)	54.0	481.4 ± 381.16 (201.0–1276.0)	79.2	561.7 ± 157.44 (343.0–853.0)	28.0	25.06	<0.001
Number of pulses	7.0 ± 2.26 (4–12)	32.2	14.9 ± 7.78 (5–33)	52.3	5.4 ± 3.11 (3–12)	57.9	7.4 ± 1.07 (6–9)	14.5	29.07	<0.001
Mean pulse period (ms)	91.5 ± 20.35 (57.0–119.6)	22.2	86.8 ± 18.46 (60.3–132.6)	21.3	100.4 ± 13.33 (78.7–116.1)	13.3	82.6 ± 20.72 (52.7–134.7)	25.1	6.26	ns
Initial pulse period (ms)	82.7 ± 20.08 (51.0–110.3)	24.3	69.2 ± 21.67 (37.8–124.8)	31.3	90.8 ± 13.35 (70.8–114.3)	14.7	76.9 ± 24.26 (52.6–141.4)	31.6	10.73	0.01
PF 1 (Hz)	132.8 ± 15.26 (117.2–164.1)	11.5	149.5 ± 30.66 (109.4–257.8)	20.5	128.9 ± 17.72 (117.2–164.1)	13.7	140.6 ± 0.00 (140.6–140.6)	0.0	8.52	0.04
PF 2 (Hz)	459.0 ± 52.37 (375.0–539.1)	11.4	488.7 ± 62.07 (398.4–585.9)	12.7	454.1 ± 48.43 (375.0–539.1)	10.7	471.1 ± 30.15 (421.9–492.2)	6.4	1.60	ns

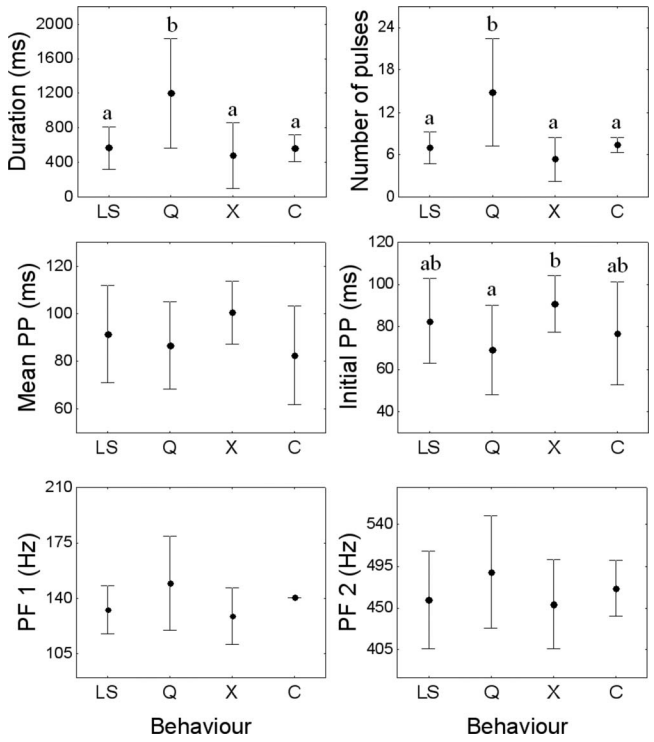


FIG. 3. Variation of courtship sound parameters in *Pseudotropheus zebra* males during lead swim (LS), quiver (Q), sounds produced with no apparent body movement (X), and circle (C). Groups that are significantly different (α=0.05) are indicated by different letters (results from Tukey tests). Both "Mean PP" and "Initial PP" refers to mean values of pulse periods; whereas the first is the mean of the pulse periods throughout the entire sound, the second indicates the mean of the first five pulses. Note that comparisons considered data pooled for all males due to the small sample sizes obtained for LS, X, and C. Only a subsample of quiver sounds was considered for the analyses (see methods).

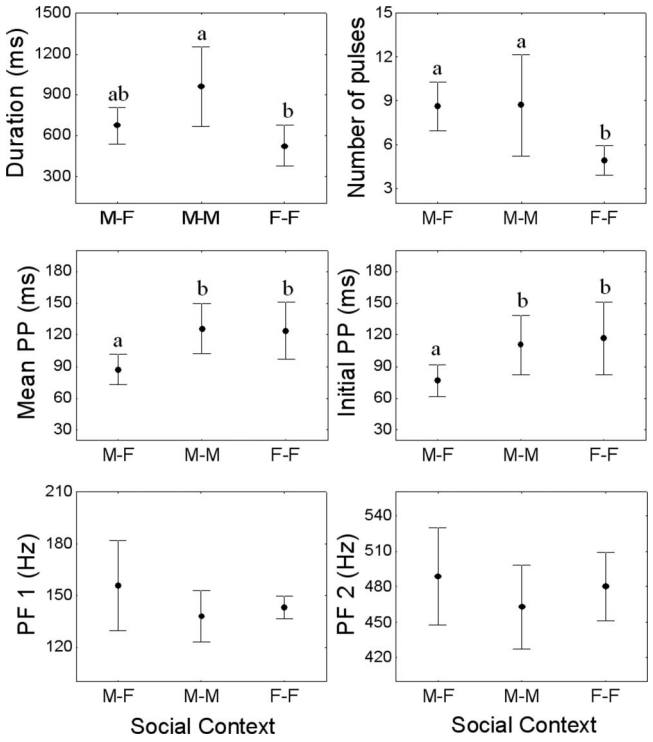


FIG. 4. Variation of the acoustic parameters of quiver sounds emitted in courtship (male–female) and agonistic interactions (male–male and female–female) by *Pseudotropheus zebra*. Groups that are significantly different (α=0.05) are indicated by different letters (results from Dunn tests).

Commonly, dominant males produced sounds during such agonistic quivering. In female-female encounters, sounds were generally produced during agonistic quivers, often by females that showed sexual readiness or during mouthbrooding, which also seemed to be more aggressive (three out of five females producing recorded sounds were mouthbrooding).

Sounds produced in male–female, male–male, and female–female encounters differed significantly in all temporal parameters but not in the frequency domain (Table I, Figs. 4 and 5). Male sounds were longer and included more pulses than those emitted by females; moreover, male sounds also differed in duration according to social context (Fig. 4). Courting male sounds also showed significantly shorter initial and mean pulse periods than agonistic sounds by either sex (Fig. 4).

IV. DISCUSSION

A. Male courtship sounds

The present study has shown that *Pseudotropheus zebra* males produce sounds not only in the early stages of courtship, during quiver, but also during courtship displays that occur closer to spawning. Moreover, the sound production in the presence of females but without any other noticeable behavioral display, consistent with observations on another Malawian haplochromine cichlid *Tramitichromis intermedius*, suggests that sound can be a purposely generated unimodal courtship display (Ripley and Lobel, 2004).

Courtship sounds varied in their characteristics according to the associated courtship behavior, being longer and with a higher pulse rate during quivering (Fig. 3). Although only a small sample size of late stage courtship sounds was

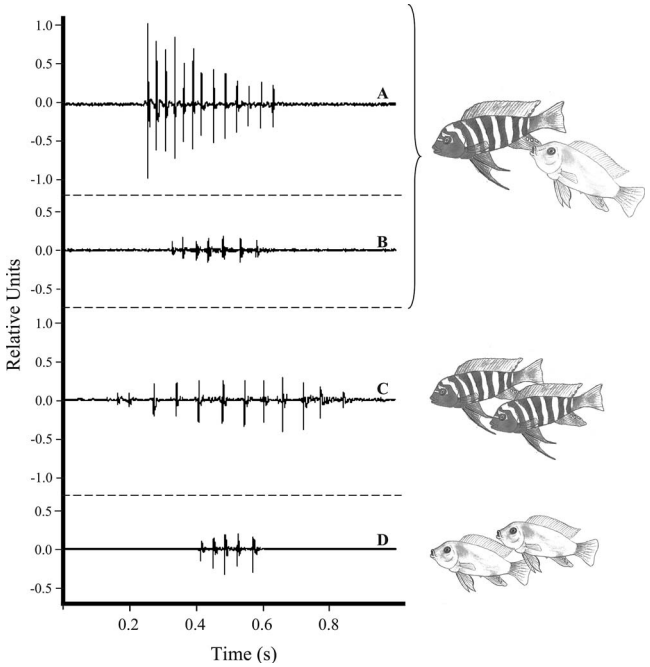


FIG. 5. Oscillograms of sounds produced associated with different contexts and gender: (**A**) male courtship quiver, (**B**) circle, (**C**) male agonistic quiver, and (**D**) female agonistic quiver. Sampling frequency 48 kHz.

recorded, the present results indicate that acoustic communication is more diversified during the courting activities than previously reported. In other cichlids, sound production seems mostly restricted to male quivering during the early stages of courtship (Ripley and Lobel, 2004, reviewed in Amorim *et al.*, 2004), except in the Mozambique tilapia *Oreochromis mossambicus* that produces sounds throughout courtship (Amorim *et al.*, 2003). In *O. mossambicus*, sounds are longer and with a faster pulse rate during the late-courtship behavior of tail wagging (Baerends and Baerends van Roon, 1950) than during other courtship activities (Amorim *et al.*, 2003). Although performed in different phases of courtship, the quivering of *P. zebra* and the tail wagging of *O. mossambicus* are probably equivalent in function. Both consist of displays in which males quiver their bodies vigorously, simultaneously emitting sounds in close proximity of the female, and may convey information of their quality and motivation. In addition, both *P. zebra* and *O. mossambicus* males may quiver and tail wag in all stages of courtship (Baerends and Baerends van Roon, 1950), particularly when females begin to wander out of a male's core spawning area.

Male motivation and quality may be advertised by higher calling rates, longer calls, and higher pulse repetition rates that are likely to be more energetically expensive. At least in some species these parameters may be assessed by females during mate choice. For example, in the gray tree frog *Hyla versicolor*, females prefer longer male calls with a higher pulse number to shorter calls (Gerhardt *et al.*, 2000), and this parameter is an indicator of male genetic quality (Welch *et al.*, 1998). In fishes, Thorson and Fine (2002) demonstrated that males *Opsanus beta* call faster at twilight, shortening and simplifying their multiboop calls, suggesting a tradeoff between call repetition rate and complexity in female choice. In invertebrates, pulse number and rate, together with sound frequency, are the most important acoustic features involved in female choice (e.g., Simmons, 1988).

Other sound parameters may transmit additional information relevant to female mate choice. Quiver sounds differed considerably among individual males, for example with larger males producing lower frequencies at PF2. This parameter was also the one that showed the least intraindividual variation (Table I), probably because it may be dependent on male size (Lobel, 2001; Amorim *et al.*, 2003) rather than motivation. Male size is often regarded as an indication of higher fitness and in cichlids may be related to social status and breeding success (e.g., Oliveira *et al.*, 1996).

The association of courtship quiver sound parameters (PF1 and pulse period) with the number of eggspots in the anal fin is less obviously explicable. Perhaps these parameters are independent indicators of some common cause, such as overall male fitness. Eggspot number was correlated with fish size in other cichlids (Goldschmidt, 1991), although it does not seem the case in *P. zebra* perhaps because of species differences or the restricted size of fish used. Females of several haplochromine cichlids are known to choose mating partners on the basis of their eggspot number (Couldridge and Alexander, 2001) and in some *Pseudotropheus* species, females prefer a larger number of eggspots (Couldridge and Alexander, 2001). In *P. zebra*, lower sound-peak frequency at PF2 and especially higher pulse rate may indicate better male condition and could be used with additional visual cues from the eggspots in mate sexual selection.

B. Agonistic sounds

Sound produced by both sexes during agonistic contexts is described in this study for *P. zebra* for the first time and has been documented for a number of cichlid species (reviewed by Lobel, 1998; Amorim, 2006). We found several significant differences in the sounds produced in agonistic context by males and females (Fig. 4). Aggressive males produced significantly longer and more pulsed sounds than females. In addition, male sounds also differed according to the social context. Courtship sounds were shorter and also had a faster pulse repetition rate than male agonistic sounds (Fig. 4). In line with our observations, in the croaking gouramis (*Trichopsis vittata*), where only females produce sounds during mating (Brittinger, 1991; Ladich, 2007), female courtship croaks are also produced at a faster rate than the aggressive croaks produced by both sexes (Brittinger, 1991). Similarly, the intervals between the double pulses that make up a croak also differ between sexes and social context (Brittinger, 1991). Although there are relatively few published quantitative comparisons of the influence of sex and social context on fish sounds, taken together, the results with *T. vittata* and the present study data with *P. zebra* suggest that temporal parameters of fish sounds may contain information on the motivation and gender of the sound producer. Sounds may carry information about male quality or motivation, which may influence the outcome of contests, in a manner similar to that proposed for female mating decisions. Playback of conspecific aggressive sounds may inhibit aggression in *Cichlasoma* (now *Archocentrus*) *centrarchus*, a Central American cichlid fish (Schwarz, 1974), whereas *Trichopsis* males that vocalized during contests had an increased chance of winning (Ladich *et al.*, 1992). In other taxa, a classical example is provided by male toads, *Bufo bufo*, that settle contests for the possession of females by signaling body size and hence fighting ability with call frequency (Davies and Halliday, 1978).

Sound production by female *P. zebra* was noted in a previous study, but not analyzed or compared with sounds made by males (Amorim *et al.*, 2004). Sound was only produced when females appeared to be sexually receptive (when the ovipositor was visible) or mouthbrooding, both situations where females typically become more aggressive. Similarly, sound production by mouthbrooding females has been documented for *O. mossambicus* (Marshall, 1971). In another cichlid fish, *A. centrarchus*, both sexes made sounds during the breeding cycle in an aggressive context (Schwarz, 1980). Females of this substrate spawning species emit sounds mainly during brood defence but also during nest preparation before spawning. It has been suggested that sound production by female fish may be more frequent than previously thought, perhaps because the sound producing apparatus is often less developed than in the male, resulting in weaker

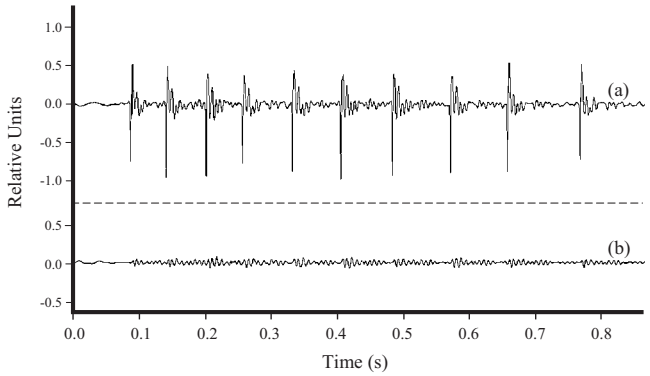


FIG. 6. Oscillograms of a courtship sound produced by a *Pseudotropheus zebra* male recorded at a distance of (a) 5 cm and (b) 40 cm from the hydrophone, in this case, sound attenuation was approximately 20 dB. Sampling frequency 48 kHz.

vocalizations, which are harder to detect (Hawkins, 1993; Ladich, 2007).

C. Concluding remarks

The variation in sounds we have documented indicates that *P. zebra* vocalizations have the potential to carry information about sex, size, motivation, and other fitness parameters that may play a role in sexual selection. Although absolute sound pressure levels have yet to be measured, it is clear that the sounds made by *P. zebra* are of low amplitudes and attenuate severely within short distances from the sender (Fig. 6), and it is unlikely that they are used to attract mating partners or to repel rivals at distance (Krebs *et al.*, 1978). More probably, and consistent with the behavioral contexts in which the sounds were observed, acoustic signals may be important during close-range encounters already initiated on the basis of visual signals. As females may reject males at this stage of a courtship sequence, and territorial rival males may decide to flee or continue fighting, close range sounds may play a major and complex role in the social behavior of *P. zebra* and other African cichlid fishes.

This study is a detailed description of sounds produced during courtship and agonistic interactions in the cichlid *Pseudotropheus zebra* and reveals an acoustic repertoire richer than previously thought. It emphasizes the need of additional research to clarify the behavioral functions of the sounds that may have also played a role in the rapid speciation of African cichlids.

ACKNOWLEDGMENTS

The authors are grateful to J. Simões senior for the logistic support, N. Wreathall for helping with the fish shipment, and all of those who helped building the tanks and maintaining the fish. This study was supported by the Natural Environment Research Council (NERC) that initially funded our collections of fish from Malawi, Program No. POCTI-ISFL-4-329/FCT, (UI&D 331/94)/FCT and a Grant No. POSI SFRH/BPD/14570/2003/FCT (MCA).

Albertson, R. C., Streelman, J. T., and Kocher, T. D. (**2003**). "Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes," Proc. Natl. Acad. Sci. U.S.A. **100**, 5252–5257.

Amorim, M. C. P. (**2006**). "Diversity of sound production in fish," *Communication in Fishes*, edited by F. Ladich, S. P. Collin, P. Moller, and B. G. Kapoor (Science Publishers, Enfield), Vol. **I**, pp. 71–104.

Amorim, M. C. P., Fonseca, P. J., and Almada, V. C. (**2003**). "Sound production during courtship and spawning of the cichlid *Oreochromis mossambicus*: male-female and male-male interactions," J. Fish Biol. **62**, 658–672.

Amorim, M. C. P., Knight, M. E., Stratoudakis, Y., and Turner, G. F. (**2004**). "Differences in sounds made by courting males of three closely related Lake Malawi cichlid species," J. Fish Biol. **65**, 1358–1371.

Baerends, G. P., and Baerends van Roon, J. M. (**1950**). "An introduction to the study of the ethology of cichlid fishes," Behaviour **1**, 1–235.

Brittinger, W. (**1991**). "The significance of the sound generating organ in the behavioural context in Trichopsis vittatus (Cuvier and Valencienne, 1831) (Belontiidae, Teleostei)," M.Sc. thesis, Univ. of Vienna, Vienna, Austria (in German).

Couldridge, V. C. K., and Alexander, G. J. (**2001**). "Color patterns and species recognition in four closely related species of Lake Malawi cichlid," Behav. Ecol. **13**, 59–64.

Davies, N. B., and Halliday, T. R. (**1978**). "Deep croaks and fighting assessment in toads, *Bufo bufo*," Nature (London) **74**, 683–685.

Genner, M. J., and Turner, G. F. (**2005**). "The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation," Fish Fish. **6**, 1–34.

Gerhardt, H. C., Tanner, S. D., Corrigan, C. M., and Walton, H. C. (**2000**). "Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*)," Behav. Ecol. **11**, 663–669.

Goldschmidt, T. (**1991**). "Egg mimics in Haplochromine cichlids (Pisces, Perciformes) from Lake Victoria," Ethology **88**, 177–190.

Hawkins, A. (**1993**). "Underwater sound and fish behaviour," *Behavior of Teleost Fishes*, 2nd ed., edited by T. Pitcher (Chapman and Hall, London), pp. 129–169.

Krebs, J., Ashcroft, R., and Webber, M. (**1978**). "Song repertoires and territory defence in the great tit," Nature (London) **271**, 539–542.

Ladich, F., Brittinger, W., and Kratochvil, H. (**1992**). "Significance of agonistic vocalization in the croaking gourami (Trichopsis vittatus, Teleostei)," Ethology **90**, 307–314.

Ladich, F. (**2007**). "Females whisper briefly during sex: context- and sex-specific differences in sounds made by croaking gouramis (Teleosts)," Anim. Behav. **73**, 379–387.

Lobel, P. S. (**1998**). "Possible species specific courtship sounds by two sympatric cichlid fishes in Lake Malawi, Africa," Environ. Biol. Fishes **52**, 443–452.

Lobel, P. S. (**2001**). "Acoustic behavior of cichlid fishes," J. Aquaricult. Aquat. Sci. **9**, 167–186.

Marshall, J. A. (**1971**). "Sound production by *Tilapia mossambica* (Pisces: Cichlidae)," Am. Zool. **11**, 632.

Oliveira, R. F., Almada, V. C., and Canário, A. V. M. (**1996**). "Social modulation of sex steroid concentrations in the urine of male cichlid fish *Oreochromis mossambicus* (Teleostei: Cichlidae)," Horm. Behav. **30**, 2–12.

Ripley, J. L., and Lobel, P. S. (**2004**). "Correlation of acoustic and visual communication in the Lake Malawi cichlid *Tramitichromis intermedius*," Environ. Biol. Fishes **71**, 389–394.

Schwarz, A. (**1974**). "Sound production and associated behavior in a cichlid, *Cichlasoma centrarchus*," Z. Tierpsychol **35**, 147–156.

Schwarz, A. L. (**1980**). "Sound production and associated behavior in a cichlid fish, *Cichlasoma centrarchus*. II. Breeding pairs," Environ. Biol. Fishes **5**, 335–342.

Simmons, L. W. (**1988**). "The calling song of the field cricket, *Gryllus bimaculatus* (De Geer): constraints on transmission and its role in intermale competition and females choice," Anim. Behav. **36**, 380–394.

Thorson, R. F., and Fine, M. L. (**2002**). "Crepuscular changes in emission rate and parameters of the boatwhistle advertisement call of the gulf toadfish, *Opsanus beta*," Environ. Biol. Fishes **63**, 321–331.

Turner, G. F. (**1999**). "Explosive speciation of African cichlid fishes," *Evolution of Biological Diversity*, edited by A. E. Magurran and R. M. May (Oxford University Press, Oxford), pp. 217–229.

Welch, A. M., Semlitsch, R. D., and Gerhardt, H. C. (**1998**). "Call duration as a reliable indicator of genetic quality in the gray tree frog," Science **280**, 1928–1930.

Zar, J. H. (**1984**). *Biostatistical Analysis*, 2nd ed. (Prentice–Hall, Englewood cliffs, NJ).

Low frequency vocalizations attributed to sei whales (*Balaenoptera borealis*)

Mark F. Baumgartner
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Sofie M. Van Parijs and Frederick W. Wenzel
Northeast Fisheries Science Center, 166 Water Street, Woods Hole, Massachusetts 02543

Christopher J. Tremblay
Bioacoustics Research Program, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850

H. Carter Esch
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Ann M. Warde
Bioacoustics Research Program, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850

(Received 15 February 2008; revised 2 May 2008; accepted 19 May 2008)

Low frequency (<100 Hz) downsweep vocalizations were repeatedly recorded from ocean gliders east of Cape Cod, MA in May 2005. To identify the species responsible for this call, arrays of acoustic recorders were deployed in this same area during 2006 and 2007. 70 h of collocated visual observations at the center of each array were used to compare the localized occurrence of this call to the occurrence of three baleen whale species: right, humpback, and sei whales. The low frequency call was significantly associated only with the occurrence of sei whales. On average, the call swept from 82 to 34 Hz over 1.4 s and was most often produced as a single call, although pairs and (more rarely) triplets were occasionally detected. Individual calls comprising the pairs were localized to within tens of meters of one another and were more similar to one another than to contemporaneous calls by other whales, suggesting that paired calls may be produced by the same animal. A synthetic kernel was developed to facilitate automatic detection of this call using spectrogram-correlation methods. The optimal kernel missed 14% of calls, and of all the calls that were automatically detected, 15% were false positives.
[DOI: 10.1121/1.2945155]

PACS number(s): 43.80.Ka [WWA] Pages: 1339–1349

I. INTRODUCTION

Passive acoustic monitoring has matured into a powerful tool for both research and conservation by allowing persistent observations of marine mammal occurrence over larger spatial scales and longer time scales than previously possible with traditional visual assessment methods. Recordings of baleen whale vocalizations have been used to assess abundance (George *et al.*, 2004), seasonal occurrence (Stafford *et al.*, 2001; Heimlich *et al.*, 2005; Mellinger *et al.*, 2007), distribution (Stafford *et al.*, 2001; Heimlich *et al.*, 2005), and behavior (Croll *et al.*, 2002; Darling *et al.*, 2006; Oleson *et al.*, 2007; Stimpert *et al.*, 2007). These studies rely on very fundamental information about which species produce particular calls. Remarkably, many calls produced by marine mammals have yet to be described and attributed to an individual species, likely because systematic collection of acoustic and visual observations to confirm the species of calling whales is uncommon, and tagging studies that use acoustic recording instrumentation have been limited to a few species (Matthews *et al.*, 2001; Madsen *et al.*, 2002; Zimmer *et al.*, 2005; Johnson *et al.*, 2006; Oleson *et al.*, 2007; Stimpert *et al.*, 2007).

Sei whales (*Balaenoptera borealis*) are found primarily in the temperate oceans of both the northern and southern hemispheres, and apparently migrate between lower latitude winter breeding grounds and higher latitude summer feeding grounds (Mizroch *et al.*, 1984; Perry *et al.*, 1999). They feed primarily on aggregations of copepods, euphausiids, and small schooling fish by filtering these prey through their baleen (Hjort and Ruud, 1929; Kawamura 1974; Flinn *et al.*, 2002). The acoustic behavior of sei whales, like most aspects of their behavior and ecology, is quite poorly described. Only four reports of sei whale calls are currently available. Thompson *et al.* (1979) described recordings of sei whales obtained in the waters between Nova Scotia and Newfoundland, Canada, as 0.7 s long bursts of seven to ten metallic pulses with peak energy at 3 kHz. Knowlton *et al.* (1991) described similar 1.4–2.6 s midfrequency vocalizations recorded in waters off southwestern Nova Scotia, Canada that consisted of two bouts of 10–20 frequency-modulated (FM) 1.5–3.5 kHz sweeps separated by 0.4–1 s. In the Southern Ocean near the Antarctic Peninsula, McDonald *et al.* (2005) recorded a number of tonal, FM, and broadband calls between 200 and 700 Hz in proximity to sei whales. The estimated source level of these calls was relatively low for baleen whales (156 dB with regard to 1 μPa at 1 m), and

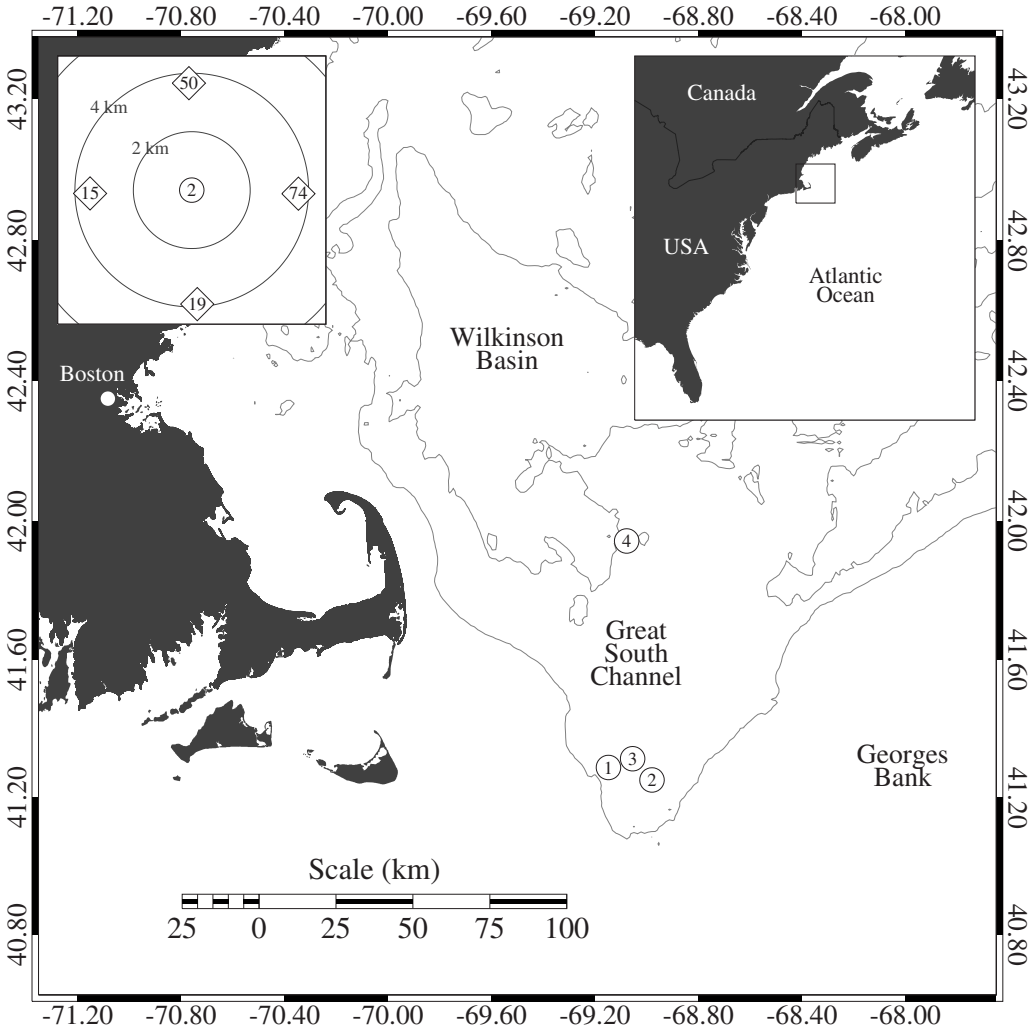


FIG. 1. Locations of the four anchor stations in the Great South Channel. The inset at the upper left depicts the configuration of MARU Nos. 15, 19, 50, and 74 around anchor station 2.

McDonald *et al.* (2005) suggested the calls were likely used for communication over short distances (a few kilometers) to facilitate feeding or social interactions with nearby conspecifics. Finally, Rankin and Barlow (2007) recorded two low frequency calls near sei whales just north of the Hawaiian Islands: a FM sweep from 100 to 44 Hz lasting 1.0 s, and a lower frequency FM sweep from 39 to 21 Hz lasting 1.3 s.

This paper describes a low frequency downsweep call attributed to sei whales in the northwestern Atlantic Ocean that is similar to the 100–44 Hz downsweep call recorded by Rankin and Barlow (2007) in the Pacific Ocean. We initially detected this call in acoustic recordings collected by autonomous ocean gliders off the coast of Cape Cod, MA during May 2005 (Fratantoni and Baumgartner, 2005; Baumgartner *et al.*, 2006; Baumgartner and Fratantoni, in press), and subsequently designed the current study to identify the species producing the call. Acoustic data were collected from an array of recorders deployed in an area frequented by right (*Eubalaena glacialis*), sei, and humpback (*Megaptera novaeangliae*) whales during the spring. Species confirmation of the downsweep call was accomplished by comparing the occurrence of these species to the occurrence of localized calls during 70 h of collocated visual and acoustic observations. Finally, a synthetic kernel (a mathematical representation of a call in frequency-time space) was developed to aid in the automated detection of the downsweep call in future studies via spectrogram cross correlation.

II. METHODS

A. Acoustic and visual observations

Collocated visual and acoustic observations were collected on four separate occasions during the spring seasons of 2006 and 2007 in the Great South Channel between Cape Cod, MA and Georges Bank (Fig. 1; Table I). For each study, observations were collected in the vicinity of a fixed geographic location called an anchor station. Initial visual surveys were conducted prior to each study to find an area of high baleen whale abundance; the anchor station was then established in this area. The primary focus of these surveys was to study the ecology of North Atlantic right whales; therefore, areas with high abundances of right whales were preferentially sought.

Acoustic recordings were collected with recoverable marine autonomous recording units (MARUs), moored instruments designed by and leased from Cornell University

TABLE I. Summary of each anchor station study, including starting date and time (local time), duration of recorder deployments, time that the anchor station was occupied by the NOAA Ship *Albatross IV* (time in parentheses indicates the duration of visual effort during daylight hours), and water depth at the anchor station.

Anchor station	Start date/time	Recorder deployments (h)	Station occupied (h)	Water depth (m)
1	5/7/06 13:30	25.5	21.0 (15.5)	103
2	5/23/06 15:30	39.0	34.5 (18.5)	137
3	5/21/07 19:00	41.5	37.5 (17.0)	160
4	6/6/07 20:00	48.0	35.5 (19.0)	192

Laboratory of Ornithology's Bioacoustics Research Program. Each MARU consists of a digital audio recorder, hard drive, and batteries encased within an 18 in. glass sphere that is positively buoyant, vacuum sealed, and rated to a depth of 6700 m. Raw audio is captured with an HTI-94-SSQ series hydrophone (2 Hz–30 kHz frequency response) and internal preamplifier (combined maximum sensitivity of −165 dB with regard to 1 V/μPa) mounted outside the plastic "hard hat" that protects the glass sphere. The MARUs were programed to sample at 10 kHz for our study, and the resulting digital audio data were low-pass filtered and decimated to 2048 Hz to facilitate analysis of low frequency baleen whale calls. Prior to the start of each anchor station study, four MARUs were deployed from the NOAA research vessel *Albatross IV* in a diamond configuration approximately 3.7 km (2 nautical miles) away from the central anchor station (Fig. 1). The MARUs were moored with sandbags so that they floated 1.5–2 m above the seafloor. The buoys were recovered 25.5–48 h after deployment (Table I) by acoustically triggering the MARU's release mechanism.

After the MARUs were deployed, the R/V *Albatross IV* continuously occupied the central anchor station for a period between 21.0 and 37.5 h (Table I) to systematically estimate whale abundance and collect continuous environmental observations. The ship was not actually anchored at the station; instead, the ship was moved to the station immediately prior to the start of each half-hourly observation period (i.e., at the top and bottom of the hour), and then allowed to drift off the station during the ensuing 15–20 min (drift from the station was typically ~500 m). During daylight hours, rotating teams of observers noted the location of all cetaceans within visual range. Three observers cooperatively scanned 360° around the ship using the naked eye and handheld 7×50 binoculars for 15 min every half hour (i.e., from 0 to 15 min past the hour and from 30 to 45 min past the hour) and for each group of cetaceans observed, recorded (1) the species, (2) group size, (3) distance and relative bearing from the ship to the group, and (4) behavioral observations. Care was taken to avoid recounting individuals or groups during a 15 min observing period. Oceanographic measurements and zooplankton abundance were collected at half-hourly intervals around the clock at the anchor station using an instrument profiler consisting of a conductivity-temperature-depth instrument, a fluorometer, a video plankton recorder, and an optical plankton counter; however, those data were collected for a different study and are not presented here.

To confirm the identity of the species producing low frequency downsweep calls, the occurrence of these calls was compared to the occurrence of the most abundant baleen whale species observed during the anchor station studies (right, sei, and humpback whales). All calls were localized (see below) so that only calls within 3 km of the anchor station were compared to the sighting data (the visual detection range from the ship was approximately 3 km for accurate species identification). Since the downsweep calls were not particularly numerous within this 3 km radius of the ship, a sampling unit was defined as 1 h of collocated visual and acoustic observations. The presence of whales was therefore noted for each sampling unit by combining the results of two successive observing periods (e.g., humpback whale presence would be noted for the sampling unit starting at 13:52:30 and ending at 14:52:30 if one or more humpback whales were sighted during either the 14:00–14:15 or the 14:30–14:45 observing periods). Sampling units with only one observing period (near dawn or dusk) were used in the analysis only if whales were present during that observing period. The presence of calls was also noted for each sampling unit, and a two-way contingency table was constructed for each species to compare the presence of whales to the presence of calls. The Cramér, or phi, coefficient (Sokal and Rohlf, 1995) was used to determine the degree of the association between the occurrence of whales and calls. This statistic varies from −1 to 1 to indicate the extent of negative or positive association, and is analogous to the correlation coefficient. Finally, a one-way Fisher's exact test was used to examine the null hypothesis that call occurrence was independent of whale presence against an alternative hypothesis that call occurrence was positively associated with whale presence.

B. Detector development

To facilitate automated detection of the downsweep call, we built a synthetic kernel that could be cross correlated with spectrograms of audio data to reliably identify call times. The methods to construct and ultimately use the kernel were based on the work of Mellinger and Clark (2000). Exemplars of the downsweep call with a high signal to noise ratio (n =60) were extracted from acoustic recordings collected in the Great South Channel by autonomous underwater vehicles (ocean gliders) in 2005 (Fratantoni and Baumgartner, 2005; Baumgartner *et al.* 2006; Baumgartner and Fratantoni, in press). Spectrograms of each exemplar were normalized, synchronized, and then averaged to produce a single representation of the downsweep call [Fig. 2(a)]. A family of detector kernels was produced from this average call and then evaluated to determine which kernel had the best performance characteristics.

Individual kernels were constructed using different amplitude contours of the averaged call (Fig. 2). For time t in the spectrogram, the start and end frequencies, f_0 and f_1, of the amplitude contour were determined. From these, two parameters were defined:

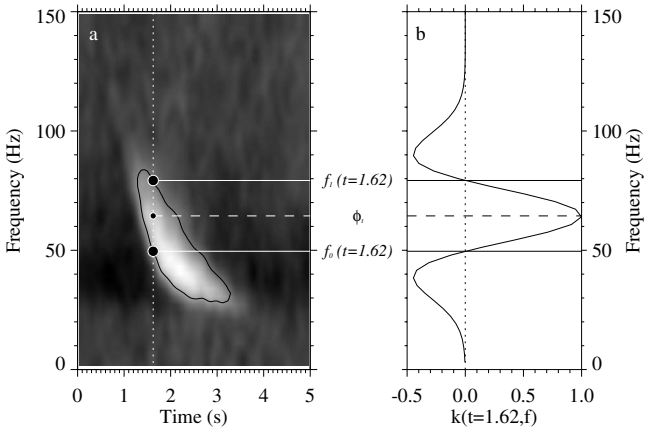


FIG. 2. (a) Spectrogram representing the average of 60 downsweep call exemplars. A single contour is shown to illustrate the construction of a kernel detection function. The large filled black circles indicate the start (f_0) and end (f_1) frequencies of the contour at time $t=1.62$ s, and the small filled black circle indicates the midpoint of these frequencies (ϕ_t). (b) The kernel detection function at $t=1.62$ s with f_0 and f_1 indicated as solid horizontal lines, and ϕ_t indicated as a dashed horizontal line.

$$\phi_t = \tfrac{1}{2}[f_0(t) + f_1(t)] \tag{1}$$

and

$$\sigma_t = \tfrac{1}{2}[f_1(t) - f_0(t)]. \tag{2}$$

The kernel was then defined [after Mellinger and Clark (2000)] as

$$k(t,f) = \left(1 - \frac{(f-\phi_t)^2}{\sigma_t^2}\right) e^{-(f-\phi_t)^2/(2\sigma_t^2)}. \tag{3}$$

This approach to kernel construction is quite similar to that described by Mellinger and Clark (2000); however, they used linear sections to approximate the shape of a call (i.e., successive values of ϕ_t fall along a line in frequency-time space) whereas our approach attempts to incorporate any nonlinearity present in the call shape (i.e., ϕ_t is allowed to vary freely according to the average call contours).

Putative vocalizations were detected by first creating spectrograms of the audio data using short-time Fourier transforms for each 640 sample frame (0.3125 s) with 80% overlap and a Hann window. Continuous quasitonal noise (e.g., ship noise) was removed using a 10 s median filter on each frequency band in the spectrogram. The synthetic kernel was then cross correlated with the spectrogram and the resulting time series of correlation coefficients was considered a detection function (Fig. 3). Mellinger and Clark (2000) used a spectrogram covariance approach; however, the spectrogram correlation method was used in this study because it accounts for changes in gain between different types of instruments (e.g., between MARUs and glider recorders). The spectrogram correlation is equivalent to the spectrogram covariance normalized by the variability in the audio data (which, in turn, is a function of gain). Peaks in the detection function indicate times of putative calls, and the amplitude of the peaks indicates the amount of agreement between the putative call and the kernel. Automated detection is accomplished by isolating only those peaks that exceed a predetermined threshold.

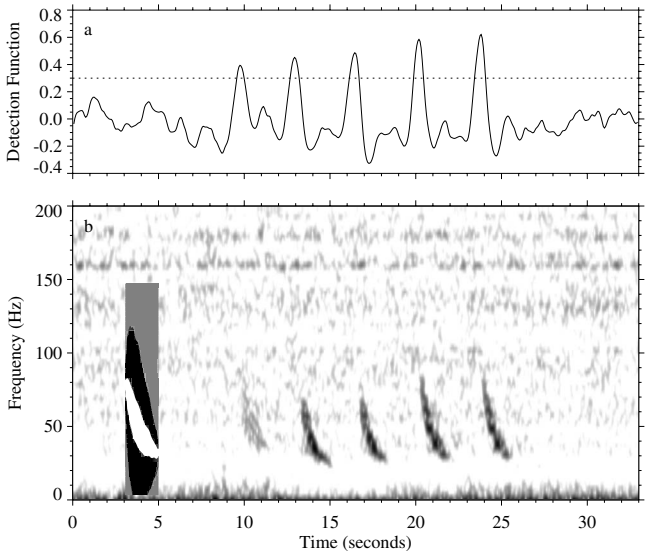


FIG. 3. (a) An example of a detection function (spectrogram-kernel correlation coefficient) derived from the spectrogram in (b). The synthetic kernel that was used to derive the detection function in (a) is shown at the left in (b) (shading indicates the sign of the kernel: white=positive, black =negative, gray=0). If a detection threshold of 0.3 was used to isolate calls [dotted line in (a)], then five peaks are present corresponding to the calls in (b). These calls consist of an initial faint call followed by two pairs of calls generated by different animals to the northeast (faint call), south (first pair), and the southeast (second pair) of the MARU array.

The performance of all the kernels was evaluated by first detecting all downsweep calls in the acoustic record of one of the buoys deployed during the study at anchor station 1 (buoy No. 15). This was accomplished by both manual review and an assisted review with a preliminary kernel detector and a very low detection threshold. The assisted review was necessary because the independent manual review missed a substantial number of calls (see Sec. III). Two experienced reviewers independently scanned spectrograms of the audio data, isolated times of low frequency downsweep calls, and confirmed the calls aurally. The independent reviews were combined to produce a single dataset of confirmed downsweep calls. The assisted review automatically identified times of potential downsweep calls using a preliminary kernel, and then each call was evaluated by a third independent reviewer. Calls from the manual review and from the assisted review were compared to evaluate the efficiency of these two approaches. Finally, all confirmed calls were used to evaluate the family of kernel detectors and to select the one with the best performance characteristics. Detection performance was evaluated for each kernel by comparing the percentage of false detections to the percentage of missed calls (traditional receiver-operator characteristic curves cannot be constructed with these data because true negatives cannot be enumerated).

C. Localization

The positions of vocalizing whales were estimated from the differences in arrival times of calls at the four MARUs deployed around each anchor station. Immediately before deployment and again after recovery of the MARUs, an impulsive sound (e.g., banging a pipe with a wrench) was simul-

taneously recorded by each of the instruments; during postprocessing, each recorder was synchronized to a common time base using these impulsive sounds. Differences in arrival times were estimated by spectrogram cross correlation. Briefly, a detected call in the spectrogram for one buoy was cross correlated with the spectrograms of the other buoys within ±10 s of the detection time. Peaks in the cross-correlation functions indicated the possible presence of the same call in the other buoy recordings, and manual verification was necessary to exclude spurious matches. The lead or lag time of the cross correlation is an estimate of the time difference of arrival. Locations could be estimated when calls were detected on three or four of the recorders using equations adapted from Watkins and Schevill (1971). In cases where calls could be detected on all four of the recorders, an iterative refinement technique was also used to minimize measurement errors in the time differences of arrival [after Foy (1976) and Freitag *et al.* 2001]. Baumgartner *et al.* (in press) described the details of this localization and iterative refinement approach in the context of tracking baleen whales tagged with an acoustic transmitter. The exact same localization methods described by Baumgartner *et al.* (in press) are used here with the assumption that the whale vocalizes near the surface (Matthews *et al.*, 2001; Oleson *et al.*, 2007).

III. RESULTS

A. Call description

Only high-quality (i.e., high signal to noise ratio), low frequency downsweep calls were analyzed for call characteristics ($n=108$ calls produced within 2 km of each MARU during all anchor station studies). On average, these calls swept from a starting frequency of 82.3 Hz (SD=15.2 Hz) to an ending frequency of 34.0 Hz (SD=6.2 Hz) over 1.38 s (SD=0.37 s) (Fig. 4). Calls farther away from the recorders tended to have lower starting frequencies (correlation between distance from recorder and starting frequency: $r=-0.375$, $p<0.0001$) and were shorter in duration (correlation between distance from recorder and call duration: $r=-0.260$, $p=0.007$), which is consistent with attenuation of the higher frequencies at the beginning of the call for distant vocalizations. Variability in call characteristics was largely driven by changes in the start frequency; the ending frequency was considerably less variable, and duration, slope (end minus start frequency divided by duration), and sweep (start minus end frequency) covaried significantly with start frequency (duration: $r=0.434$, $p<0.001$, slope: $r=-0.622$, $p<0.001$; sweep: $r=0.916$, $p<0.001$).

Calls occasionally occurred in pairs [Fig. 4(b)], and for those pairs localized inside the array, the average distance between position estimates for each individual call comprising the pair was 66 m (SD=47 m, $n=14$ pairs localized within 3.7 km of the anchor station during all anchor station studies). Although positional error in the localization procedure was not directly evaluated, it was likely on the order of several tens of meters. Similarity between the first and last calls in the pairs was compared to the similarity between the first call in each pair and contemporaneous calls made by other whales in the area; similarity was measured as the correlation between call spectrograms. Only contemporaneous calls that could not have been produced by the whale making the first call in a pair were used (i.e., only calls produced sufficiently far apart that a whale would have to travel at an unrealistic speed of over 15 m s^{-1} to produce both the paired call and the contemporaneous call). Each of the 14 call pairs detected within 3.7 km of the anchor stations had between zero and ten contemporaneous calls available for comparison (one pair was omitted because of a low signal to noise ratio). Of the 40 resulting comparisons, we found only one case in which the correlation coefficient between the first call in a pair and a contemporaneous call was significantly higher than the corresponding correlation coefficient between the two calls comprising that pair ($p<0.05$, one-sided z-test for comparing correlation coefficients, no adjustment for multiple testing). These results indicate that the calls comprising the pairs were more similar to one another than to calls produced by other whales nearby. Based on this similarity and the short distance between calls comprising the pairs, we presume that call pairs are produced by the same individual whale. The average interval between individual calls in a pair, measured from the start of the first call to the start of the second call, was 3.5 s (SD=0.36, $n=112$ pairs localized inside or outside the array during all anchor station studies). On rare occasions, triplets were recorded [Fig. 4(c)] with similar intercall intervals as the pairs.

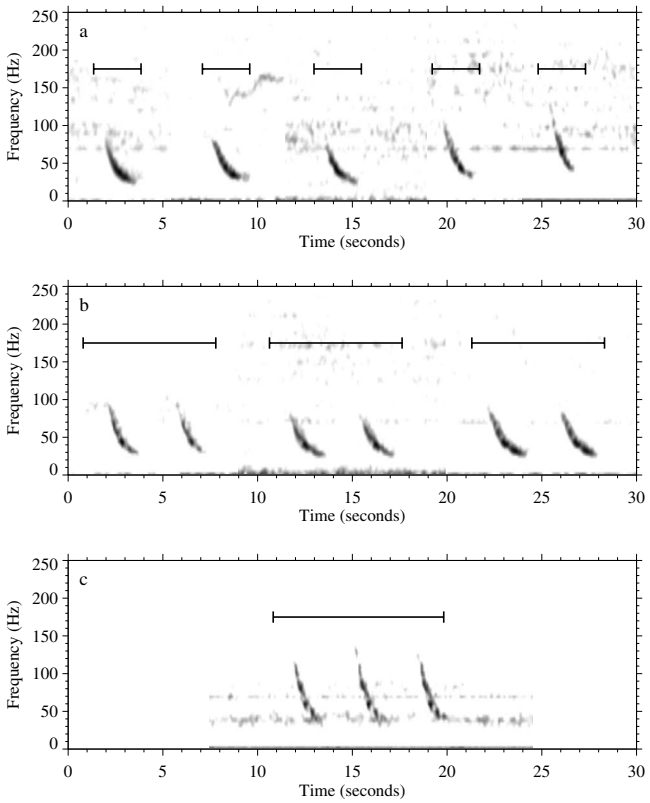


FIG. 4. Examples of low frequency downsweep calls occurring in (a) singles, (b) pairs, and (c) triplets. The bars indicate individual calls or sets of calls. Each single in (a) and pair in (b) were recorded at different times on different recorders and concatenated here (i.e., calls were not recorded contiguously as shown).

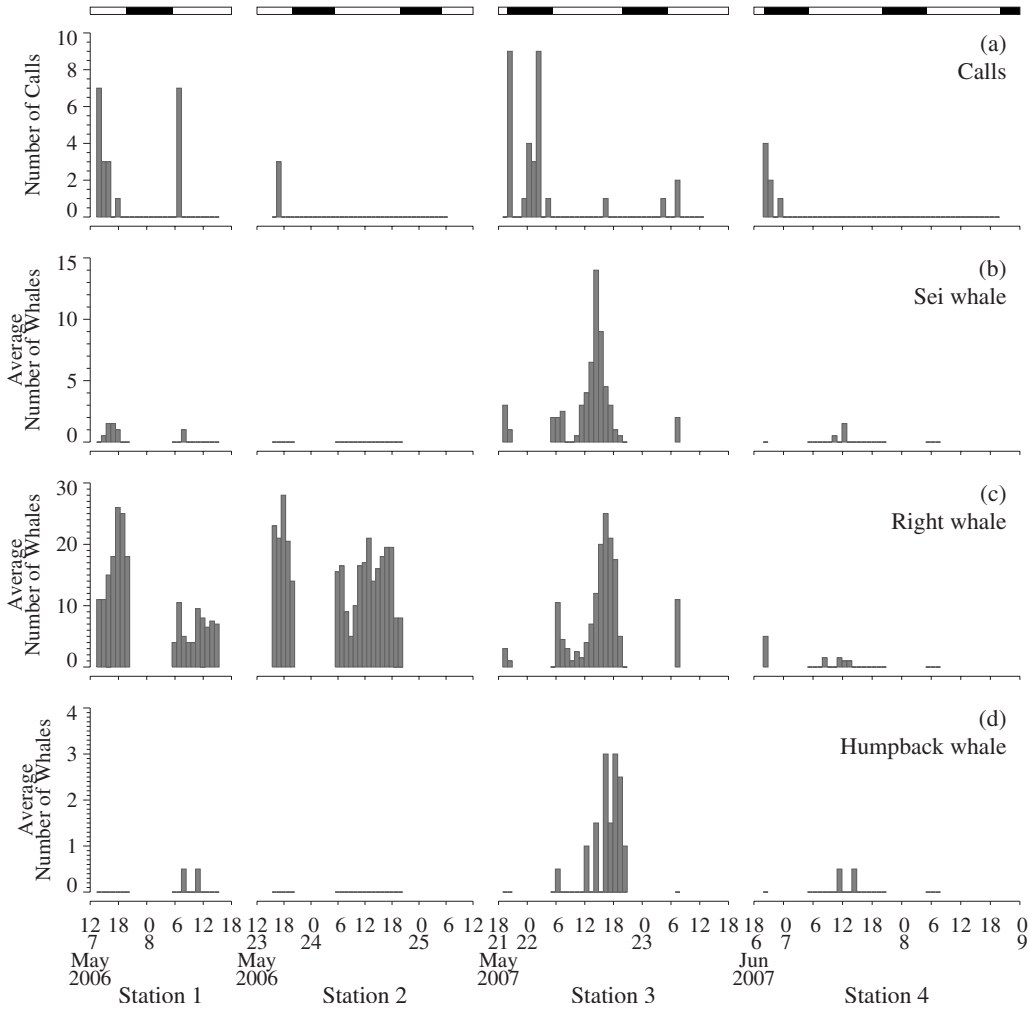


FIG. 5. Hourly time series of (a) downsweep calls detected and localized within 3 km of the anchor station and [(b)–(d)] average whale abundance for (b) sei, (c) right, and (d) humpback whales within 3 km of the anchor station during each study. The white and black bars at the top of (a) indicate periods of day and night, respectively. Each hour of visual and acoustic efforts is depicted as vertical bars (number of calls or whales >0) or horizontal lines (number of calls or whales=0). Abscissa is shown in local time.

B. Species identification

Low frequency downsweep calls were detected and localized within 3 km of the anchor station during every study period (Fig. 5). Calls tended to be clustered into periods of high calling rates separated by long periods of silence, which is consistent with whales moving in and out of the 3 km radius around the station as well as possible diel periodicity. Whale abundance derived from the visual sightings was highly variable. Right whales were nearly always present at anchor stations 1–3, but not at station 4. The continuous presence of right whales was not particularly surprising considering that the station locations were chosen primarily based on the initial high abundance of right whales. A large multispecies aggregation of right, sei, and humpback whales was encountered at station 3 on the afternoon of May 22, 2007, but sei and humpback whale abundance was otherwise generally low at the stations. Fin whales (*Balaenoptera physalus*) were sighted at the anchor stations occupied during 2007 only (including in the May 22 multispecies aggregation), but sightings were uncommon and abundance was very low. The coefficient of association between call occurrence and whale presence was highest for sei whales (ϕ =0.315; Table II) and the null hypothesis of independence was rejected only for sei whales (Fisher's exact test; p =0.016; Table II), which strongly suggest that sei whales produce the low frequency downsweep call.

C. Detector performance

Manual review (visual and aural) of the recorder data from one of the MARUs deployed near anchor station 1 yielded 302 low frequency sei whale calls over the 25.5 h deployment. During the assisted review, all but 5 of these 302 calls (98.3%) were detected with the preliminary kernel detector and confirmed to be low frequency sei whale calls. The preliminary kernel detector also isolated 620 additional vocalizations confirmed to be low frequency sei whale calls that were not detected during manual review of the recorder data. The use of the preliminary synthetic kernel allowed a substantial increase in sensitivity by isolating calls that were nearly impossible to recognize visually in a spectrogram, yet were clearly detectable by ear.

The performance of kernels derived from separate amplitude contours of the averaged exemplar calls was evaluated by comparing detections for a range of detection thresh-

TABLE II. Species-specific, two-way contingency tables comparing hourly occurrence of whales from visual observations (present or absent) to occurrence of calls from acoustic observations (call or no call) within 3 km of the anchor stations. The accompanying coefficient of association (ϕ, Sokal and Rohlf, 1995) and p-value (p) for a one-tailed Fisher's exact test of independence are provided. Only those incomplete visual survey units with whales present are included in these tallies; therefore, the total number of hourly survey units varies between species (see text).

	Call	No call	Total	ϕ	p
Humpback whale					
Present	1	11	12		
Absent	5	48	53		
Total	6	59	65	−0.015	0.722
Right whale					
Present	10	49	59		
Absent	0	15	15		
Total	10	64	74	0.199	0.087
Sei whale					
Present	6	17	23		
Absent	2	42	44		
Total	8	59	67	0.315	0.016

olds with the more than 900 confirmed calls identified by the assisted review (Fig. 6). The largest kernels (both in duration and bandwidth) performed similarly, but performance generally decreased with kernel size. Detection performance was quite good for the detector that minimized both missed calls and false detections (Fig. 7); for a threshold of 0.25, the percentage of missed calls and false detections were 13.5% and 15.3%, respectively, while only two false detections were encountered for every ten true detections (Table III).

IV. DISCUSSION

Using systematic collocated visual and acoustic observations, we documented a low frequency downsweep call previously unreported in the Atlantic Ocean and attributed these calls to sei whales. This call was by no means uncommon; during the first anchor station study, the call was recorded on one MARU an average of 37 times/h. Baumgartner and Fratantoni (in press) observed calling rates of over 500 per hour from ocean glider deployments in the Great South Channel during May 2005. This call is present in other recent acoustic recordings from the Gulf of Maine (Van Parijs unpublished data), New England Shelf (D. Fratantoni, personal communication), mid-Atlantic Bight (J. Lynch, personal communication), and in Davis Strait between Greenland and Canada (K. Stafford, personal communication). The call is similar to a low frequency call attributed to sei whales in the Pacific Ocean by Rankin and Barlow (2007). The Pacific recordings were collected at low latitudes during late

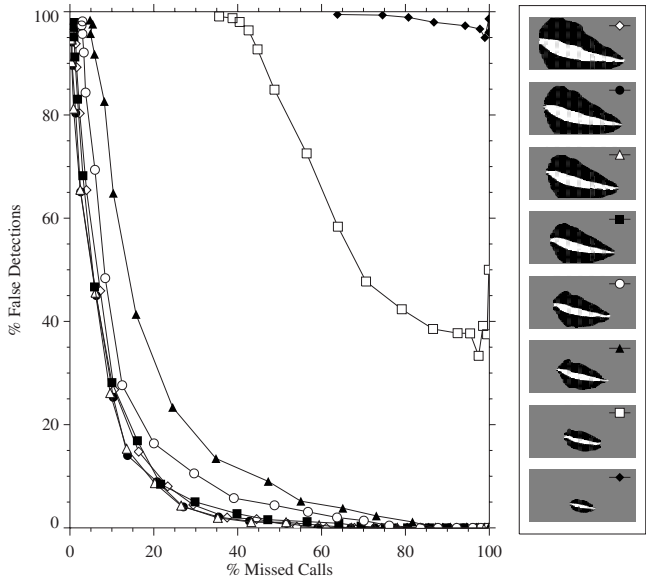


FIG. 6. Performance of each kernel detector measured by the percentages of missed calls and false detections. Kernels that minimize both missed calls and false detections (i.e., pass closest to the origin) are considered optimal for most applications. Representations of each kernel are shown at the right in frequency-time space as in Fig. 7 (the x-axis represents time ranging from 0 to 3 s, the y-axis represents frequency ranging from 0 to 150 Hz, and the shading indicates the sign of the kernel: white=positive, black=negative, gray=0).

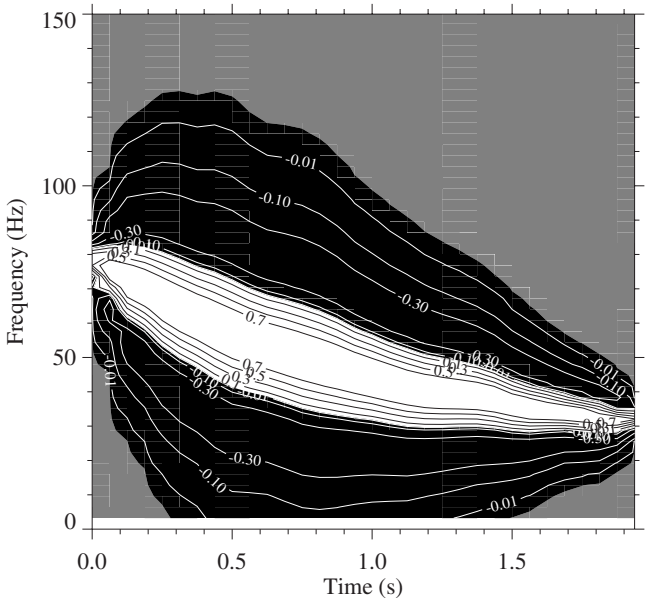


FIG. 7. Synthetic kernel with the best performance characteristics (depicted with open triangles in Fig. 6) for detecting sei whale, low frequency downsweep calls.

TABLE III. Performance results for the optimal kernel detector (depicted with open triangles in Fig. 6). True positive indicates the number of correctly detected calls, false negative indicates the number of calls missed by the detector, and false positive indicates the number of false detections.

Detection threshold	True positive	False negative	False positive	Total detections	Missed calls (%)	False detections (%)	False:True detections
0.05	933	6	38 739	39 672	0.6	97.6	41.5
0.10	935	4	16 415	17 350	0.4	94.6	17.6
0.15	931	8	4 028	4 959	0.9	81.2	4.3
0.20	882	57	738	1 620	6.1	45.6	0.8
0.25	812	127	147	959	13.5	15.3	0.2
0.30	690	249	31	721	26.5	4.3	0.0
0.35	533	406	6	539	43.2	1.1	0.0
0.40	381	558	2	383	59.4	0.5	0.0
0.45	259	680	0	259	72.4	0.0	0.0
0.50	155	784	0	155	83.5	0.0	0.0
0.55	87	852	0	87	90.7	0.0	0.0
0.60	24	915	0	24	97.4	0.0	0.0
0.65	1	938	0	1	99.9	0.0	0.0

fall, and when comparing these calls to previously reported midfrequency calls recorded during summer in the Atlantic and the Antarctic, Rankin and Barlow (2007) speculated that the low frequency downsweeps may be geographically distinct or only produced on the wintering grounds. If the calls recorded by both us and Rankin and Barlow (2007) serve the same function, then the occurrence of these calls on temperate feeding grounds during the spring suggests that these calls are not restricted to any region or season.

Despite our reliance on recordings collected in the presence of right and humpback whales, it is unlikely that either of these species produce the low frequency downsweep call. North Atlantic right whales make a variety of low frequency tonal and FM calls, including a higher frequency downsweep call (Parks and Clark, 2007). A considerable amount of acoustic data have been collected in the exclusive presence of right whales and from sound-recording tags attached to right whales (Matthews *et al.*, 2001; Parks and Tyack, 2005), yet a low frequency downsweep call similar to the one described here has never been detected in those recordings (S. Parks, personal communication). Humpback whales make a stunning variety of sounds organized in both song and individual calls (e.g., Payne and McVay, 1971; D'Vincent *et al.*, 1985; Clark and Clapham, 2004), including vocalizations in the 30–90 Hz frequency band. The rich repertoire and loquaciousness of humpback whales can certainly confound efforts to isolate and identify calls made by other species; however, our analysis indicated that the occurrence of the low frequency call was not at all related to the occurrence of humpback whales (ϕ=−0.015, p=0.722, Table II).

Because the downsweep call described here is at low frequency, FM, and produced often, we suggest that it likely functions as a contact call. Attenuation of an acoustic signal increases exponentially with the frequency; therefore, lower frequency calls can be detected at comparatively longer distances than higher frequency calls. Frequency modulation of the call further improves signal discrimination from background noise and competing sounds (Wiley and Richards, 1978). Sei whales do not tend to aggregate in tightly associated groups (groups tend to be comprised of only a few animals when assessed visually; Perry *et al.*, 1999); therefore, acoustic contact over long distances via a low frequency call may allow dispersed animals to coordinate activities such as feeding or breeding (Payne and Webb, 1971). Downsweep calls at the perimeter of the array (where localization errors are relatively small) were easily detected on all of the recorders, indicating that detection range was at least 7.5 km (the diameter of the area bounded by the array); although localization errors undoubtedly increased with distance from the array, maximum detection distances estimated from calls outside the array were realistically in the range 10–15 km and possibly as high as 20 km.

While some of the variability in the starting frequency and duration of the calls is attributable to attenuation of higher frequency components for distant calls, subtle differences between calls were observed [Fig. 4(a)]. Variability in calls was largely a function of the start frequency; calls with higher starting frequencies tended to be longer, steeper in slope, and swept through a larger range of frequencies. The ending frequency was much less variable than the starting frequency. These characteristics suggest that the ending frequency of the call is relatively fixed, whereas the starting frequency can vary and this variation influences the duration of the call (lower starting frequencies result in shorter calls). This variation may largely be due to differences between individuals. We found that individual calls within pairs [e.g., Figs. 3(b) and 4(b)] were more similar to one another than to calls made by other whales. Although the close spatial proximity of calls produced within pairs and their similarity strongly suggests that these paired calls are produced by a single whale, it is possible that associated animals (e.g., mother and calf, and paired adults) may countercall when nearby to one another. However, the loudness of these paired calls suggests otherwise: Why would a whale produce a call that can be heard ~10 km away when the intended listener is only, on average, 66 m away? If pairs are indeed produced by a single whale, then call structure may be far less variable between calls made by the same whale than between calls by

TABLE IV. Comparisons of both call rate (CR) and call occurrence (CO) with visual abundance (VA) and visual occurrence (VO) using generalized linear models. Models with call rate as the dependent variable are Poisson regression models and those with call occurrence as the dependent variable are logistic regression models. Each drop in deviance statistic has one degree of freedom. In each case, β_1 indicates the nature and magnitude of the relationship between the variables, and the p-value (p) indicates the significance of the relationship. Visual occurrence was treated as an indicator variable in the models (i.e., 0 for absent and 1 for present).

Comparison	Model	β_1	Drop in deviance	p
Call rate versus visual abundance	$\log(\mathrm{CR})=\beta_0+\beta_1(\mathrm{VA})$	−0.0756	0.364	0.546
Call rate versus visual occurrence	$\log(\mathrm{CR})=\beta_0+\beta_1(\mathrm{VO})$	1.29	11.7	0.0006
Call occurrence versus visual abundance	$\mathrm{logit}(\mathrm{CO})=\beta_0+\beta_1(\mathrm{VA})$	0.0535	0.106	0.745
Call occurrence versus visual occurrence[a]	$\mathrm{logit}(\mathrm{CO})=\beta_0+\beta_1(\mathrm{VO})$	2.00	6.33	0.0118

[a]This comparison is equivalent to the Fisher's exact test reported in Sec. III.

different whales. Similar observations in bottlenose dolphin vocalizations by Caldwell and Caldwell (1965) suggested that dolphins produce individually distinctive contact calls (signature whistles), and research in recent decades has provided support for this hypothesis (Smolker *et al.*, 1993; Janik and Slater, 1998, Sayigh *et al.*, 2007). Although no work has been published on signature contact calls in baleen whales (likely owing to the difficulties of experimental manipulation), the prevalence of individually identifiable contact calls in other taxa (reviewed in Boughman and Moss, 2003) suggests that such calls are probably used by baleen whales.

Despite these subtle differences between vocalizations, the prevalence of the downsweep call, its low frequency, and its stereotypic character make this call extremely useful for detecting the presence of vocalizing sei whales. The excellent performance of the synthetic kernel when compared to manual analysis was surprising. The automated method clearly highlighted calls that were impossible to detect visually during routine inspection of a spectrogram unless prompted to carefully review a specific time period both visually and aurally. The increased sensitivity of the automated detector not only allowed identification of faint calls, but also of calls that were missed by a reviewer because of nonoptimal spectrogram viewing parameters (e.g., contrast and brightness) or fatigue. While the automated detector works well for detecting calls when they occur, no detector (human or automated) can detect the presence of whales in an acoustic record when whales are not vocalizing. There were periods at anchor stations 3 and 4 (Fig. 5) when daytime abundances of sei whales were relatively high, yet very few or no calls were produced. During May 2005, Baumgartner and Fratantoni (in press) observed diel periodicity in the calling rates of sei whales in this same region (higher calling rates by day than night). There is some suggestion that calling rates within 3 km of the anchor stations during 2007 also exhibited diel periodicity, but in the opposite direction (higher calling rates by night than day). These preliminary observations are the subject of ongoing research, but clearly temporal variability in calling rates will have a profound impact on estimates of occurrence.

There is often a strong desire in passive acoustic monitoring applications to generate abundance estimates from detection data, yet the nature of the relationship between vocalization behavior and abundance has yet to be elucidated for nearly all marine mammals. For obligate callers, such as echolocating odontocetes, there is hope that vocalization rates may be correlated with abundance, but for baleen whales that vocalize primarily for social reasons, this relationship remains unclear. In our study, hourly vocalization rates of the downsweep call localized within 3 km of the anchor stations (treated as a Poisson process) were not correlated with sei whale abundance determined from the visual surveys, nor was the hourly occurrence of calls related to sei whale abundance (Table IV). Both call rate and call occurrence were significantly related to sei whale occurrence, which simply indicates that the odds of detecting a call as well as the rate of calling increase when sei whales are present (a rather self-evident result when one accepts that sei whales produce the downsweep call). From these observations, it is clear that the abundance of sei whales cannot be predicted from vocalization rates at hourly time scales. Even occurrence, although strongly related to vocalization rates and the occurrence of calls, is not perfectly predicted by vocalizations. Accurate detection of a vocalization is certainly conclusive evidence of the presence of a whale, but silence is not always indicative of an absence of whales; in our study, sei whales were present during 17 (29%) of the 59 hourly periods during which calls were not detected within 3 km of the anchor stations (Table II). Interestingly, there were two hourly periods when sei whale calls were detected, yet no sei whales were seen. In each of these periods, the calls were likely produced by a single animal (based on the pattern of localizations: clumped within 250 m of one another in one case, and in a linear sequence indicative of a traveling whale in the other) approximately 2 km from the ship during the observing period. Visual observations are clearly the only reliable method to confirm the identity of calling whales, but not all vocalizing whales, particularly single whales, can be seen and identified using the methods employed in this study.

Comparisons between the occurrence of localized low frequency downsweep calls and the presence of baleen whales around the anchor stations provided strong evidence that sei whales produced these calls. A relatively large dataset of collocated visual and acoustic observations (70 h) was required to successfully attribute the downsweep call to a particular species because of the high abundance and cooc-

currence of three potential sources: right, sei, and humpback whales. Previous studies have relied on relatively short-term recordings of a particular call in the presence of only a single species for attribution of the call to that species; however, this approach precludes attribution in areas frequented by several species, is susceptible to confounding by audible animals of other species that are not in visual range, and can be thwarted by temporal variability in calling behavior (i.e., encounters with silent whales prevent attribution of their calls). Longer collocated visual and acoustic time series, while expensive and not particularly glamorous to collect, are critical to the successful identification of species-specific calls. Such information will allow progressive acoustic studies, which (1) utilize a larger suite of calls per species, (2) characterize calling variability (in terms of the types of vocalizations) and the environmental and social context of the calls, and (3) describe the community composition of the vocalizing animals.

ACKNOWLEDGMENTS

We are indebted to the many observers that collected the visual sighting data for this study, including Ingrid Biedron, Shonda Gaylord, Nicole Gilles, Elizabeth Josephsen, Betty Lentell, Nadine Lysiak, Sarah Mussoline, John Nicolas, Richard Pace, Melissa Patrician, David Potter, and Elizabeth Vu. We are also grateful for the able help of the captain, officers, and crew of the NOAA R/V *Albatross IV* for their help at sea. Cornell University Laboratory of Ornithology's Bioacoustics Research Program graciously made the MARUs available to us for lease. Fred Serchuk provided helpful criticisms of this paper, as did two anonymous reviewers. Funding was provided by the NOAA National Marine Fisheries Service and the WHOI Ocean Life Institute.

Baumgartner, M. F., and Fratantoni, D. M. (**2008**). "Diel periodicity in both sei whale vocalization rates and the vertical migration of their copepod prey observed from ocean gliders," Limnol. Oceanogr. (in press).

Baumgartner, M. F., Fratantoni, D. M., and Clark, D. W. (**2006**). "Investigating baleen whale ecology with simultaneous oceanographic and acoustic observations from autonomous underwater vehicles," Eos Transactions, American Geophysical Union 87(36), Ocean Sciences Meeting Supplement, Abstract OS24E-05, Honolulu, Hawaii, 20–24 February.

Baumgartner, M. F., Freitag, L., Partan, J., Ball, K., and Prada, K. (**2008**). "Tracking large marine predators in three dimensions: The real-time acoustic tracking system," IEEE J. Ocean. Eng. (in press).

Boughman, J. W., and Moss, C. F. (**2003**). "Social sounds: Vocal learning and development of mammal and bird calls," in *Acoustic Communication*, edited by A. M. Simmons, A. N. Popper, and R. R. Fay (Springer, New York), pp. 138–224.

Caldwell, M. C., and Caldwell, D. K. (**1965**). "Individualized whistle contours in bottlenose dolphins (*Tursiops truncatus*)," Nature (London) **207**, 434–435.

Clark, C. W., and Clapham, P. J. (**2004**). "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring," Proc. R. Soc. London, Ser. B **271**, 1051–1057.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Fores, S., Gedamke, J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.

Darling, J. D., Jones, M. E., and Nicklin, C. P. (**2006**). "Humpback whale songs: Do they organize males during the breeding season?," Behaviour **143**, 1051–1101.

D'Vincent, C. G., Nilson, R. M., and Hanna, R. E. (**1985**). "Vocalization and coordinated feeding behavior of the humpback whale in southeastern Alaska," Sci. Rep. Whales Res. Inst. **36**, 11–47.

Flinn, R. D., Trites, A. W., Gregr, E. J., and Perry, R. I. (**2002**). "Diets of fin, sei, and sperm whales in British Columbia: An analysis of commercial whaling records, 1963–1967," Marine Mammal Sci. **18**, 663–679.

Foy, W. H. (**1976**). "Position-location solutions by Taylor-series estimation," IEEE Trans. Aerosp. Electron. Syst. **12**, 187–194.

Fratantoni, D. M., and Baumgartner, M. F. (**2005**). "AUV-based physical, biological, and acoustic observations in support of marine mammal ecology studies," *International Ocean Research Conference*, 6–10 June (The Oceanography Society, Paris, France).

Freitag, L., Johnson, M., Grund, M., Singh, S., and Preisig, J. (**2001**). "Integrated acoustic communication and navigation for multiple UUVs," Proceedings of MTS/IEEE OCEANS 2001, Vol. **4**, pp. 2065–2070.

George, J. C., Zeh, J., Suydam, R., and Clark, C. (**2004**). "Abundance and population trend (1978–2001) of western Arctic bowhead whales surveyed near Barrow, Alaska," Marine Mammal Sci. **20**, 755–773.

Heimlich, S. L., Mellinger, D. K., and Nieukirk, S. L. (**2005**). "Types, distribution, and seasonal occurrence of sounds attributed to Bryde's whales (*Balaenoptera edeni*) recorded in the eastern tropical Pacific, 1999–2001," J. Acoust. Soc. Am. **118**, 1830–1837.

Hjort, J., and Ruud, J. T. (**1929**). "Whales and plankton in the North Atlantic," Conseil Permanent International pour l'Exploration de la Mer. Rapports et Procès Verbaux des Réunions, Vol. **56**, pp. 5–123.

Janik, V. M., and Slater, P. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.

Johnson, M., Madsen, P. T., Aguilar Soto, N., and Tyack, P. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.

Kawamura, A. (**1974**). "Food and feeding ecology in the southern sei whale," Sci. Rep. Whales Res. Inst. **26**, 25–144.

Knowlton, A. R., Clark, C. W., and Kraus, S. D. (**1991**). "Sounds recorded in the presence of sei whales, *Balaenoptera borealis*," Abstract in the Ninth Biennial Conference on the Biology of Marine Mammals, Chicago, IL, p. 40.

Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., and Moehl, B. (**2002**). "Sperm whale sound production studied with ultrasound time/depth-recording tags," J. Exp. Biol. **205**, 1899–1906.

Matthews, J. N., Brown, S., Gillespie, D., Johnson, M., McLanaghan, R., Moscrop, A., Nowacek, D., Leaper, R., Lewis, T., and Tyack, P. (**2001**). "Vocalisation rates of the North Atlantic right whale (*Eubalaena glacialis*)," J. Cetacean Res. Manage. **3**, 271–282.

McDonald, M. A., Hildebrand, J. A., Wiggins, S. M., Thiele, D., Glasgow, D., and Moore, S. E. (**2005**). "Sei whale sounds recorded in the Antarctic," J. Acoust. Soc. Am. **118**, 3941–3945.

Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.

Mellinger, D. K., Nieukirk, S. L., Matsumoto, H., Heimlich, S. L., Dziak, R. P., Haxel, J., Fowler, M., Meinig, C., and Miller, H. V. (**2007**). "Seasonal occurrence of North Atlantic right whale (*Eubalaena glacialis*) vocalizations at two sites on the Scotian Shelf," Marine Mammal Sci. **23**, 856–867.

Mizroch, S. A., Rice, D. W., and Breiwick, J. M. (**1984**). "The sei whale, *Balaenoptera borealis*," Mar. Fish. Rev. **46**, 25–29.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.

Parks, S. E., and Clark, C. W. (**2007**). "Acoustic communication: Social sounds and the potential impacts of noise," in *The Urban Whale: North Atlantic Right Whales at the Crossroads*, edited by S. D. Kraus and R. M. Rolland (Harvard University Press, Cambridge, MA), pp. 310–332.

Parks, S. E., and Tyack, P. L. (**2005**). "Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups," J. Acoust. Soc. Am. **117**, 3297–3306.

Payne, R., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 585–597.

Payne, R., and Webb, D. (**1971**). "Orientation by means of long range acoustic signaling in baleen whales," Ann. N.Y. Acad. Sci. **188**, 110–142.

Perry, S. L., DeMaster, D. P., and Silber, G. K. (**1999**). "The great whales: History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973," Mar. Fish. Rev. **61**, 1–74.

Rankin, S., and Barlow, J. (**2007**). "Vocalizations of the sei whale *Balaenoptera borealis* off the Hawaiian Islands," Bioacoustics **16**, 137–145.

Sayigh, L. S., Esch, H. C., Wells, R. S., and Janik, V. M. (**2007**). "Facts about signature whistles of bottlenose dolphins, *Tursiops truncates*," Anim. Behav. **74**, 1631–1642.
Smolker, R. A., Mann, J., and Smuts, B. B. (**1993**). "Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants," Behav. Ecol. Sociobiol. **33**, 393–402.
Sokal, R. R., and Rohlf, F. J. (**1995**). *Biometry*, 3rd ed. (W. H. Freeman, New York).
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**2001**). "Geographic and seasonal variation of blue whale calls in the North Pacific," J. Cetacean Res. Manage. **3**, 65–76.
Stimpert, A. K., Wiley, D. N., Au, W. W. L., Johnson, M. P., and Arsenault, R. (**2007**). "'Megapclicks': Acoustic click trains and buzzes produced during night-time foraging of humpback whales (*Megaptera novaeangliae*)," Biotechnol. Lett. **3**, 467–470.
Thompson, T. J., Winn, H. E., and Perkins, P. J. (**1979**). "Mysticete sounds," in: *Behavior of Marine Animals, Vol. 3: Cetaceans*, edited by H. E. Winn and B. L. Olla (Plenum, New York), pp. 403–431.
Watkins, W. A., and Schevill, W. E. (**1971**). "Four hydrophone array for acoustic three-dimensional location," Technical Report No. 71–60, Woods Hole Oceanographic Institution.
Wiley, R. H., and Richards, D. G. (**1978**). "Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations," Behav. Ecol. Sociobiol. **3**, 69–94.
Zimmer, W. M. X., Johnson, M., Madsen, P. T. M., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.

Temporal scales of auditory objects underlying birdsong vocal recognition

Timothy Q. Gentner[a)]
Department of Psychology, Neurosciences Graduate Program, University of California, San Diego, La Jolla, California 92093

(Received 20 November 2007; revised 5 May 2008; accepted 28 May 2008)

Vocal recognition is common among songbirds, and provides an excellent model system to study the perceptual and neurobiological mechanisms for processing natural vocal communication signals. Male European starlings, a species of songbird, learn to recognize the songs of multiple conspecific males by attending to stereotyped acoustic patterns, and these learned patterns elicit selective neuronal responses in auditory forebrain neurons. The present study investigates the perceptual grouping of spectrotemporal acoustic patterns in starling song at multiple temporal scales. The results show that permutations in sequencing of submotif acoustic features have significant effects on song recognition, and that these effects are specific to songs that comprise learned motifs. The observations suggest that (1) motifs form auditory objects embedded in a hierarchy of acoustic patterns, (2) that object-based song perception emerges without explicit reinforcement, and (3) that multiple temporal scales within the acoustic pattern hierarchy convey information about the individual identity of the singer. The authors discuss the results in the context of auditory object formation and talker recognition. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2945705]

PACS number(s): 43.80.Lb, 43.66.Gf [JAS] Pages: 1350–1359

I. INTRODUCTION

Male European starlings, *Sturnus vulgaris*, produce behaviorally relevant vocalizations (songs) that are spectrally and temporally complex. Both male and female adult starlings can learn to recognize large sets of these songs, even when sung by several different conspecific males (Gentner *et al.*, 2000). This recognition learning, in turn, drives neuronal plasticity in regions of the auditory forebrain analogous to mammalian auditory cortex, where neurons respond most strongly to the songs that birds have learned to recognize (Gentner and Margoliash, 2003). Several lines of evidence point to the importance of short, stereotyped, spectrotemporal patterns called motifs, in guiding song recognition. For example, one can closely control behavioral and physiological responses to songs by manipulation of song acoustics at the motif level, which equates to timescales on the order of several 100's of milliseconds (Gentner, 2004). It is not understood, however, how spectrotemporal acoustic structures defined on more precise timescales contribute to song recognition. One hypothesis is that song recognition is driven explicitly by submotif level features, rather than whole motifs. Alternatively, information about the identity of the singer may be coded at multiple temporal scales within songs. As an initial step in understanding the detailed relationships between vocal recognition and song acoustics we report here on how the short timescale acoustic structure of songs governs recognition behavior.

Starling song is hierarchically structured. Males tend to sing in long continuous episodes called *bouts*. Song bouts are composed of much shorter acoustic units referred to as *motifs* (Adret-Hausberger and Jenkins, 1988; Eens *et al.*, 1991) (Fig. 1) that, in turn, are composed of still shorter units called *notes*. Notes can be broadly classified by the presence of continuous energy in their spectrotemporal representations. The note pattern within a given motif is largely stereotyped across successive motif renditions, and each motif is often repeated two or more times in a song before a different motif is sung. Thus, starling songs appear (acoustically) as sequences of iterated motifs, where each motif is a spectrotemporally complex event (Fig. 1). Different song bouts from the same male are not necessarily composed of the same set of motifs. A complete repertoire of motifs can, however, be characterized over many song bouts, and for a mature male starling can exceed 50 or more unique motifs (Eens *et al.*, 1989; Eens, 1992; 1997; Gentner and Hulse, 1998).

Although some sharing of motifs does occur among captive males (Hausberger and Cousillas, 1995; Hausberger, 1997), the motif repertoires of different males living in the wild are generally unique (Adret-Hausberger and Jenkins, 1988; Eens *et al.*, 1989, 1991; Chaiken *et al.*, 1993; Gentner and Hulse, 1998). Thus, learning which males sing which motifs can provide a diagnostic cue for individual recognition. At least to a first approximation, this strategy does a good job of describing how starlings learn to recognize conspecific songs. Using operant trainings techniques, starlings can easily learn to recognize many songs sung by different individuals, and can maintain this accurate recognition when classifying novel song bouts from the training singers (Gentner and Hulse, 1998; Gentner *et al.*, 2000). When the novel song bouts have *no* motifs in common with the training songs, however, performance in this recognition task falls to chance (Gentner *et al.*, 2000). Also consistent with a motif-

a)Electronic mail: tgentner@ucsd.edu

 0001-4966/2008/124(2)/1350/10/$23.00

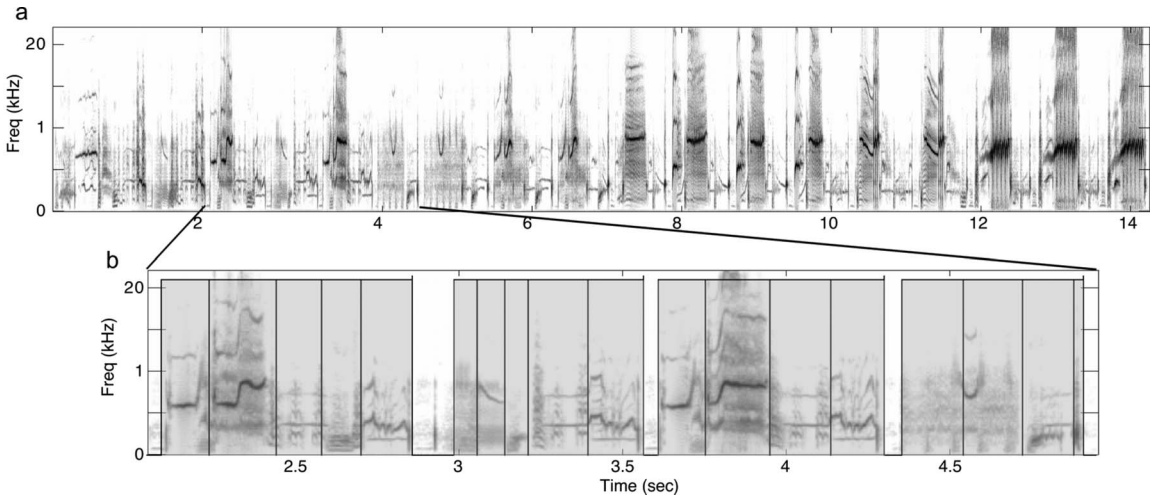


FIG. 1. Hierarchical organization of starling song. Spectrograms showing spectral power as a function of time for (a) a 14-s long excerpt from a much longer bout of singing by one male European starling and (b) a region from the same song with the temporal axis expanded. Motifs in (b) are denoted by the gray shading, and the boundaries of submotif features (see Sec. II) are shown by the overlaid black boxes. Note the highly variable but still repetitive structure, and the hierarchy of groupings for spectrotemporal acoustic patterns.

memorization strategy, the recognition of hybrid, chimeric, songs based on the natural songs of two familiar males shows a linear relationship to the relative proportions of familiar motifs in each bout (Gentner and Hulse, 2000).

Taken together, the results of these studies suggest that when starlings learn to recognize conspecific songs from different singers, they memorize large numbers of unique motifs corresponding to individual singers. The spectrotemporal structure and scale of motifs can vary widely, however, even within a single bird, and the apparent recognition of "whole motifs" may result from recognition of spectrally and/or temporally restricted acoustic features that are diagnostic of (or uniquely covariant with) the larger event (motif). That is, starlings might learn notes rather than motifs. Although motifs form an operationally and phenomenologically useful unit for manipulations that alter song recognition, the necessity of the motif as the minimal perceptual unit for song recognition has not been established empirically. Here we explore the temporal lower bound on acoustic pattern recognition in starlings by testing the recognition of conspecific songs that have been systematically manipulated at submotif temporal scales.

II. GENERAL METHODS

A. Subjects

Seven adult European starlings, *Sturnus vulgaris*, served as subjects in this study. All subjects were wild and caught in southern California in May 2006. All had full adult plumage at the time of capture, and thus were at least one year old. From the time of capture until their use in this study, all subjects were housed in large mixed sex, conspecific aviaries with *ad libitum* access to food and water. The photoperiod in the aviary and the testing chambers followed the seasonal variation in local sunrise and sunset times. No significant sex differences have been observed in previous studies of individual vocal recognition (Gentner *et al.*, 2000), and the sex of subjects in this study was not controlled.

B. Apparatus

Starlings learned to classify the training stimuli using a custom-built operant apparatus, housed in a $61 \times 81 \times 56$ cm inner diameter sound attenuation chamber (Acoustic Systems). Inside the chamber, a subject was held in a weld-wire cage ($41 \times 41 \times 35$ cm) that permitted access to a 30×30 cm operant panel mounted on one wall. The operant panel contained three circular response ports spaced 6 cm center-to-center, aligned in a row with the center of each port roughly 14 cm off the floor of the cage and with the whole row centered on the width of the panel. Each response port was a PVC housed opening in the panel fitted with an IR receiver and transmitter that detected when the bird broke the plane of the response port with its beak. This "poke-hole" design allows starlings to probe the apparatus with their beak, in a manner akin to their natural appetitive foraging behavior. Independently controlled light emitting diodes (LEDs) could illuminate each response port from the rear. Directly below the center port, in the section of the cage floor immediately adjacent to the panel, a fourth PVC lined opening provided access to food. A remotely controlled hopper, positioned behind the panel, moved the food into and out of the subject's reach beneath the opening. Acoustic stimuli were delivered through a small full-range audio speaker mounted roughly 30 cm behind the panel and out of the subject's view. The sound-pressure level (SPL) inside all chambers was calibrated to the same standard broadband signal. Custom software monitored the subject's responses, and controlled the LEDs, food hoppers, chamber light, and auditory stimulus presentation according to procedural contingencies.

C. Stimuli

1. Song recording

Recordings of four male European starlings were used to generate all the stimuli for this experiment. The procedures for obtaining high-quality song recordings from male starlings have been detailed elsewhere (Gentner and Hulse,

1998). Briefly, a minimum of 0.5 h of song was recorded from each male when housed individually in a large sound-attenuating chamber. During recording, males had visual and auditory access to a female starling (the same female was used to induce song from all the males). All the songs were recorded on digital audiotape (16 bit, 44.1 kHz) using the same microphone (Sennheiser ME66-K6), and high-pass filtered at 250 Hz to remove low frequency background noise. The multiple songs of each bird were parsed into roughly 15-s exemplars of continuous singing taken from the beginning, middle, or end of a typically much longer song bout, and then sorted into sets based on the presence or absence of motifs that were shared with other 15-s song exemplars from the same bird. Human observers labeled the motifs in each 15-s exemplar. These same stimuli have been used to explore the role of motif familiarity in the recognition of individual songs in several studies (Gentner and Hulse, 1998; 2000; Gentner *et al.*, 2000). None of the males whose songs were used to generate the stimuli for the present study served as subjects in the operant testing described here.

2. Baseline training stimuli

To avoid issues related to psuedoreplication (Kroodsma, 1989), we used three different stimulus sets for the baseline song classification. Each stimulus set consisted of eight song exemplars drawn from the library of song bouts sampled from a single bird, and eight exemplars drawn from the songs of another bird (16 exemplars total). The singer of each set of songs and the assignment of those songs as either S+ or S− was counterbalanced across test subjects. Each exemplar was 15 ± 0.5 s of continuous song taken from either the beginning, middle, or end of a song bout, as described previously. Many of the exemplars sampled from the beginning of a song bout included whistles, along with other "warble" motifs [i.e., "variable" motifs, rattles, and high-frequency motifs, for motif nomenclature see Adret-Hausberger and Jenkins (1988) and Eens *et al.* (1991)]. Those sampled at later time points in a bout comprised only warble song motifs. Previous data indicate that recognition is easily learned with this length of a song exemplar, and is unaffected by the relative position within a longer song bout from which the exemplar is sampled and/or the broader motif classes it may or may not contain (Gentner and Hulse, 1998).

3. Submotif permutations

From each of the four original sets of 15-s song exemplars we selected 16 additional 15-s song stimuli that served as the basis for further testing. Eight of these 16 song stimuli had motifs in common with the baseline training exemplars from the same singer and eight were composed entirely of novel motifs sung by the same male (i.e. motifs that did not appear in any baseline training songs). We refer to these two types of song stimuli as "familiar-motif" and "unfamiliar-motif" song, respectively. The familiar-motif songs shared, on average, 89.1% of their motifs with the baseline training songs.

We parsed the 8 familiar-motif and 8 unfamiliar-motif songs from each singer into constituent submotif segments

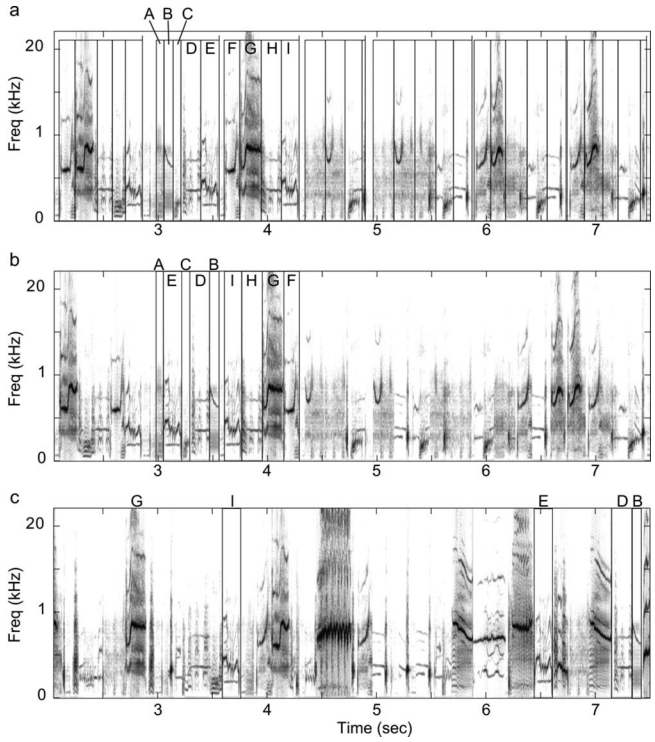


FIG. 2. Submotif song features and permutations. Example spectrograms from excerpts of three test exemplars where ordering of the submotif features, denoted by the black overlaid boxes, is (a) unpermuted (i.e., naturally ordered), (b) permuted within the temporal boundaries of their original motif, or (c) permuted over the entire exemplar. The submotif features for the second and third motifs in (a) are labeled with letters, so that their permuted positions in (b) and (c) can be more easily seen. Because only an excerpt of the entire stimulus exemplar is shown, not all of the submotif features labeled in (a) appear in (c).

with a semiautomated procedure, detailed as follows. We converted the waveform of each 15-s song stimulus to the frequency domain, by computing the spectrogram of the original stimulus (9.27 ms window, 90% overlap; 512 fft points; see Fig. 2). We obtained an estimate of the noise over a single time slice in a set of songs by manually identifying periods of silence in the songs from each individual, and then computed the mean magnitude of the power spectra (512 fft points) for the "silence" intervals separately for each bird. We then recursively compared each time slice of the stimulus to the silence interval (by taking the normalized inner product) to find points where the song changed from "signal" to silence. On each recursive call, the similarity threshold used to classify the given time slice as signal or "noise" was progressively lowered so that we detected increasingly subtle temporal breaks within the segments yielded by the preceding call. We halted the recursion when we reached a minimum threshold value (held constant for all songs), and marked the start and stop times for all segments parsed in this way. In principle, of course, the recursive parsing procedure could be carried on until arbitrarily small segments were delineated (up to the temporal resolution of the spectrogram).

Using dynamic time warping (DTW) (see Anderson *et al.*, 1996; Kogan and Margoliash, 1998) we compared each parsed song segment to a library of segment templates for that bird, and determined the template (or set of tem-

plates) that yielded the closest match. From this match we then modified (and confirmed by hand) the segment start and stop times in the parsed songs as necessary to ensure that different iterations of the same motif were parsed into similar sets of segments. The reference library of segment templates for each bird was obtained from a random sample of all the motifs that that bird produced, and optimized so that it produced a minimum error when classifying (using similar DTW techniques) all of the motifs in that bird's repertoire (data not published).

To permute the temporal sequence of each song, we shuffled the order of submotif segments either (1) within the temporal boundaries of their original motif [Fig. 2(b)], or (2) over the whole song [Fig. 2(c)]. In the former case, we constrained the permutation so that the original ordering of segments within a motif was not replicated. The segmentation, dynamic time warping, and permutation routines were written in MATLAB (v7.4). The threshold level for noise/signal and the resulting minimum segment duration were established empirically as part of a larger (unpublished) endeavor using DTW to optimally match a minimum number of spectral-temporal templates to all the submotif components in the songs from a single male starling. We extracted a total of 5141 submotif segments from all of the songs. The mean ($\pm$sem) number of motif and submotif segments extracted from a given 15-s song stimulus was 15.61 ± 0.39 and 80.33 ± 1.94, respectively. The mean ($\pm$sem) submotif segment duration was 154.66 ± 1.24 ms with a range from 10 to 663 ms. The mean ($\pm$sem) motif duration over all song stimuli was 873.12 ± 27.74 ms, with a range from 191 to 2762 ms.

For each of the 16 test songs from a given singer we created four permutations with submotif segments randomly distributed over the whole song and four permutations with submotif segments reordered within their original motif boundaries. This yielded a total of 256 permutation sequences per baseline stimulus set. Because many more than four permutations are possible, we selected a subset that coved the possible range of permutations as uniformly as possible assuming all possible positions for a given submotif segment are equally salient (which is a strong but valid *a priori* assumption given the available data). To quantify the complexity of any given permutation we took the number of submotif segment transitions that differed from the original sequence, normalized by the number of possible transitions. Suppose S is the original stimulus and T is the permuted stimulus and that $S(i)$ represents ith feature in S, then the "R distance" is calculated as the number of times $S(i+1)$ does not follow $S(i)$ in T. For example, when $S=[1\ 2\ 3\ 4\ 5]$ and $T=[4\ 5\ 1\ 2\ 3]$, the R distance for $T=1/(5-1)=0.25$, as only one transition (from 3 to 4) is absent from T. If $T=[3\ 1\ 5\ 2\ 4]$, the R distance is $4/(5-1)=1$ as all 4 original transitions are absent.

D. Procedure

1. Shaping

Subjects learned to work the apparatus through a series of successive shaping procedures. Upon initially entering the operant chamber, the subject was given unrestricted access to the food hopper, and then taught through autoshaping to peck the center port to gain access to the food. Once the subject pecked reliably at the center port to obtain food, the center LED ceased flashing, while the requirement to peck at the same location remained in effect. Shortly thereafter, pecks to the center port initiated the presentation of a song stimulus, and the trial proceeded as described in Sec. II D 2. In all cases, initial shaping occurred within one to two days, and was followed immediately by the start of song recognition training.

2. Song recognition training

Each subject was trained initially to classify sixteen 15-s song exemplars (8 exemplars from 2 singers) using a "go-nogo" operant procedure. In this procedure, subjects initiated a trial by pecking at the center response port to trigger the immediate presentation of a training song. Following stimulus presentation the animal was required to either peck the center response port again, or to withhold responses altogether. Responses to half of the stimuli (S+) were reinforced positively with 2-s access to the food hopper. Responses to the other half of the stimuli (S−) were punished by extinguishing the house light for 2–10 s and denying food access. Failure to respond to either S+ or S− stimuli had no operant consequence. For performance evaluation, we considered a response to an S+ stimulus and the withholding of a response to an S− stimulus as "correct." Withholding a response to an S+ stimulus and responding to an S− stimulus were considered "incorrect." Subjects could freely peck at the center response port throughout stimulus presentation, but only the first response within a 2-s response window beginning at stimulus offset triggered reinforcement or punishment. Responses prior to completion of the stimulus were ignored.

The stimulus exemplar presented on any given trial was selected randomly with uniform probability from the pool of all 16 stimuli the animal was learning to classify. The intertrial interval was 2 s. Water was always available. Subjects were on a closed economy during training, with daily sessions lasting from sunrise to sunset, and each subject could run as few or as many trials as they were able. Food intake was monitored daily to ensure each subject's well being. The explicit pairings of songs for baseline training was counterbalanced across subjects. All procedures were approved by the UCSD institutional animal care and use committee whose policies are consistent with the Ethical Principles of the ASA.

3. Test procedure

Prior to initiation of the first test session, the rate of food reinforcement for correct responses to S+ stimuli was lowered from 100% (where it had been during baseline training) to 80%, and the rate of "punishment" (dimmed house lights) for incorrect responses to S− stimuli was lowered to 95%. After performance again stabilized, typically within one or two sessions, we began presenting test stimuli on roughly 20% of the trials. The test stimuli were 32 naturally ordered songs composed of either familiar or novel motifs and 256

other versions of these songs (8/singer/familiarity type) where the submotif ordering was permuted (see Sec. II C 3). The test stimulus for a given trial was selected randomly from the set of all possible test stimuli for that subject, and balanced so that equal numbers of permuted and unpermuted songs were presented. We reinforced responses to test stimuli nondifferentially regardless of accuracy as follows: each response to a test stimulus had a 40% chance of eliciting a food reward, a 40% chance of eliciting punishment (timeout without food), and a 20% chance of eliciting no operant consequence. Because reinforcement of the test stimuli was random and nondifferential with respect to response outcome, subjects had no opportunity to learn to associate a given test stimulus with a given response. Thus, the correct classification of test stimuli can be taken as strong evidence for generalization rather than learning rote sets of specific training exemplars. If there was no generalization, classification accuracy would be at chance and all responses would be the same.

E. Analysis

We used d' to estimate the sensitivity for classification of baseline training song stimuli, and the various test stimuli as given by

$$d' = z(H) - z(F),$$

where H gives the proportion of responses to an S+ stimulus, F gives the proportion of responses to an S− stimulus, and $z(\,)$ denotes the z score of those random variables. The measure d' is convenient because it eliminates any biases in the response rates (e.g., due to guessing) that may vary across individuals and within individuals over time. To gauge the effect of various song manipulations during the test sessions, we compared d' values for different stimulus classes using repeated measures analysis of variance (ANOVA), and where appropriate used post-hoc analyses to quantify the significance of specific differences between mean d' measures. Identical analyses conducted on mean percent correct scores yielded the same results.

III. RESULTS

All seven subjects easily learned the initial song classification task, sorting the songs of two conspecific males into separate classes with high accuracy. The mean performance over all subjects showed significant improvement over the course of training ($F_{(6,39)}=7.65$, $p<0.0001$, repeated measures ANOVA), with individual birds requiring 1300–3800 trials to achieve reliably accurate classification (mean d' over five consecutive blocks >1.0). At asymptote, the mean ($\pm$sem) d' for all birds was 3.6 ± 0.32 (Fig. 3), and the percent correct combined for both classes of songs was 87.5 ± 0.01.

Once subjects achieved stable and accurate performance on the baseline training songs, we presented test stimuli on a subset of all trials (see Sec. II D 3). As expected from previous studies, subjects were significantly better at classifying novel songs composed of familiar motifs compared to those composed entirely of unfamiliar motifs from the same singers (i.e., motifs that did not appear in any of the baseline training songs). The mean ($\pm$sem) d' for the naturally ordered familiar-motif songs was 1.47 ± 0.21, whereas that for the naturally ordered unfamiliar-motif songs was 0.59 ± 0.13, and these values were significantly different from each other ($F_{(1,6)}=28.66$, $p<0.005$, repeated measures ANOVA). Interestingly, responding to both classes of naturally ordered songs was significantly above chance ($d'=0$) ($t=7.07$, $p<0.0005$; $t=4.60$, $p<0.005$, for songs with familiar and unfamiliar motifs, respectively). The mean proportions of correct responses to the test songs composed of familiar and unfamiliar motifs were 0.70 ± 0.04 and 0.59 ± 0.03, respectively, and both means are significantly above chance (t-test, $p\leqslant0.02$, both cases).

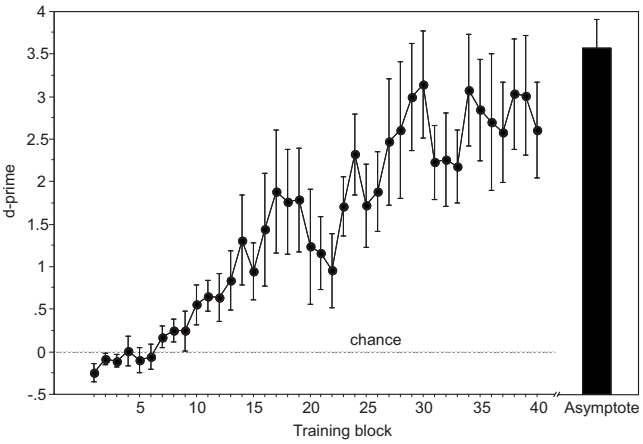


FIG. 3. Acquisition of song recognition. Mean d-prime ($\pm$sem) shown over the course of successive 100-trial training blocks showing the gradual improvement in classification of songs from two different singers. Performance is above chance after seven blocks, which includes acquisition of all operant task requirements (e.g., when to peck) as well as knowledge about the stimuli. Mean ($\pm$sem) performance at asymptote, just prior to testing (see methods), is shown by the black bar on the right.

The foregoing results are consistent with previous studies (Gentner and Hulse, 1998; Gentner *et al.*, 2000) in showing a clear advantage for song recognition when familiar motifs are present in novel, to-be-recognized, song bouts compared to when a bout is composed entirely of novel motifs sung by an otherwise familiar singer. This recognition advantage may result from a representation of motifs as holistic acoustic patterns or from the representation of diagnostic acoustic features centered below the level (or temporal scale) of the motif. Previous manipulations, that have only altered songs at the level of the motif, cannot distinguish between these two recognition strategies. To test the availability of acoustic pattern information at temporal scales shorter than a single motif we also asked birds to classify versions of the naturally ordered test stimuli that had been parsed into submotif features, whose order was permuted either within a motif or within the entire song (see Sec. II C 3). Permuted and naturally ordered songs were presented in the same test sessions.

The permuted songs were significantly more difficult for the subjects to recognize than the naturally ordered songs ($F_{(2,12)}=9.91$, $p<0.005$, main effect of permutation), but this effect was restricted to the familiar songs. That is, permuting

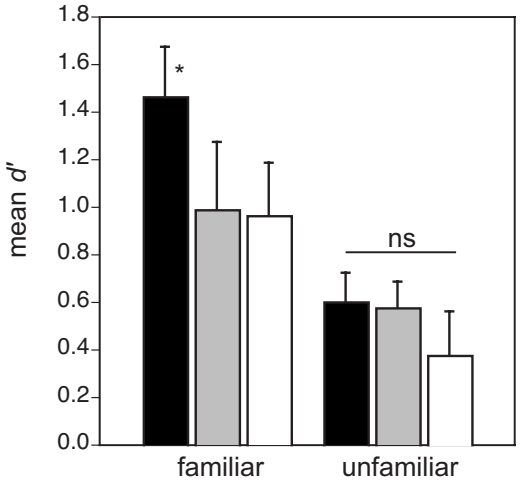


FIG. 4. Permutation test results. Mean (±sem) classification performance for the various test stimuli composed of submotif features derived from familiar (left-most bars) and unfamiliar (right-most bars) motifs. Black bars show the classification of stimuli in which the sequence of submotif features is not permuted from its natural order, gray bars when the permutation displacement is constrained by the temporal boundaries of the original motif (see methods), and white bars when the permutation is allowed over the entire exemplar. The asterisk denotes a significant difference from the other familiar-motif songs.

the submotif features in songs composed of familiar motifs lead to significant impairments in recognition ($F_{(2,12)}=8.98$, $p<0.005$, repeated measures ANOVA), whereas the same permutations of songs composed of novel motifs had no effect ($F_{(2,12)}=1.63$, NS; Fig. 4). The mean (±sem) d' values for the permutations of songs composed of familiar and unfamiliar motifs were 0.97 ± 0.18 and 0.47 ± 0.11, respectively, and both of these means were significantly greater than chance ($d'=0$) (one tailed t-test, $p\leqslant0.0008$, both cases; Fig. 4). Collapsing across permutation type (i.e., motif level or song level), the overall classification for permuted versions of the familiar-motif songs was significantly better than that for the permuted versions of unfamiliar-motif songs ($F_{(1,6)}=11.70$, $p<0.05$, repeated measures ANOVA; Fig. 4).

Surprisingly, subjects classified the permuted songs with similar proficiency regardless of whether the submotif features were shuffled over the range of their original motif or over the whole song sample. For songs composed of familiar motifs, the mean (±sem) d' values associated with motif- and song-shuffled permutations were 0.99 ± 0.29 and 0.96 ± 0.23, respectively, and the difference between these means was not significant. For songs composed of unfamiliar motifs, the mean (±sem) d' values associated with motif- and song-shuffled permutations were 0.57 ± 0.11 and 0.38 ± 0.19, respectively, and the difference between these means was not significant. The mean d' values of all of the permutations, except the song-level permutations with unfamiliar motifs, were above chance ($d'=0$) ($p<0.05$, all cases; Fig. 4).

The similar levels of recognition observed for both motif- and song-shuffled permutations of songs with familiar motifs suggests that any violation of motif-segment sequencing leads to similar deficits. This may reflect a deficit in object recognition or in the recognition of explicit sequence of submotif features. To examine these hypotheses, we asked how the severity of a given submotif permutation, measured with R distance (see Sec. II), was related to recognition. If subjects had learned the explicit sequence of submotif segments, then more severe permutations (i.e., those that broke most of the segment transitions) should be harder to recognize than those that were less severe. In fact, we observed the opposite. For songs composed of familiar motifs, recognition of the motif-shuffled songs was positively correlated with R distance ($r=0.219$, $p<0.005$). That is, as the motif-level permutations in these songs became more severe, recognition improved. For the song-shuffled versions of familiar-motif songs and both types of permutations of the unfamiliar songs there was no correlation between R distance and the recognizability of a given song ($r<0.05$, all cases, NS). The mean (±sem) R distance for familiar and novel-motif songs combined was 0.82 ± 0.003 and 0.98 ± 0.001, for permutations within motifs and over the whole song, respectively.

Given that all of the test songs were recognized at levels above chance, it is helpful to consider the degree of acoustic similarity between the various stimulus sets. DTW provides a tool for assessing similarities between complex waveforms. To assess the motif-segment similarity across songs, we parsed the baseline stimuli into constituent submotif segments using the same procedures as for the test stimuli (see Sec. II), and then found the best DTW match. We expressed the best match as the minimum path length or "distance," $D_{\min}$, (a unitless number) between each segment in the baseline songs and all the segments in the naturally ordered test songs. Smaller values for $D_{\min}$ denote greater similarity. As expected the baseline submotif segments more closely matched the segments that made up the familiar-motif test songs than the unfamiliar-motif test songs from the same singer [Fig. 5(a)]. The mean (±sem) minimum distance ($D_{\min}$) between baseline and familiar-motif songs was 37.11 ± 0.52 and that between baseline and unfamiliar-motif songs was 48.75 ± 0.64. These values are significantly different ($p<0.0001$, $t=-28.38$, paired t-test). Even though the "unfamiliar" test songs had no motifs in common with the training stimuli, the submotif segments from these songs were significantly more similar to those from the training songs of the same singer than to the segments from the training songs of the opposing singer in the training set [Fig. 5(b); see Sec. II]. The mean (±sem) minimum distance between baseline and unfamiliar-motif songs from the opposing singer was 53.14 ± 0.64, and this is significantly larger than the best matches between baseline and unfamiliar-motif segments from the same singer ($p<0.0001$, $t=-14.97$, paired t-test). The data for the individual stimulus sets are shown in Fig. 5(b).

IV. CONCLUSIONS

The song recognition system of European starlings has emerged as a valuable model for auditory processing at behavioral and physiological levels (Leppelsack, 1974; Leppelsack and Vogt, 1976; Leppelsack, 1983; Hausberger *et al.*, 2000; Gentner *et al.*, 2001; Gentner and Margoliash, 2002; George *et al.*, 2003; Gentner, 2004; Cousillas *et al.*, 2005;

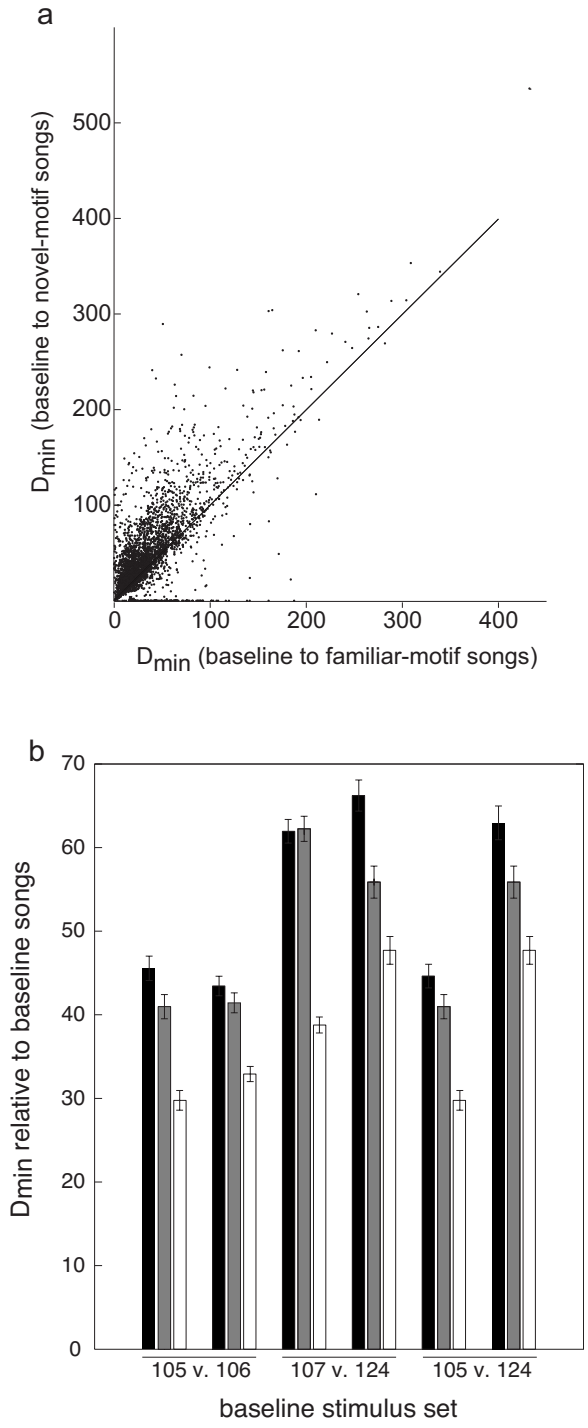


FIG. 5. Submotif feature similarity. Similarity expressed as the minimum DTW path (D_{min}; see methods) between submotif features. (a) Similarity between each submotif feature in the baseline training songs compared to those in the novel-motif test songs and to those in the familiar-motif test songs. The displacement of the distribution above the unity line reflects the fact that the features in the familiar-motif songs are more similar to the baseline submotif features, than are those in the novel-motif songs. (b) Mean similarity ($D_{min} \pm$ sd) among songs within each of the three different versions of the training and test stimuli. For a given stimulus set (e.g., "105 vs 106") the two numbers refer to the singers from which the songs were drawn and thus that the subjects learned to recognize (see methods). The three bars for each singer show the similarity of the training (baseline) songs from that bird relative to the training songs from its paired singer (black), relative to the same singer's novel-motif songs (gray), and relative to the same singer's familiar-motif songs (white). Note that the gray and white bars for a given singer are always the same, but are replotted to facilitate easy comparisons within each stimulus set.

George *et al.*, 2005; Gentner *et al.*, 2006). To date, however, studies of song recognition have relied on relatively long timescale manipulations of motif sequences, over several hundreds to thousands of milliseconds, to control recognition. The present results address how temporal patterning at the submotif level, in the range of tens to hundreds of milliseconds, effects the recognition of individual starling songs. We show that song recognition suffers significant impairments when the temporal sequencing of submotif level acoustic features is permuted, and that the effects of altering the sequencing of submotif features are restricted to familiar, i.e., learned, motifs. These results support two main conclusions. First, learned temporal patterns of submotif features, i.e., motifs, are perceptually salient auditory objects that emerge without explicit reinforcement, and which are positioned within a hierarchy of acoustic patterns. Second, starlings can readily access information at multiple levels within the acoustic pattern hierarchy for individual vocal recognition.

A. Auditory objects and song

The concept of an auditory object has received substantial research attention in recent years (Griffiths and Warren, 2004 for review), but remains somewhat controversial as a framework for understanding auditory perception. Often times, the term "auditory object" is used in the context of scene analysis (Bregman, 1990) to refer to the mental representation of an environmental sound source rather than the source itself or the sound it produces (e.g. Alain and Arnott, 2000). More generalized examples of auditory objects are common in the words and syllables that make up human language (Warren and Bashford, 1993; Warren, 1999). Thus, the concept of an auditory object has broad intuitive appeal. Indeed, by analogy to vision where the notion of an object is more intuitive (and dominant), it may be that consideration of natural auditory perception in the absence of objects is unrealistic. Pragmatically, the empirical questions concern the features of the acoustic signal that guide object structure, and the conditions under which different sets of these structural constraints hold. For example, the same acoustic communication signal may contain information about a range of relevant external events including the location of the sound source, its species, sex, individual identity, and more specific semantic content (e.g., food quality). As in vision, auditory objects are likely to exist at multiple scales spanning the relevant physical dimensions of the signal, i.e., time and spectral frequency, and features that define object boundaries in one context may be irrelevant in another. This line of reasoning suggests that a discussion of auditory objects requires explicit links to well-defined goals of the perceptual/cognitive system under investigation.

From this position, we define an auditory object as a set of coincident, or closely coincident, acoustic events that can be perceived as a whole, and that carries with it behaviorally relevant information. The results of the present study support the idea that starling song motifs form salient auditory objects of this sort. When starlings are trained to classify large sets of songs according to singer, they are significantly better

at recognizing novel songs from the training singers in which the natural sequences of submotif level features is preserved, compared to songs in which the natural sequence of submotif features is permuted. If the starlings had learned to recognize songs by attending only to salient submotif features, that is by learning sets of notes rather than sets of motifs, then permuting the sequence of notes should not have impaired recognition. Instead, the effects of the note level permutations suggest that motifs constitute auditory objects that convey a significant portion of the acoustic information required for individual song recognition. We also show that the effects of permuting the sequence submotif features are restricted to familiar, i.e., learned, motifs. Recognition of songs comprising unfamiliar motifs from the training singers was not significantly affected by the same permutations. Importantly, although the perception of these auditory objects appears dependent on learning, reinforcement was never explicitly tied to an object-based solution strategy. Thus the perceptual sensitivity to these auditory objects is an emergent product, rather than target, of learning. This emergence may reflect a parsimonious solution to the recognition of spectrotemporally complex signals and/or the effects of segmentation and phrasing constraints imposed by song production mechanisms.

The pattern of errors associated with different permutations of the learned submotif feature sequences is also consistent with an object-level perception of motifs. We found no significant differences in the submotif feature permutations that operated over the whole song and those that were confined to the boundaries of each segment's original motif. The similar level of recognition observed for both kinds of permutations—those over the motif and those over the song—indicates that the knowledge of temporal patterning within a motif is very precise, but not necessarily explicit. Had the starlings learned each motif as a perceptually explicit sequence of submotif features, one would expect local (within motif boundary) permutations to be less disruptive to recognition than permutations over the entire song stimulus. Instead, any violation of the learned temporal sequence of submotif features, at least down the to the level tested here, appears to be disruptive for recognition. Likewise, the observation of a *positive* correlation between *R* distance (our measure for permutation severity) and recognition of familiar-motif-shuffled songs runs contrary to the notion that starlings learned explicit sequences of submotif features. Had they done so, one would expect to see a negative relationship between *R* distance and performance, because stronger violations of the training sequence should be, if anything, harder to recognize. Instead, the smallest changes to the submotif pattern structure exact the greatest toll on recognition.

Therefore, we conclude that submotif feature sequences form "temporal compounds," or global auditory patterns, of the sort described for human audition (Warren and Bashford, 1993; Warren, 1999).

Exploring the precise lower bound for the length of component features within a temporal sequence is topic for future research. At some temporal scale, one might predict that sequence permutations would simply abolish all recognition. In any case, the present results demonstrate that sensitivity to temporal pattering within the boundaries of a motif extends well down into the range of tens of milliseconds. This approaches the thresholds for classic psychophysical tests of temporal integration of roughly 2 ms (Klump and Maier, 1989). Interestingly, a 10 ms lower bound to temporal order sensitivity is consistent with the optimal time window found for discrimination of patterned spike trains elicited by song in the avian analog to primary auditory cortex (Narayan *et al.*, 2006), and integration windows at higher regions in the sensory hierarchy more closely reflect timescales close to mean segment and motif duration, ~200 and 1000 ms, respectively (Thompson and Gentner, 2007). In general, the neural bases that underlie object level representation of temporally patterned acoustic sequences are unexplored. Single neuron extracellular recordings from caudo-medial mesopallium (CMM), a region in the songbird forebrain analogous to mammal auditory cortex show strong responses to motifs in the songs that birds have learned to recognize (Gentner and Margoliash, 2003). This work provides a phenomenological correlate to such objects, and suggests that starlings and other songbirds will be useful organisms within which to explore these issues.

Although the strongest recognition is reserved for songs comprising naturally ordered renditions of familiar motifs, motif percepts do not appear to be holistic in the strictest sense of that term (see Warren, 1999). In the extreme, a holistic object percept is one that cannot be recognized by any one of (or any sequentially permuted subset of) its constituent parts, and the present pattern of results fails to meet this definition. Subjects suffered significant impairment when the sequencing of submotif features in familiar-motifs songs was permuted, but recognition of these permuted songs remained reliably above chance (Fig. 4). The only explanation for above chance recognition of the permuted familiar-motif songs is that all of the task-relevant diagnostic information was not abolished by the submotif permutation. Here again, as for motif level recognition, the effects of learning are evident as even the permuted versions of familiar-motif songs were more easily recognized than any of the unfamiliar-motif songs, including those with naturally ordered submotif features. Thus, the submotif features themselves, as well as their temporal patterning, are learned, and starlings have access to both levels of acoustic information when making classification judgments that reflect individual singer identity. It is unclear whether one or both temporal scales can carry information beyond individual identity (e.g. Hausberger *et al.*, 1995; Hausberger, 1997). The "classic" view on talker recognition in humans holds that individual identity information is coded separately in non-linguistic components of the vocal signal (Bricker and Pruzansky, 1976), but phonetic components can also contribute to talker recognition (Sheffert *et al.*, 2002). It is interesting to speculate that the presence of perceptually separable temporal scales in a non-human vocal communication signal, as we have shown here, may set the stage for increasingly complex forms of encoding observed in human vocal communication.

B. Assessing auditory object similarity

Taken together, the results of these and earlier behavioral studies (Gentner, 2004) suggest that when starlings learn to recognize conspecific songs from different singers, they memorize large numbers of unique motifs corresponding to each individual singer. However, although the temporal organization of submotif features plays a clear role in guiding song recognition, information coded at the level of the motif cannot explain song recognition entirely. Even when motif identity and the corresponding submotif temporal structure are abolished, recognition is above chance, albeit only modestly, at a severely impaired level (Fig. 4). This result conflicts slightly with previous results showing that recognition falls to chance when starlings are presented with song composed entirely of novel motifs sung by otherwise familiar singers (Gentner *et al.*, 2000). One possible explanation for this difference in recognition of novel motif songs is methodological. Previous studies used a slightly different training protocol in which the subject's task was to give one response to songs of a single male and another response to songs from four other males. Compared to the one-versus-one design used in the present study, the one-versus-many version of the recognition task necessarily involves lower motif overlap within the set of "many" songs. The lower motif overlap, in turn, likely makes the previous task more difficult which may mask the very subtle recognition of unfamiliar-motif songs observed in the present study. The different studies also drew subjects from different populations of starlings, and it is theoretically possible that perceptual sensitivity to the features that permit the modest recognition of novel motifs from familiar singers varies across populations.

Regardless of the source, starlings are able to recognize songs comprising unfamiliar motifs from otherwise familiar singers. The fact that natural motif organization provided no measurable improvement to the recognition of these unfamiliar-motif songs indicates that the information used to achieve this marginal recognition is carried by (and only by) the submotif features. This information may reflect true "voice characteristics," i.e., singer-invariant acoustic properties, imparted to all or a subset of the motifs sung by a single individual. To examine the features that guide the very subtle recognition of novel motifs one can look to the similarity measures employed in the present study. Unique motifs may result from individually specific groupings of submotif features that are shared among all starlings or from submotif features that are themselves specific to each individual. If submotif features are shared between birds, then two sets of novel motifs from one singer should be approximately "as similar" to a third set of motifs from another singer as the features making up all three sets of motifs are drawn from a common pool. If, however, submotif features are specific to each individual, then two distinct sets of motifs from the same singer may be more similar than two sets of motifs from different singers, as a bird may be more likely to use the same submotif feature in more than one motif. We found that for each of the three different stimulus sets, disjoint sets of motifs from the same singer were significantly more similar than sets of motifs from different singers. This suggests that submotif features are specific to each individual bird, and may be taken as evidence for a weak form of "voice characteristic" imparted to a least a subset of the notes that a bird produces. The use of voice characteristics (e.g., vocal timbre, the frequency of glottal pulsation, and spectral contours imparted by laryngeal morphology) is well documented for individual talker recognition in humans (Bricker and Pruzansky, 1976). A more precise understanding of the information that starlings used to recognize the unfamiliar-motif songs awaits further investigation.

Although starling song recognition is guided largely by the memorization of unique notes and their temporal patterns, the presence of and perceptual sensitivity to voice characteristics in any nonhuman animal, however subtle, is important theoretically. For humans, the rich semantic content of our language necessitates that most words are shared among speakers, and so precludes a "repertoire memorization" strategy for individual recognition. Instead, the voice characteristics used in speaker recognition appear to be coded in acoustic parameters of the signal that are predominantly non-linguistic (Remez *et al.*, 1997). Our results indicate that an independent communication channel exists in at least one other species. It is interesting to speculate that its subtle role in songbird vocal recognition may represent an unexploited, and unnecessary, capacity of vocal communication signals that lack a rich combinatorial semantic structure.

ACKNOWLEDGMENTS

Grant No. R01DC008358-01A1 from NIH NIDCD supported this research. The author thanks Satoru Fukushima for assistance in programming a portion of the DTW algorithm.

Adret-Hausberger, M., and Jenkins, P. F. (**1988**). "Complex organization of the warbling song in starlings," Behaviour **107**, 138–156.

Alain, C., and Arnott, S. R. (**2000**). "Selectively attending to auditory objects," Front. Biosci. **5**, D202–212.

Anderson, S., Dave, A., and Margoliash, D. (**1996**). "Template-based automatic recognition of birdsong syllables from continuous recordings.," J. Acoust. Soc. Am. **100**, 1209–1219.

Bregman, A. S. (**1990**). "The auditory scene," in *Auditory Scence Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge).

Bricker, P. D., and Pruzansky, S. (**1976**). "Speaker recognition," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic, New York), pp. 295–326.

Chaiken, M., Böhner, J., and Marler, P. (**1993**). "Song Acquisition in European Starlings, Sturnus vulgaris: a comparison of the songs of live-tutored, tape-tutored, untutored, and wild-caught males," Anim. Behav. **46**, 1079–1090.

Cousillas, H., Leppelsack, H. J., Leppelsack, E., Richard, J. P., Mathelier, M., and Hausberger, M. (**2005**). "Functional organization of the forebrain auditory centres of the European starling: a study based on natural sounds," Hear. Res. **207**, 10–21.

Eens, M. (**1992**). "Organization and functin of the song in the European starling *Sturnus vulgaris*," University of Antwerp, Thesis.

Eens, M. (**1997**). "Understanding the complex song of the European starling: An integrated approach," Advances in the Study of Behavior **26**, 355–434.

Eens, M., Pinxten, M., and Verheyen, R. F. (**1989**). "Temporal and sequential organization of song bouts in the European starling," Ardea **77**, 75–86.

Eens, M., Pinxten, R., and Verheyen, R. F. (**1991**). "Organization of Song in the European Starling—Species—Specificity and Individual-Differences," Belg. J. Zoolog. **121**, 257–278.

Gentner, T. Q. (**2004**). "Neural systems for individual song recognition in adult birds," Ann. N.Y. Acad. Sci. **1016**, 282–302.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (**2006**). "Recursive syntactic pattern learning by songbirds," Nature (London) **440**, 1204–1207.
Gentner, T. Q., and Hulse, S. H. (**1998**). "Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris," Anim. Behav. **56**, 579–594.
Gentner, T. Q., and Hulse, S. H. (**2000**). "Perceptual classification based on the component structure of song in European starlings," J. Acoust. Soc. Am. **107**, 3369–3381.
Gentner, T. Q., Hulse, S. H., Bentley, G. E., and Ball, G. F. (**2000**). "Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds," J. Neurophysiol. **42**, 117–133.
Gentner, T. Q., Hulse, S. H., Duffy, D., and Ball, G. F. (**2001**). "Response biases in auditory forebrain regions of female songbirds following exposure to sexually relevant variation in male song," J. Neurophysiol. **46**, 48–58.
Gentner, T. Q., and Margoliash, D. (**2002**). *The Neuroethology of Vocal Communication: Perception and Cognition* (Springer, Berlin).
Gentner, T. Q., and Margoliash, D. (**2003**). "Neuronal populations and single cells representing learned auditory objects," Nature (London) **424**, 669–674.
George, I., Cousillas, H., Richard, J. P., and Hausberger, M. (**2003**). "A new extensive approach to single unit responses using multisite recording electrodes: application to the songbird brain," J. Neurosci. Methods **125**, 65–71.
George, I., Cousillas, H., Richard, J. P., and Hausberger, M. (**2005**). "State-dependent hemispheric specialization in the songbird brain," J. Comp. Neurol. **488**, 48–60.
Griffiths, T. D., and Warren, J. D. (**2004**). "What is an auditory object?," Nat. Rev. Neurosci. **5**, 887–892.
Hausberger, M. (**1997**). "Social influences on song acquisition and sharing in the European starling (*Sturnus vulgaris*)," in *Social Influences on Vocal Development*, edited by C. Snowden and M. Hausberger (Cambridge University Press, Cambridge), pp. 128–156.
Hausberger, M., and Cousillas, H. (**1995**). "Categorization in birdsong: From behavioural to neuronal responses," Behav. Processes **35**, 83–91.
Hausberger, M., Leppelsack, E., Richard, J., and Leppelsack, H. J. (**2000**). "Neuronal bases of categorization in starling song," Behav. Brain Res. **114**, 89–95.
Hausberger, M., Richard-Yris, M.-A., Henry, L., Lepage, L., and Schmidt, I. (**1995**). "Song sharing reflects the social organization in a captive group of European starlings (*Sturnis vulgaris*)," J. Comp. Psychol. **109**, 222–241.
Klump, G. M., and Maier, E. H. (**1989**). "Gap detection in the starling (*Sturnus vulgaris*): I Psychophysical thresholds," J. Comp. Physiol., A **164**, 531–538.
Kogan, J. A., and Margoliash, D. (**1998**). "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study," J. Acoust. Soc. Am. **103**, 2185–2196.
Kroodsma, D. E. (**1989**). "Pseudoreplication external validity and the design of playback experiments," Anim. Behav. **38**, 715–719.
Leppelsack, H.-J. (**1974**). "Functional Properties of the Acoustic Pathway in the Field L of the Neostratum caudale of the Starling," J. Comp. Physiol. **88**, 271–320.
Leppelsack, H.-J. (**1983**). "Analysis of song in the auditory pathway of song birds," in *Advances in Vertebrate Neuroethology*, edited by J. P. Evert, B. R. Capranica, and D. J. Ingle (Plenum, New York), pp. 783–799.
Leppelsack, H. J., and Vogt, M. (**1976**). "Response to auditory neurons in the forebrainof a song bird to stimulation with species-specific sounds," J. Comp. Physiol. **107**, 263–274.
Narayan, R., Grana, G., and Sen, K. (**2006**). "Distinct time scales in cortical discrimination of natural sounds in songbirds," J. Neurophysiol. **96**, 252–258.
Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identification based on phonetic information," J. Exp. Psychol. Hum. Percept. Perform. **23**, 651–666.
Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. E. (**2002**). "Learning to recognize talkers from natural, sinewave, and reversed speech samples," J. Exp. Psychol. Hum. Percept. Perform. **28**, 1447–1469.
Thompson, J. V., and Gentner, T. Q. (**2007**). "Temporal- and rate-coding schemes, and the emergence of recognition for complex acoustic communication signals," in *Society for Neuroscience* (Society of Neuroscience Abstracts, San Diego, CA).
Warren, R. M. (**1999**). *Auditory Perception: A New Analysis and Synthesis* (Cambridge University Press, New York).
Warren, R. M., and Bashford, J. A. (**1993**). "When acoustic sequences are not perceptual sequences: The global perception of auditory patterns," Percept. Psychophys. **54**, 121–1261.

The inner ears of Northern Canadian freshwater fishes following exposure to seismic air gun sounds

Jiakun Song[a)]
Department of Biology, Neuroscience and Cognitive Science Program, and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742 and Institute for Marine Biosystem and Neurosciences, Shanghai Fisheries University, Shanghai 200090, China

David A. Mann[b)]
College of Marine Science, University of South Florida, 140 7th Avenue South, St. Petersburg, Florida 33701

Peter A. Cott[c)] and Bruce W. Hanna[d)]
Fisheries and Oceans Canada, 101, 5204-50th Avenue, Yellowknife, Northwest Territories X1A 1E2, Canada

Arthur N. Popper[e)]
Department of Biology, Neuroscience and Cognitive Science Program, and Center for Comparative and Evolutionary Biology of Hearing,University of Maryland, College Park, Maryland 20742

(Received 5 September 2007; revised 5 April 2008; accepted 28 May 2008)

An earlier study examined the effects of exposure to seismic air guns on the hearing of three species of fish from the Mackenzie River Delta in Northern Canada [Popper *et al.* (2005). "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**, 3958–3971]. The sound pressure levels to which the fishes were exposed were a mean received level of 205–209 dB re 1 μPa (peak) per shot and an approximate received mean SEL of 176–180 dB re 1 μPa^2 s per shot. In this report, the same animals were examined to determine whether there were effects on the sensory cells of the inner ear as a result of the seismic exposure. No damage was found to the ears of the fishes exposed to seismic sounds despite the fact that two of the species, adult northern pike and lake chub, had shown a temporary threshold shift in hearing studies.
 [DOI: 10.1121/1.2946702]

PACS number(s): 43.80.Lb, 43.80.Nd, 43.64.Wn [MCH] Pages: 1360–1366

I. INTRODUCTION

There is growing concern that human-generated (anthropogenic) sounds may have an impact on the health and/or survival of fishes (e.g., Myrberg, 1980; Popper, 2003; Popper *et al.*, 2004; Hastings and Popper, 2005). The types of sounds of interest vary greatly in their acoustic parameters but include noise from shipping, low- and midfrequency sonars, pile driving, and a range of other anthropogenic sources.

One of the anthropogenic sources of primary concern has been air guns used in seismic surveys. These are used extensively by the oil and gas industry for exploration and by geologists for subsea studies. Although most work with air guns has taken place in marine environments, there has been an increase in seismic exploration in rivers and lakes, and as a result, concerns have expanded to include freshwater fish species.

Air guns produce a compressed air bubble that, once released, collapses under the pressure of water. This results in a sharp concussive sound with peak sound levels possibly equal to or exceeding 230 dB (re 1 μPa) at 1 m from a single air gun. The general procedure for this type of seismic survey is to trail an air gun array behind a boat and set off frequent "shots." These sounds are directed down toward the substrate and reflect off geologic formations below the water-substrate interface. The reflected signals are picked up by hydrophones that are towed behind the vessel or laid on the bed of the waterbody (Bott, 1999).

Although it has been suggested that the sound energy from air guns may harm fishes (reviewed by Popper, 2003; Popper *et al.*, 2004), there are few studies that have directly examined the potential effects on fish physiology or behavior. McCauley *et al.* (2003) found that exposure of caged pink snapper (*Pagrus auratus*) to multiple emissions of a single seismic air gun resulted in substantial damage to the sensory cells in a small region of the saccule, one of the end organs in the inner ear of fishes. McCauley *et al.* (2003) also found that although some damage was found several hours after exposure, the damage to the sensory tissue apparently continued to increase in the days postexposure, although it never was found over more than a small portion of any saccular epithelium.

[a)]Electronic mail: jksong@umd.edu and jksong@shfu.edu.cn
[b)]Electronic mail:dmann@marine.usf.edu
[c)]Electronic mail: pete.cott@dfo-mpo.gc.ca
[d)]Electronic mail: bruce.hanna@dfo-mpo.gc.ca
[e)]Author to whom corresponence should be addressed. Tel.: 301-405-1940. FAX: 301-314-9358. Electronic mail: apopper@umd.edu

TABLE I. Experimental groups and number of animals used per group for SEM.

Species	Chub	Whitefish	Adult pike	Young pike
Baseline	2	2	3	2
5 shot	5	4	4	4
5 shot, 18/24 h	2 (18 h)	⋯	3 (24 h)	⋯
Control	4	4	4	4
20 shot	4	⋯	⋯	5
20 shot, 18 h	4	⋯	⋯	⋯

More recently, Popper *et al.* (2005) examined the effects of exposure to a 730-in^3. seismic air gun array on the hearing capabilities of several different fish species found in the Mackenzie River near Inuvik, Northwest Territories, Canada (see map by Popper *et al.*, 2005; Mann *et al.*, 2007). The investigators showed that there was some temporary hearing loss as a result of air gun exposure in lake chub (*Couesius plumbeus*) and adult (but not young of the year) northern pike (*Esox lucius*) but no effects on hearing in broad whitefish (*Coregonus nasus*). All three fishes are freshwater species and represent the first nonmarine species tested for effects from seismic air gun noise. In addition, the broad whitefish is an important subsistence fish in the region, and there is concern that damage to this species could have impacts to subsistence harvesting.

This paper extends the work of Popper *et al.* (2005) by examining the inner ear tissues of the same specimens that were exposed to air gun sounds in that study. In addition, since the ears of these three species have not yet been described, their anatomy and ultrastructure are presented briefly.

II. METHODS

Animals were obtained from the Mackenzie River and held at the Fisheries and Oceans Canada facilities in Inuvik, Northwest Territories. Fish capture, holding methodology, and seismic exposure were described elsewhere (Popper *et al.*, 2005). Fish were exposed to 5 or 20 shots from a 730-in^3. air gun array to emulate an exposure comparable to what fish would experience from a seismic survey in a river (see Table I). The mean received sound pressure levels of sounds to which the fishes were exposed for each shot were from 205 to 209 dB re 1 μPa (peak), which was equivalent to an approximate mean received level per shot sound exposure level (SEL) of 176–180 dB re 1 μPa^2 s. Particle motion levels of the received sound are presented by Popper *et al.* (2005).

Fishes were exposed to air guns and then examined for hearing capabilities prior to the ears being fixed for ultrastructural examination (Popper *et al.*, 2005). Ears were examined from each exposed species (broad whitefish, lake chub, and adult and young northern pike). The experimental groups and number of animals in each group are presented in Table I.

Within 15 min following hearing tests, the fish were euthanized with an overdose of buffered tricaine methane sulfonate (MS-222). The heads were quickly opened and fixative (4% gluteraldehyde and 2.5% paraformaldehyde in buffer, pH 7.4) was squirted into the cranial cavity. The brain was then carefully removed and the extraneous tissue was dissected away, and at all times, the ear region was kept wet with fixative. The ear regions were then placed in jars of cold fixative and shipped overnight to the University of Maryland where the ears were dissected and examined.

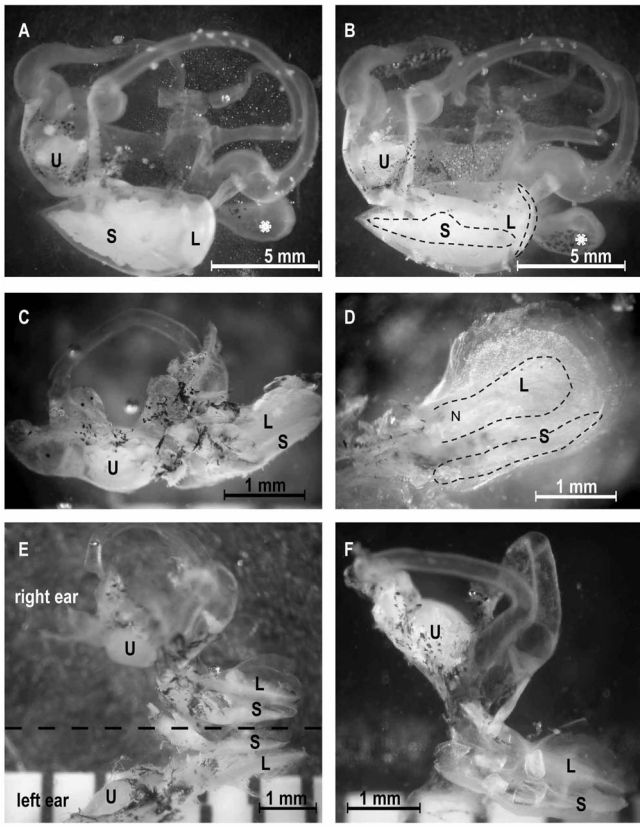


FIG. 1. Light micrographs of the ears of unexposed northern pike and lake chub. In each case, anterior is to the left and dorsal to the top. Lateral (A) and medial (B) views of the left ear of northern pike (note image of the medial view is electronically mirrored so that anterior remains to the left). In addition to having three otolithic organs, the lagena (L), saccule (S), and utricle (U), each ear has three semicircular canals and three sensory areas or ampullae associated with the canals [(note that asterisk at the posterior end of the saccule is a chamber that has only been seen in the northern pike ear—see text) (A) and (B)]. The saccular epithelium is outlined with a dashed line in (B). (C) Lateral view of the left ear of lake chub. (D) Medial view of the saccule and lagena of the left ear of lake chub (electronically mirrored). Dashed lines indicate the locations of the sensory epithelia of the lagena and saccule, and the innervating eighth cranial nerve (N). (E) Dorsal view of both ears of the lake chub (brain removed to expose the ears which lie below it). The dashed line is down the midline of the fish. Note how close the two ears are to one another. Dorsomedial view of the right ear of the lake chub (F) with the saccular otolith broken into two pieces. Note that the saccule and lagena of the lake chub, an otophysan, are different from the other two species in this study and typical of all otophysans.

Once at the University of Maryland, the ears were then removed from the heads and photographed in a dissection microscope, and the otoliths were then removed. The tissue was then washed in three changes of phosphate-buffered saline (PBS) (0.1*M*, pH 7.4), postfixed in 1% osmium tetroxide (in PBS), and further dissected to reveal the sensory epithelia of the three otolithic end organs, the saccule, lagena, and utricle (Fig. 1). The tissue was then dehydrated through a graded ethanol series (30%, 50%, 70%, 80%, 90%, and 95%)

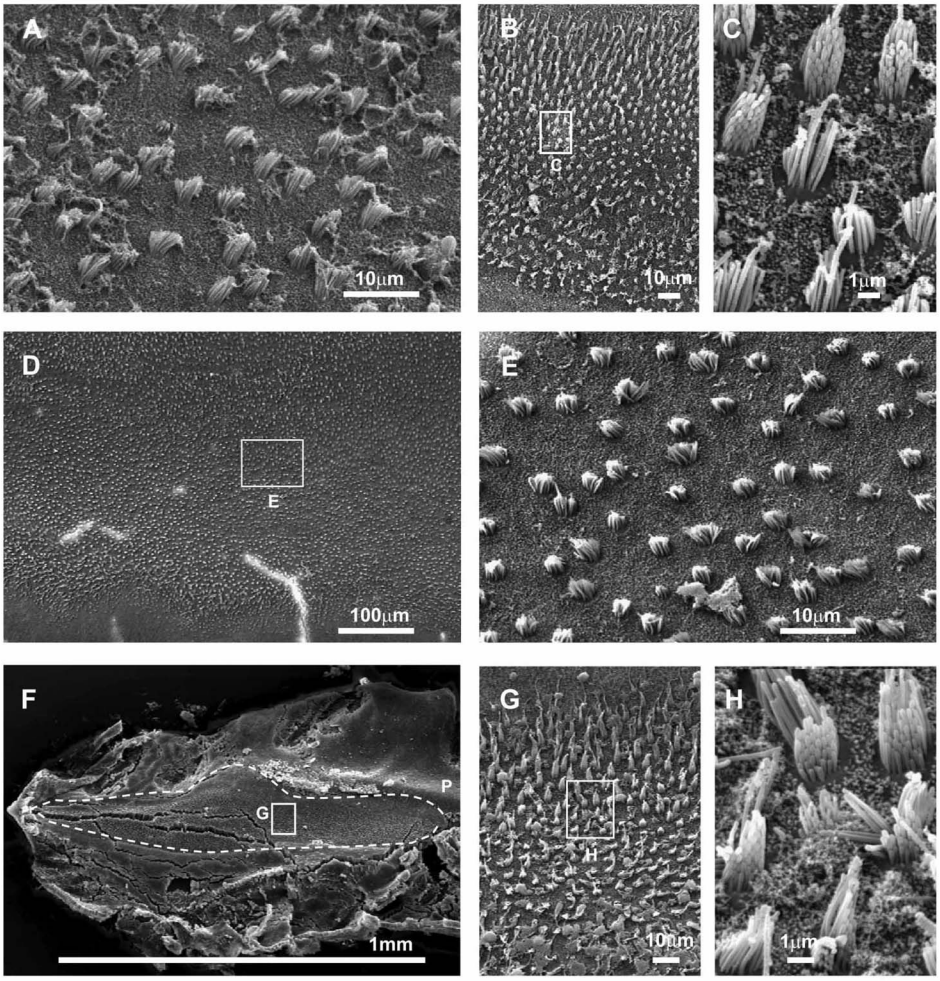


FIG. 2. Scanning electron micrographs of the saccular sensory epithelia of adult and young northern pike. (A) Adult pike 41E-baseline control. There was no damage to tissue in control and baseline. [(B) and (C)] Young pike 88C-control. (C) High magnification of a region in (B). [(D) and (E)] Adult pike 73E-exposed to 5 shots and then held 24 h before hearing tests were conducted and fish was sacrificed. (D) Low-magnification image showing that the field of ciliary bundles is intact and there is no indication of damage. (E) High magnification of boxed area in (D) showing the intact nature of the ciliary bundles. [(F)–(H)] Young pike 94E-exposed to 20 shots and then tested for hearing and sacrificed within 1 h. (F) Scanning electron micrograph showing that the whole epithelium (dashed area) is intact (anterior to the left, dorsal to the top). [(G)–(H)] Successively higher magnifications showing no effect of seismic exposure. Compare (E)–(G) with comparable images in controls of (B) and (C). It should be noted that there are cracks in various tissues as in (F). However, these cracks are widely seen in SEM tissue of fish ears (e.g., Popper, 1977) and are artifacts of the fixation and drying process and do not represent effects of noise exposure.

and up to three changes of 100% ethanol. The tissue was subsequently critical point dried using liquid CO_2 as the intermediary fluid, mounted on aluminum stubs, coated with a 25-nm-thick layer of gold/palladium (60:40) using a Denton DV503 device, and viewed with an AMRAY 840 scanning electron microscope.

In scanning electron microscopy, all regions of the sensory epithelia of the otolithic organs from both ears were examined at 2000–5000× for all fish from each treatment group. Low- and high-power electron micrographs were taken to document tissue from different epithelial regions. While tissue from the saccules was always usable, it was not always possible to get useful information from the lagena of the northern pike or the utricle of the broad whitefish due to poor quality tissue since the epithelia surfaces could not be easily seen. The same problem arose in all specimens of these species including baseline controls, controls, and experimental animals, suggesting some kind of fixation problem.

Tissue was examined for substantive differences between exposed and control animals because there were no predetermined criteria for what might constitute damage. Moreover, the types of damage that might have been expected ranged from extensive areas of missing ciliary bundles on sensory hair cells to regions in the epithelia where there were "holes" from which sensory hair cells had been lost. These types of damage have been seen in earlier studies (e.g., Enger, 1981; Hastings *et al.*, 1996; McCauley *et al.*, 2003).

III. RESULTS

A. General ear structure

The ears of fishes include three semicircular canals and three otolith organs, the saccule, lagena, and utricle (Fig. 1). Each of the otolithic organs has a single dense calcareous otolith that lies adjacent to the sensory epithelium. The epithelium contains a large number of sensory hair cells (Figs. 2–6) from which arise ciliary bundles that are made up of a single kinocilium and multiple stereocilia. The ciliary bundles project into the lumen of the otolithic chamber so that the tips of the cilia come close to or contact the otolith.

The ears of the northern pike [Figs. 1(a) and 1(b)] and broad whitefish are typical of species that do not have specializations for hearing in that the saccule is much larger than the lagena (e.g., Popper *et al.*, 2003; Song *et al.*, 2006). The only unique feature of the ears is the presence of a small sac, of unknown function, associated with the ear in both young and adult northern pike [Figs. 1(a) and 1(b)].

In contrast to the northern pike and broad whitefish, the lake chub [Figs. 1(c)–1(f)] is classified as "hearing specialist" because it has a series of bones, the Weberian ossicles, that connect the swim bladder to the inner ear. The lake chub lagena is substantially larger than the saccule and located

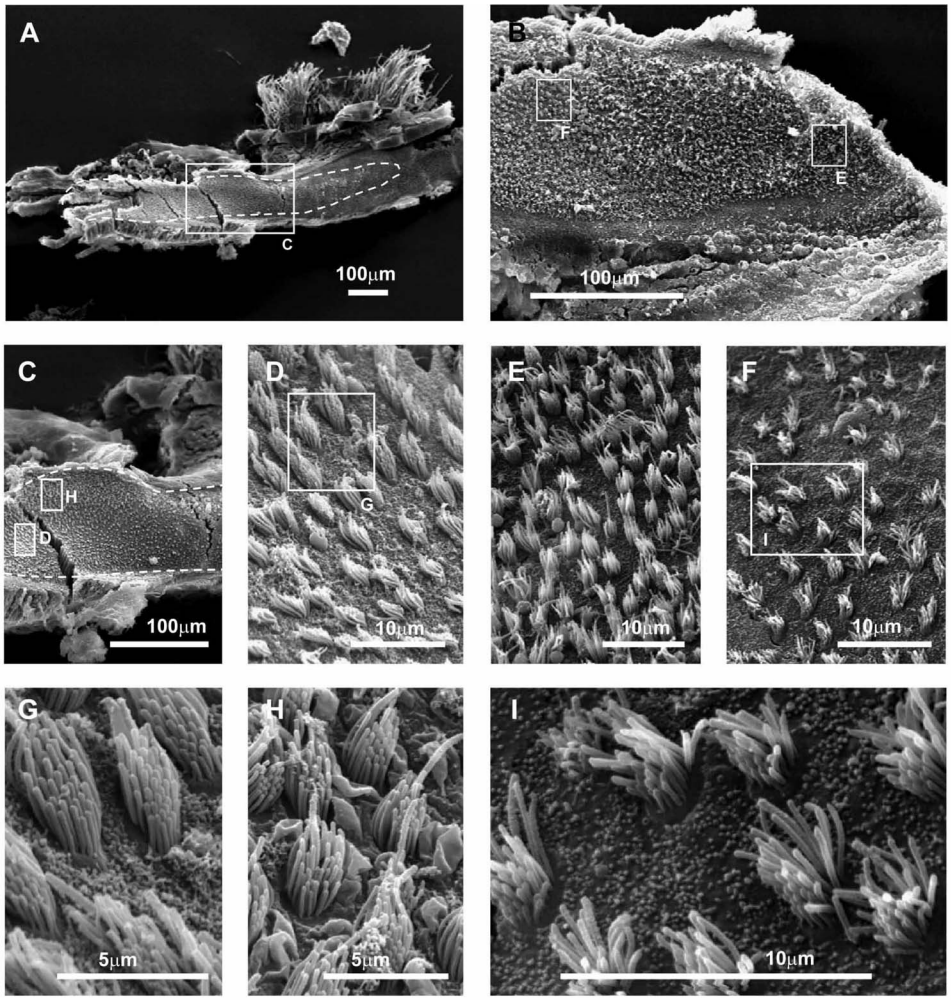


FIG. 3. Scanning electron micrographs of the saccule from an experimental lake chub. Left column [(A), (C), (D), (G), and (H)] Animal 80E-exposed to 20 seismic shots and sacrificed 18 h later. (A) Whole left saccule (anterior to the left, dorsal to the top) showing outline of epithelial area (dashed line). (C) High magnification of anterior region. (D) An even higher magnification of a region shown in the square in (C). [(G) and (H)] Higher magnifications from regions shown in (C) and (D). The cells in (H) are more peripherally located than those in (G). Right column [(B), (E), (F), and (I)] Lake chub 60E-exposed to 5 shots and sacrificed 1 h later. (B) Low power of anterior region of the right saccular epithelium. [(E) and (F)] Higher magnification from the central region of (B). (I) Higher magnification of a portion of (F).

medial and slightly posterior to the saccule [Figs. 1(c) and 1(d)]. The ear of the broad whitefish is very similar to several other salmonids described earlier (Popper, 1976, 1977), and the shape and topographic relationship of saccule and lagena [Fig. 5(a)] are similar to those in the northern pike.

The ears of all species studied lie in the cranial cavity just lateral to the posterior part of the brain [Figs. 1(e) and 1(f)]. Each of the sensory epithelia (including those of the semicircular canals) is innervated by branches of the eighth cranial nerve that then projects into the medulla of the brain.

B. Effects of air gun exposure on experimental fishes

Analysis of each fish showed no damage to any of the sensory tissues of the three otolithic organs (Figs. 2–6). Because there was no apparent damage to any of the fishes, the micrographs presented here were chosen to demonstrate the lack of damage in each species exposed to emissions from air guns, the different numbers of shots to which fish were exposed, and the different time intervals postexposure that fishes were allowed to survive. All sensory epithelia were intact and no differences were observed between baseline controls, controls, and the exposed groups. (For examples of what damage might look like, readers are referred to Enger 1981; Hastings *et al.*, 1996; and McCauley *et al.*, 2003.)

Data for the saccule of northern pike are shown in Fig. 2 for a fish exposed to 5 shots and then sacrificed 1 or 24 h after exposure and for the utricle of young fish exposed to 20 shots [Figs. 6(c) and 6(d)] (Table I). There were no differences in the structure of the sensory epithelia in the experimental fish compared to control and baseline control fish.

Similarly, there was no apparent damage to sensory cells in the lake chub as compared to baseline and control animals [saccule shown in Fig. 3, lagena in Fig. 4, and utricle in Figs. 6(a) and 6(b)]. Moreover, there were no apparent effects in lake chub exposed to 20 shots and allowed to survive for 18 h (e.g., Figs. 2–4 right column) or to fish that received 5 shots and were allowed to survive for 1 h postexposure (Figs. 2–4 left column). Similarly, there was no effect on the utricle of lake chub exposed to 20 shots [Figs. 6(a) and 6(b)].

Finally, there was no damage seen to the saccule and lagena of the broad whitefish exposed to five shots (Fig. 5). No data are available for the utricle.

IV. DISCUSSION

This study examined the gross structure of the ear of three species of freshwater fishes from northern Canada that had been exposed to seismic air guns. The ears of the three species show no distinct differences from the ears of other species studied to date (e.g., Popper *et al.*, 2003; Ladich and

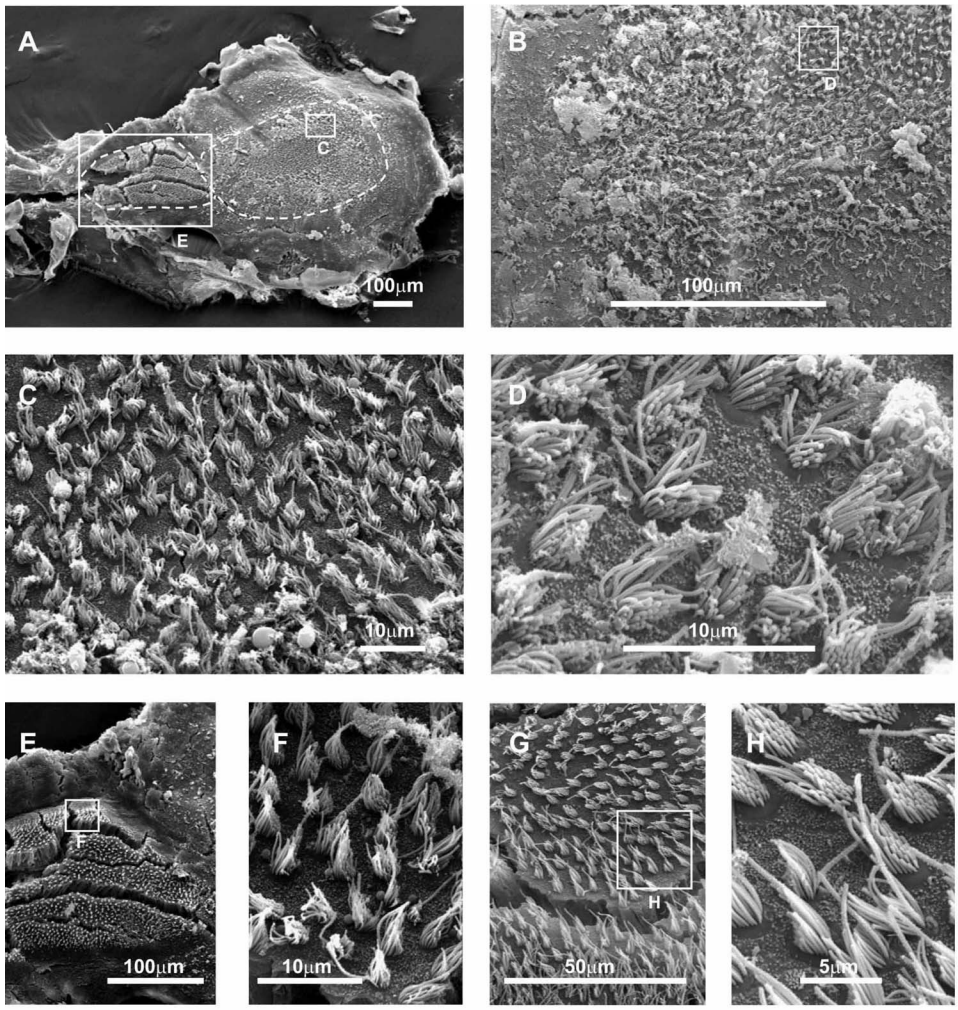


FIG. 4. Scanning electron micrographs of the lagena of experimental lake chub. Left column [(A), (C), (E), and (F)] Chub 60E-exposed to 5 shots and sacrificed 1 h later. (A) Whole lagena epithelium, anterior to the left and dorsal to the top. Dashed line shows outline of the lagenar epithelium. (C) Higher magnification of the posterior region. [(E) and (F)] Anterior region of the epithelium showing both low (E) and high (F) magnifications to demonstrate that the tissue shows no damage resulting from seismic exposure. Right column [(B), (D), (G), and (H)] Lake chub 80E-exposed to 20 shots and sacrificed 18 h later. (B) Low-power image of the posterior region. (D) Higher power of the central region of (B). [(G) and (H)] Higher magnifications of the anterior region. There is no evidence of any damage in this tissue.

Popper, 2004; Song *et al.*, 2006) except for a small sac, of unknown function, associated with the ear of the northern pike [Figs. 1(a) and 2(b)]. There is no evidence of any damage to the surfaces and ciliary bundles of the sensory cells of the otolithic organs of any of the species.

A. General ear anatomy

The lake chub has an ear typical of other members of its group, the Otophysi (Popper and Platt, 1983), and is a hearing specialist based on both ear structure and hearing capabilities (Popper *et al.*, 2005). The ear structures support the suggestion that both the northern pike and broad whitefish are hearing "generalists" since neither species has characteristics in gross or ultrastructure that might be found in any hearing specialist (e.g., Popper *et al.*, 2003; Ladich and Popper, 2004). While the function of the small sac associated with the ear of the northern pike is not known, no connection to the swim bladder was observed, and so it is more likely that the chamber is filled with endolymph from the ear than with air. Moreover, hearing data for this species (Mann *et al.*, 2007) supported an argument that the northern pike does not have the hearing range or sensitivity one would expect in a species where there is an air bubble associated with the ear.

B. Effects of seismic exposure

The results show that there was no damage to the sensory epithelia studied in any of the otolithic end organs of any of the three fish species exposed to seismic air guns, including lake chub and adult northern pike held up to 18–24 h postexposure (Table I). In particular, there was no damage to the saccular sensory epithelia, the otolithic end organ of fish thought to be most involved in hearing (Popper *et al.*, 2003). The examined tissues from all exposed fish showed that there were no differences from tissues of controls that were placed in the exposure apparatus but not exposed to air gun noise or from fish that were just kept in holding cages and not moved to the test apparatus. At the same time, both adult northern pike and lake chub exhibited temporary hearing loss (temporary threshold shift; see Popper *et al.*, 2005), showing that hearing loss in fishes is not necessarily accompanied by morphological effects on the sensory hair cells, at least at the level of scanning electron microscopy and at least for the time duration postexposure used in this study.

In contrast to the findings here, several earlier studies have reported damage to sensory epithelia in fishes exposed to pure tones (Enger, 1981; Hastings *et al.*, 1996) and a seismic source (McCauley *et al.*, 2003). In each case, the tissue showed ciliary bundles sheared away from the surfaces of

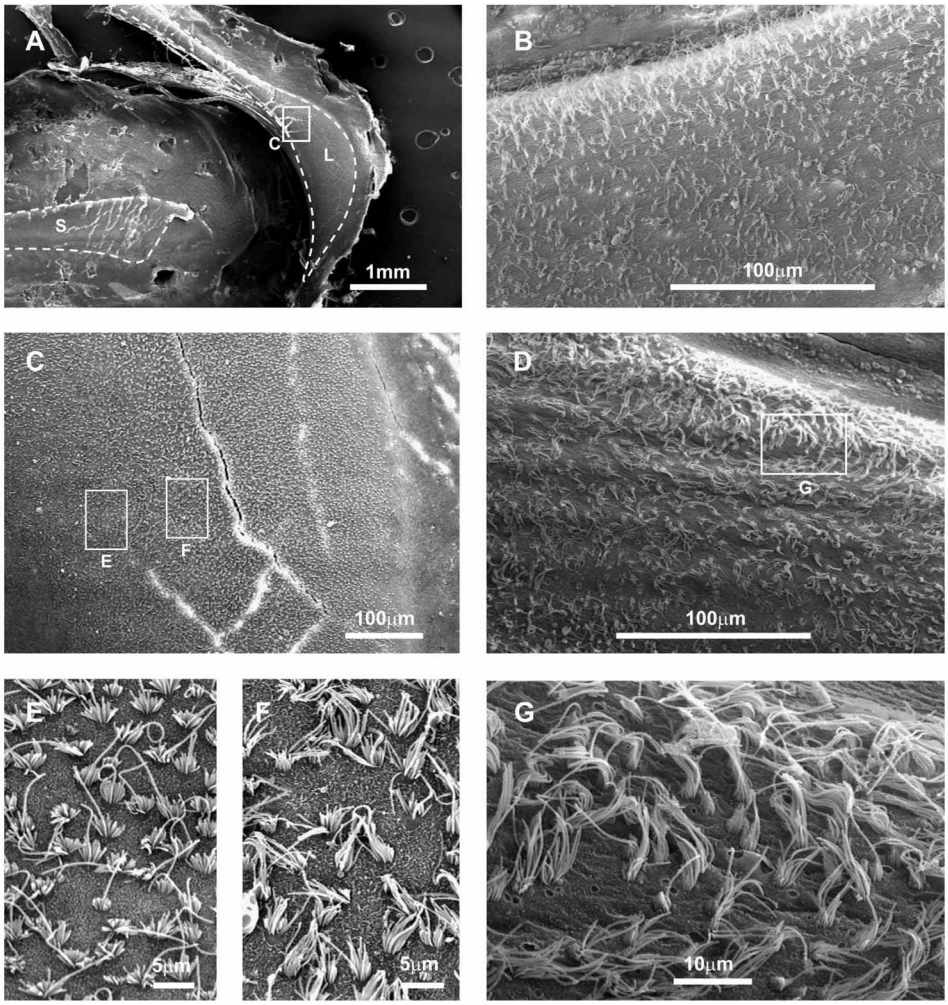


FIG. 5. Scanning electron micrographs of the lagena of broad whitefish. Left column [(A), (C), (E), and (F)] Broad whitefish 23C-control. (A) Low-power image of the posterior region of the left saccular epithelium and the left lagena (dashed outlines). (C) High magnification of area in the lagena shown in (A). [(E) and (F)] Higher magnifications of (C), with (E) being more peripheral and (F) more central. Right column [(B), (D), and (G)] Broad whitefish 35E-exposed to 5 shots and sacrificed within 1 h. (B) Right lagena. (D) Left lagena. (G) High magnification of (D). There was no damage to any of this tissue.

the sensory cells and/or holes in the epithelia where sensory cells apparently died or were "blown out of" the tissue. None of these effects were seen in the tissues reported here.

These results are in notable contrast to the findings of McCauley *et al.* (2003) that showed substantial morphological damage to the ears of pink snapper. However, as pointed out earlier (Popper *et al.*, 2005), there are significant differences between these two seismic studies including air gun size, the number of air guns, operating pressure to the guns, the sound exposure and recovery time of fishes, and the environment in which the study was conducted. Another difference was the shallower freshwater environment of the Mackenzie River (average of about 1.9 m; Popper *et al.*, 2005) compared with the deeper marine environment of the harbor (Jervoise Bay) in Perth, Australia (average 9 m; McCauley *et al.*, 2003). Due to shallow-water propagation characteristics at the location of the fish cage, there was relatively less low-frequency acoustic energy but more high-frequency acoustic energy in the present study than in the Perth investigation (see power spectra by Popper *et al.*, 2005 and McCauley *et al.*, 2003). It is possible that lower frequency sounds are more likely to elicit damage than higher frequencies, although there are no data to support such an argument. Also, McCauley *et al.* (2003) found maximum loss of sensory tissue at 54 days postexposure. In contrast, we were only able to examine several species held no more than 24 h postexposure. However, McCauley *et al.* (2003) also looked at the tissue 18 h postexposure, and although there was no damage at a level that was statistically different from that in control tissue (probably due to small sample size), it was clear that the 18-h tissues were qualitatively very different from those in control tissues. This was not the case in our study where tissues from fish sacrificed postexposure were no different from control and tissues.

ACKNOWLEDGMENTS

This study was supported by Fisheries and Oceans Canada and Indian and Northern Affairs Canada, the Program of Energy Research and Development (PERD), the Inuvialuit Fisheries Joint Management Committee (FJMC), and WesternGeco. Additional support was provided by Grant No. P30 DC-04664 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. Numerous people helped with various aspects of this project: Dave Tyson, Marty Bergmann, Don Cobb, Ron Allen, and the DFO Inuvik office staff; Steve Whidden from WesternGeco; Les Harris from the Gwich'in Renewable Resource Board; Kevin Bill, Andrea Hoyt, and the FJMC mentoring program students Gerald Kisoun, Noel Cockney, Candice Cockney, and Angus Alunik; Edward Dillon; Merik Allen; and Ms. Moe Grant who permitted us to use her prop-

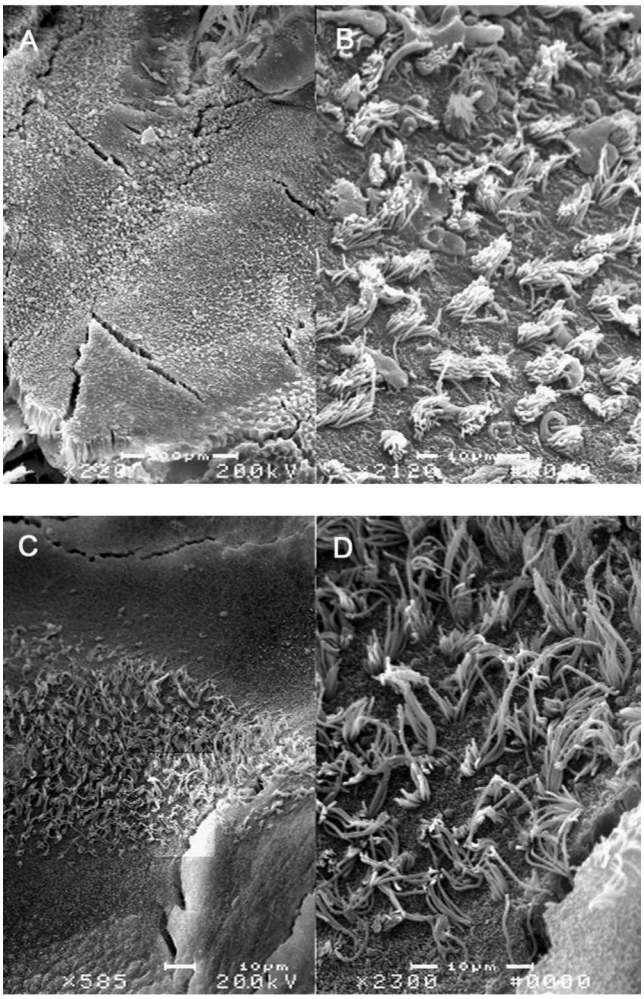


FIG. 6. Scanning electron micrographs of utricles of exposed fish. (A) Lake chub 80E exposed to 20 shots at low magnification. (B) Same tissue at higher magnification. (C) Young pike 94E exposed to 20 shots at low magnification. (D) Same tissue at higher magnification. There was no damage to any of this tissue as compared to controls and baseline animals. The poor quality of fixation was found in the utricles of all northern pike used in these experiments.

erty for this study. We thank Dr. Elena Sanovich for considerable help with the images and the plates and Helen Popper for editorial review of the manuscript. Finally, we thank Dr. Mardi Hastings and several excellent reviewers for valuable comments on earlier versions of the manuscript. This study was approved by the Fisheries and Oceans Canada Animal Care Committee and conducted in accordance with Animal Use Permit No. 04-05-009, Scientific Research License No. 13608, and License to Collect Aquatic Plants, Animals and Fish for Scientific Purposes No. SLE-04/05-213.

Bott, R. (**1999**). *Our Petroleum Challenge: Exploring Canada's Oil and Gas Industry*, 6th ed. (Petroleum Communication Foundation, Calgary, Canada).

Enger, P. S. (**1981**). "Frequency discrimination in teleosts—central or peripheral?," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 243–255.

Hastings, M. C., and Popper, A. N. (**2005**). "Effects of sound on fish," Report to California Department of Transportation Contract No. 43A0139, Task Order 1 http://www.dot.ca.gov/hq/env/bio/files/Effects_of_Sound_on_Fish23Aug05.pdf (last viewed Sept. 5, 2007).

Hastings, M. C., Popper, A. N., Finneran, J. J., and Lanford, P. J. (**1996**). "Effect of low frequency underwater sound on hair cells of the inner ear and lateral line of the teleost fish *Astronotus ocellatus*," J. Acoust. Soc. Am. **99**, 1759–1766.

Ladich, F., and Popper, A. N. (**2004**) "Parallel evolution in fish hearing organs," in *Evolution of the Vertebrate Auditory System*, edited by G. A. Manley, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 95–127.

Mann, D. A., Cott, P. A., Hanna, B. W., and Popper, A. N. (**2007**) "Hearing in eight species of northern Canadian freshwater fishes: implications for seismic surveys," J. Fish Biol. **70**, 109–120.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). "High intensity anthropogenic sound damages fish ears," J. Acoust. Soc. Am. **113**, 638–642.

Myrberg, A. A., Jr. (**1980**). "Ocean noise and the behavior of marine animals," in *Advanced Concepts in Ocean Measurements for Marine Biology*, edited by F. P. Diemer, F. J. Vernberg, and D. V. Mirkes (University of South Carolina Press, Columbia, SC), pp. 461–491.

Popper, A. N. (**1976**). "Ultrastructure of the auditory regions in the inner ear of the lake whitefish," Science **192**, 1020–1023.

Popper, A. N. (**1977**). "A scanning electron microscopic study of the sacculus and lagena in the ears of fifteen species of teleost fishes," J. Morphol. **153**, 397–418.

Popper, A. N. (**2003**). "Effects of anthropogenic sound on fishes," Fisheries **28**, 24–31.

Popper, A. N., and Platt, C. (**1983**). "Sensory surface of the saccule and lagena in the ears of ostariophysan fishes," J. Morphol. **176**, 121–129.

Popper, A. N., Fay, R. R., Platt, C., and Sand, O. (**2003**). "Sound detection mechanisms and capabilities of teleost fishes," in *Sensory Processing in Aquatic Environments*, edited by S. P. Collin and N. J. Marshall (Springer-Verlag, New York), pp. 3–38.

Popper, A. N., Fewtrell, J., Smith, M. E., and McCauley, R. D. (**2004**). "Anthropogenic sound: effects on the behavior and physiology of fishes," Mar. Technol. Soc. J. **37**(4), 35–40.

Popper, A. N., Smith, M. E., Cott, P. A., Hanna, B. W., MacGillivray, A. O., Austin, M. E., and Mann, D. A. (**2005**). "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**, 3958–3971.

Song, J., Mathieu, A., Soper, R. F., and Popper, A. N. (**2006**). "Structure of the inner ear of bluefin tuna (*Thunnus thynnus*)," J. Fish Biol. **68**, 1767–1781.

Ultrasound attenuation estimation using backscattered echoes from multiple sources

Timothy A. Bigelow[a)]
Department of Electrical Engineering, University of North Dakota, P.O. Box 7165, Grand Forks, ND 58202

(Received 21 February 2008; revised 21 May 2008; accepted 29 May 2008)

The objective of this study was to devise an algorithm that can accurately estimate the attenuation along the propagation path (i.e., the total attenuation) from backscattered echoes. It was shown that the downshift in the center frequency of the backscattered ultrasound echoes compared to echoes obtained in a water bath was calculated to have the form $\Delta f = m f_o + b$ after normalizing with respect to the source bandwidth where m depends on the correlation length, b depends on the total attenuation, and f_o is the center frequency of the source as measured from a reference echo. Therefore, the total attenuation can be determined independent of the scatterer correlation length by measuring the downshift in center frequency from multiple sources (i.e., different f_o) and fitting a line to the measured shifts versus f_o. The intercept of the line gives the total attenuation along the propagation path. The calculations were verified using computer simulations of five spherically focused sources with 50% bandwidths and center frequencies of 6, 8, 10, 12, and 14 MHz. The simulated tissue had Gaussian scattering structures with effective radii of 25 μm placed at a density of 250/mm^3. The attenuation of the tissue was varied from 0.1 to 0.9 dB/cm-MHz. The error in the attenuation along the propagation path ranged from $-3.5 \pm 14.7\%$ for a tissue attenuation of 0.1 dB/cm-MHz to $-7.0 \pm 3.1\%$ for a tissue attenuation of 0.9 dB/cm-MHz demonstrating that the attenuation along the propagation path could be accurately determined using backscattered echoes from multiple sources using the derived algorithm.
 [DOI: 10.1121/1.2949519]

PACS number(s): 43.80.Vj, 43.80.Ev, 43.80.Qf [CCC] Pages: 1367–1373

I. INTRODUCTION

In recent years, there has been an explosion of new applications for medical ultrasound. These applications range from ultrasound therapy to tissue characterization. Ultrasound therapy often involves thermal ablation where tissue necrosis is induced by elevated temperatures during ultrasound exposure (Foley *et al.*, 2004; Hynynen, 1997; Lizzi *et al.*, 1992; Otsuka *et al.*, 2005; Souchon *et al.*, 2003; Takegami *et al.*, 2005; Wu *et al.*, 2004). However, ultrasound has also been used for removing tissue in a controlled manner (Parsons *et al.*, 2006; Xu *et al.*, 2005; Xu *et al.*, 2004), gene transfection (Christiansen *et al.*, 2003; Duvshani-Eshet and Machluf, 2005; Feril *et al.*, 2005; Frenkel *et al.*, 2002; Liang *et al.*, 2004; Manome *et al.*, 2005; Ogawa *et al.*, 2002; Wei *et al.*, 2004; Zarnitsyn and Prausnitz, 2004), localized drug delivery (Shortencarier *et al.*, 2004; van Wamel *et al.*, 2004), and the acceleration of thrombolysis (Everbach and Francis, 2000). In tissue characterization, backscattered ultrasound echoes have been used to estimate the characteristic scatterer correlation length for tissue microstructure (Insana and Hall, 1990; Insana *et al.*, 1990; Lizzi *et al.*, 1997; Lizzi *et al.*, 1983; Oelze *et al.*, 2004; Oelze and Zachary, 2006) as well as to estimate tissue displacement/stiffness after using acoustic radiation force (ARFI) to push the tissue (Fahey *et al.*, 2005; Fahey *et al.*, 2004; Hsu *et al.*, 2007; Nightingale *et al.*, 2002).

In all of these applications, an accurate estimate of attenuation along the propagation path could improve the clinical utility of the technique. For the therapy applications, knowing the attenuation would improve therapy planning by allowing for an estimate of ultrasound fields at the sight of the therapy prior to initializing the therapy. When performing ARFI, an estimate of attenuation along the propagation path would allow for an estimate of the radiation force pushing the tissue to be determined, allowing for the stiffness of the tissue to be quantified. Also, when estimating the scatterer correlation length, an accurate estimate of the frequency dependence of total attenuation along the propagation path is critical in order to compensate for the spectral changes due to attenuation prior to estimating the scatterer correlation length.

Despite these benefits, an accurate estimation of total attenuation along the propagation path from backscattered ultrasound echoes has remained an elusive goal for many years. The challenge results from both attenuation and scatterer correlation length modifying the backscattered power spectrum (Bigelow *et al.*, 2005; Insana *et al.*, 1990). Some previous approaches have included quantifying the attenuation based on changes in the backscattered intensity with depth (He and Greenleaf, 1986; Tu *et al.*, 2006), estimating the local attenuation of the tissues along the propagation path and then summing these attenuations to obtain estimates of total attenuation (Lizzi *et al.*, 1992; Sidney, 1997), estimating

[a)]Author to whom correspondence should be addressed. Tel.: (701) 777-3368. FAX: (701) 777-5253. Electronic mail: timothybigelow@mail.und.nodak.edu.

the attenuation by comparing the echoes at two different frequencies assuming a coarse model for scattering (Lu *et al.*, 1995), and simultaneously estimating the scatterer correlation length and attenuation along the propagation path from backscattered ultrasound echoes (Bigelow and O'Brien, 2005a; Bigelow and O'Brien, 2005b; Bigelow *et al.*, 2005). However, these approaches have many deficiencies. The estimates based on changes in backscatter intensity assume that the tissue along the propagation path is homogeneous. Summing multiple estimates of local attenuation to obtain an estimate of total attenuation is highly computationally intensive and prone to errors as the errors can accumulate with increasing tissue depth. Lastly, simultaneously estimating both scatterer size and total attenuation results in poor precision and is highly dependent on an accurate model for the tissue scattering structure prior to obtaining the estimate. Therefore, a new algorithm is needed if attenuation along the propagation path is to be accurately estimated.

In this paper, a new algorithm for estimating the attenuation along the propagation path is derived. It combines backscattered echoes from multiple sources. Unlike the previous approach based on changes in backscattered intensity, the proposed algorithm does not assume that the tissue is homogeneous along the propagation path. Also, unlike the previous approach that sums estimates of local attenuation resulting in cumulative errors and high computational demands, the new algorithm estimates the attenuation based on the echoes received from a single region of the tissue. Last, unlike the previous algorithms that attempt to find the total attenuation and correlation length simultaneously, the proposed algorithm has only a weak dependence on the scattering structure of the tissue. After deriving the new algorithm, the performance of the algorithm is validated using computer simulations where the attenuation of the tissue is varied from 0.1 to 0.9 dB/cm–MHz.

II. DERIVATION OF ALGORITHM

The backscattered power spectrum received by an ultrasound source from a random distribution of weak scatterers at the focus that satisfy the Born approximation is given by (Bigelow and O'Brien, 2004b)

$$E[|V_{\text{scat}}(f)|^2] \propto \frac{k^4|V_{\text{plane}}(f)|^2}{A_{\text{comp}}(f)} F_\gamma(f, a_{\text{eff}}), \tag{1}$$

where k is the wave number. Also, $V_{\text{plane}}(f)$ is the voltage spectrum that would be returned from a rigid plane placed at the focal plane in a water bath and is obtained independently to calibrate the echoes from the tissue. $F_\gamma(f, a_{\text{eff}})$ is the form factor describing the correlation function for the tissue and depends on both frequency and the effective scatterer size, a_{eff}, of the tissue. $A_{\text{comp}}(f)$ is a generalized attenuation-compensation function that accounts for focusing, local attenuation, and total attenuation and is given by (Bigelow and O'Brien, 2004b)

$$A_{\text{comp}} = \frac{e^{4\alpha_{\text{tot}} z_T}}{\int_{-L/2}^{L/2} ds_z (g_{\text{win}}(s_z) e^{-4s_z^2/w_z^2} e^{4\alpha_{\text{loc}} s_z})}, \tag{2}$$

where α_{tot} is the total attenuation along the propagation path from the source to the focus and is the parameter being estimated by the algorithm developed in this paper. For inhomogeneous propagation paths, α_{tot} is the effective attenuation weighted by the attenuation of each tissue along the path and the path length in each tissue. Therefore, Eqs. (1) and (2) are equally valid for inhomogeneous tissues. Also, z_T is the distance from the aperture plane to the focal plane for the transducer, L is the length of the windowing function, g_{win}, in millimeters that is used to gate the backscattered waveform, and α_{loc} is the local attenuation in the focal region. w_z is the effective Gaussian depth of focus that results from approximating the field in the focal region with a Gaussian function and can be used to quantify the ultrasound field for both spherically focused and array sources. w_z depends linearly on wavelength and is given by $w_z = 6.01\lambda(f\#)^2$ for an ideal spherically focused source where $f\#$ is the f-number of the source (Bigelow and O'Brien, 2004b).

The value for w_z for a focused ultrasound source can be determined experimentally prior to imaging the tissue by obtaining echoes from a rigid plane scanned through the focal region in a water bath (Bigelow and O'Brien, 2004a). Furthermore, an earlier paper demonstrated that

$$\alpha_{\text{loc}} \equiv \sqrt{(\alpha_{\text{high}}^2 + \alpha_{\text{low}}^2)/2} \tag{3}$$

is a good choice for α_{loc} when there is uncertainty in the value of local attenuation where α_{high} and α_{low} are the largest and smallest attenuation values expected for the tissue (Bigelow and O'Brien, 2006). Therefore, the impact of local attenuation and focusing can be removed by dividing $E[|V_{\text{scat}}(f)|^2]$ by $\int_{-L/2}^{L/2} ds_z (g_{\text{win}}(s_z) e^{4s_z^2/w_z^2} e^{4\alpha_{\text{loc}} s_z})$ after obtaining echoes from the region of interest, yielding

$$E[|V_{\text{scat}}(f)|^2]_{\text{compensated}} \propto k^4 |V_{\text{plane}}(f)|^2 e^{-4\alpha_{\text{tot}} z_T} F_\gamma(f, \alpha_{\text{eff}}). \tag{4}$$

A similar expression to Eq. (4) can also be found after removing the focusing/diffraction effects for array sources. If we then split the attenuation term into a forward propagating contribution (i.e., prior to scattering), z_{T+}, and a backward propagating contribution (i.e., after scattering), z_{T-}, Eq. (4) becomes

$$E[|V_{\text{scat}}(f)|^2]_{\text{compensated}} \propto k^4 |V_{\text{plane}}(f)|^2 e^{-2\alpha_{\text{tot}} z_{T+}} F_\gamma(f, a_{\text{eff}}) e^{-2\alpha_{\text{tot}} z_{T-}}. \tag{5}$$

This split is necessary because some modifications of the spectrum due to attenuation occur prior to scattering and some occur after.

Equation (5) can be simplified further if we assume that the backscattered spectrum is approximately Gaussian such that

$$k^4|V_{\text{plane}}(f)|^2 \propto \exp\left(-\frac{(f-f_o)^2}{2\sigma_\omega^2}\right), \tag{6}$$

and we assume that the correlation function for the tissue has the form

$$F_\gamma(f,a_{\text{eff}}) \propto e^{-Af^n} \tag{7}$$

at least over a frequency range of interest. Equation (7) has been used as a general form for several form factors, including Gaussian and spherical shell, in the past (Bigelow *et al.*, 2005). In this equation, A is a constant that depends on a_{eff} as well as the correlation function while n gives the frequency dependence of the correlation function. $n\sim 2$ for most form factors. Based on these approximations and assuming that $\alpha_{\text{tot}}=\alpha_o f$ (i.e., strictly linear dependence on frequency), Eq. (5) can be written as

$$E[|V_{\text{scat}}(f)|^2]_{\text{compensated}} \propto \exp\left(-\frac{(f-f_o)^2}{2\sigma_\omega^2} - 2\alpha_o f z_{T+} - Af^n - 2\alpha_o f z_{T-}\right). \tag{8}$$

Therefore, the impact of total attenuation and scattering can be considered as a series of Gaussian transformations on a Gaussian function. A linear dependence on frequency may limit the usefulness of the derived algorithm in some applications. However, over a sufficiently small bandwidth, the power-law frequency dependence of some tissues can be approximated as $\alpha_o f+\alpha_b$ (Jongen *et al.*, 1986). The complete impact of this approximation when the tissue has a strong power-law dependence will be investigated in detail in the future.

The first Gaussian transformation corresponding to the propagation prior to scattering can be simplified to yield

$$\exp\left(-\frac{(f-f_o)^2}{2\sigma_\omega^2} - 2z_{T+}\alpha_o f\right) = \exp\left[-\left(\frac{[f-(f_o-2z_T\alpha_o\sigma_\omega^2)]^2 + 4z_T\alpha_o\sigma_\omega^2 f_o - (2z_T\alpha_o\sigma_\omega^2)^2}{2\sigma_\omega^2}\right)\right]. \tag{9}$$

Therefore, the expected backscattered power spectrum can be written as

$$E[|V_{\text{scat}}(f)|^2]_{\text{compensated}} \propto \exp\left(-\frac{(f-\tilde{f}_o)^2}{2\sigma_\omega^2} - Af^n - 2\alpha_o f z_{T-}\right), \tag{10}$$

where the terms independent of frequency are absorbed by the proportionality and $\tilde{f}_o=f_o-2z_T\alpha_o\sigma_\omega^2$. Hence, the attenuation due to forward propagation only modifies the center frequency of the power spectrum. The second transformation corresponding to scattering can also be simplified while neglecting higher order terms and assuming $n<3$, as was done in an earlier publication, yielding (Bigelow *et al.*, 2005)

$$E[|V_{\text{scat}}(f)|^2]_{\text{compensated}} \propto \exp\left(-\frac{(f-\tilde{f}_o')^2}{2\tilde{\sigma}_\omega^2} - 2\alpha_o f z_{T-}\right), \tag{11}$$

where $\tilde{f}_o'=\tilde{f}_o-\tilde{\sigma}_\omega^2 An\tilde{f}_o^{n-1}$ and $\tilde{\sigma}_\omega^2=[1/\sigma_\omega^2+An(n-1)\tilde{f}_o^{n-2}]^{-1}$. Therefore, scattering modifies both the center frequency and the bandwidth of the power spectrum. The last Gaussian transformation corresponding to back propagation can then be simplified similar to the first propagation corresponding to forward propagation to yield

$$E[|V_{\text{scat}}(f)|^2]_{\text{compensated}} \propto \exp\left(-\frac{[f-(\tilde{f}_o'-2z_T\alpha_o\tilde{\sigma}_\omega^3)]^2}{2\tilde{\sigma}_\omega^2}\right) = \exp\left(-\frac{(f-\tilde{f}_o'')^2}{2\tilde{\sigma}_\omega^2}\right). \tag{12}$$

Therefore, attenuation due to back propagation also only modifies the center frequency of the power spectrum. Only this time, it is the center frequency after scattering from the tissue.

Based on the Gaussian transformations corresponding to forward propagation, scattering, and back propagation, the change in center frequency of the spectrum scattered from the tissue as compared to the center frequency of the reference spectrum obtained from a rigid plane at the focus is given by

$$f_o-\tilde{f}_o'' = 4z_T\alpha_o\left(\frac{\sigma_\omega^2+\tilde{\sigma}_\omega^2}{2}\right) + \tilde{\sigma}_\omega^2 An(f_o-2z_T\alpha_o\sigma_\omega^2)^{n-1}. \tag{13}$$

Equation (13) can be further simplified by assuming that $n\sim 2$, yielding

$$f_o-\tilde{f}_o'' = 4z_T\alpha_o\left(\frac{\sigma_\omega^2+\tilde{\sigma}_\omega^2}{2} - \overbrace{\frac{\sigma_\omega^2\tilde{\sigma}_\omega^2 An}{2}}^{\text{small}}\right) + \tilde{\sigma}_\omega^2 An f_o. \tag{14}$$

However, $\sigma_\omega^2\tilde{\sigma}_\omega^2 An/2$ is typically much smaller than $(\sigma_\omega^2+\tilde{\sigma}_\omega^2)/2$ for diffuse scattering. Therefore, Eq. (14) is approximately given by

$$f_o-\tilde{f}_o'' \cong 4z_T\alpha_o\left(\frac{\sigma_\omega^2+\tilde{\sigma}_\omega^2}{2}\right) + \tilde{\sigma}_\omega^2 An f_o. \tag{15}$$

In Eq. (15), the downshift in center frequency is linearly dependent on the center frequency, f_o, of the reference spectrum $k^4|V_{\text{plane}}(f)|^2$. Also, the intercept of this line only depends on the attenuation along the propagation path, $z_T\alpha_o$,

the bandwidth of the backscattered spectrum from the tissue, $\tilde{\sigma}_\omega^2$, and the bandwidth of the reference spectrum σ_ω^2. It does not depend on the form or size of the scattering structures. Therefore, the attenuation along the propagation path can be determined regardless of the scattering structure in the region of interest by scanning the same tissue region by multiple sources with different center frequencies (i.e., f_o's) and measuring the downshift in center frequency normalized with respect to the average bandwidth, as given by

$$\Delta f_{\text{norm}} = \frac{f_o - \tilde{f}_o''}{\left(\frac{\sigma_\omega^2 + \tilde{\sigma}_\omega^2}{2}\right)} \cong 4 z_T \alpha_o + \frac{\tilde{\sigma}_\omega^2 An}{\left(\frac{\sigma_\omega^2 + \tilde{\sigma}_\omega^2}{2}\right)} f_o \tag{16}$$

for each source. Fitting a line to Δf_{norm} vs f_o and finding the intercept of that line yields the attenuation along the propagation path. Theoretically, Eq. (16) will yield the effective attenuation along the propagation path even for inhomogeneous tissues while only requiring echoes from a specific region of interest.

III. COMPUTER SIMULATIONS

A. Simulation parameters

The algorithm was verified by simulating the echo waveforms corresponding to five spherically focused ultrasound sources exposing exactly the same homogeneous attenuating half-spaces. The sources had focal lengths of 5 cm and f-numbers of 4. The center frequencies of the sources were 6, 8, 10, 12, and 14 MHz, and each source had a −3 dB bandwidth of 50%. The sound speed of the half-spaces was 1540 m/s, and the attenuation of the half-spaces was varied from 0.1 to 0.9 dB/cm-MHz in order to assess the dependence of the algorithm on attenuation. The attenuation dependence is critical when verifying the performance of the algorithm because an earlier algorithm also based on Gaussian approximations of the spectrum failed to give accurate attenuation estimates for attenuation values greater than 0.05 dB/cm MHz for tissue depths of ~5 cm (Bigelow *et al.*, 2005). The scattering structures within each half-space had Gaussian correlation functions (i.e., form factor of $F_\gamma(f, a_{\text{eff}}) = \exp[-0.827(ka_{\text{eff}})^2]$) with an a_{eff} of 25 μm and were positioned at a density of 250/mm^3 (approximately five scatterers per resolution cell for a 14 MHz transducer). Echoes were generated for 1000 random scatterer distributions for each value of half-space attenuation.

Estimates for $E[|V_{\text{scat}}(f)|^2]_{\text{compensated}}$ from each source were obtained by first windowing the simulated echoes in the time domain by a rectangular windowing function. A rectangular windowing function was selected because it yielded the best performance when compensating the backscattered power spectrum for windowing, focusing, and local attenuation (Bigelow and O'Brien, 2006). After windowing the echoes in the time domain, the power spectrum was found for each gated window, and a set of independent spectra (number in set varied from 5 to 100) was averaged together to yield ten estimates $E[|V_{\text{scat}}(f)|^2]_{\text{windowed}}$. Depending on the number of spectra, not all 1000 echoes were used in the estimate. The broadening of the spectra by windowing was then reduced by

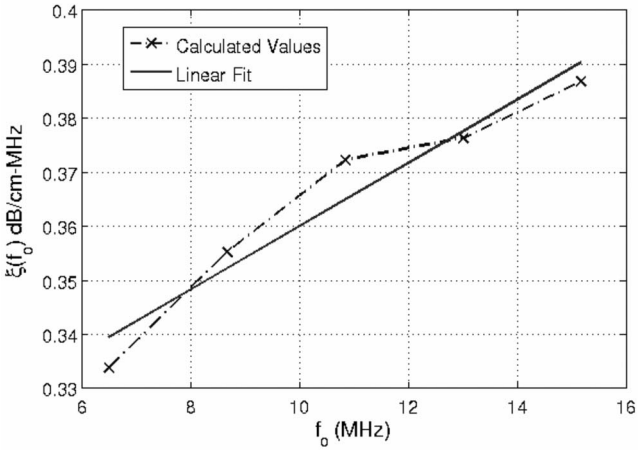


FIG. 1. A sample plot of $\xi(f_o)$ in dB/cm-MHz vs f_o for the five different sources used in the simulations with the corresponding linear fit. This example is for a half-space attenuation of 0.3 dB/cm-MHz and was obtained using 25 independent rf echoes and a rectangular gating window length of 4.36 mm (i.e., 17λ).

$$E[|V_{\text{scat}}(f)|^2] \cong E[|V_{\text{scat}}(f)|^2]_{\text{windowed}} \times \exp\left(-\frac{(f-\tilde{f}_a)^2 \sigma_g^2}{2(\tilde{\sigma}_a^2 + \sigma_g^2)\tilde{\sigma}_a^2}\right), \tag{17}$$

as has been described previously (Bigelow and O'Brien, 2005b). In Eq. (17), $\tilde{f}_a$ and $\tilde{\sigma}_a$ are the approximate values for the spectral peak frequency and Gaussian bandwidth of the backscattered power spectrum (i.e., $E[|V_{\text{scat}}(f)|^2]_{\text{windowed}} \propto \exp(-(f-\tilde{f}_a)^2/2\tilde{\sigma}_a^2)$) and σ_g^2 is the bandwidth found by fitting a Gaussian function to the magnitude squared of the Fourier transform of the rectangular windowing function. $E[|V_{\text{scat}}(f)|^2]$ was then divided by $\int_{-L/2}^{L/2} ds_z (g_{\text{win}}(s_z) e^{-4s_z^2/w_z^2} \times e^{4\alpha_{\text{loc}} s_z})$, where $\alpha_{\text{loc}} = 0.64$ dB/cm MHz, as given by Eq. (3) with α_{high} and α_{low} equal to 0.9 and 0.1 dB/cm MHz, respectively, to yield an estimate of $E[|V_{\text{scat}}(f)|^2]_{\text{compensated}}$.

Once $E[|V_{\text{scat}}(f)|^2]_{\text{compensated}}$ was determined, estimates for $\tilde{f}_o''$ and $\tilde{\sigma}_\omega^2$ were obtained for each source by fitting a Gaussian function to the power spectrum using frequencies in the −20 dB bandwidth of the backscattered power spectrum. While noise was not considered in this preliminary analysis, we anticipate that the main impact of noise will be to reduce the frequency range used to fit the Gaussian to the power spectra and hence should not significantly impact our results. Also, f_o and σ_ω^2 were found for each source by fitting a Gaussian function to the backscattered power spectrum of the echo from a rigid plane placed at the focal plane. From these quantities,

$$\xi(f_o) = \frac{2(f_o - \tilde{f}_o'')}{4 z_T (\sigma_\omega^2 + \tilde{\sigma}_\omega^2)} \cong \alpha_o + (\cdots) f_o \tag{18}$$

was calculated for each source. A linear fit to $\xi(f_o)$ vs f_o for the different sources yields an estimate for the slope of the attenuation along the propagation path. As an example, Fig. 1 shows a plot of $\xi(f_o)$ in dB/cm- MHz vs f_o for the five different sources used in the simulations with the corresponding linear fit. The attenuation for half-space was 0.3 dB/cm MHz, and this example was obtained using 25

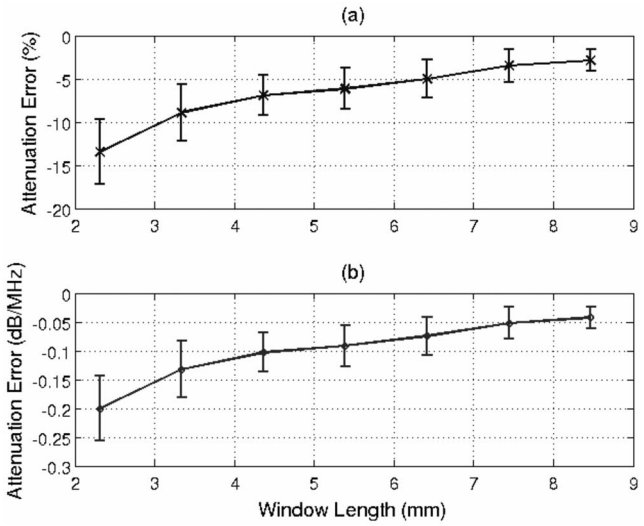


FIG. 2. Plot of mean and standard deviation of error in attenuation slope in (a) percent and in (b) dB/MHz vs the rectangular window length used to gate the time domain waveforms in millimeters. The attenuation of the half-space for this simulation was 0.3 dB/cm-MHz, and 100 independent echoes were used when estimating the power spectrum.

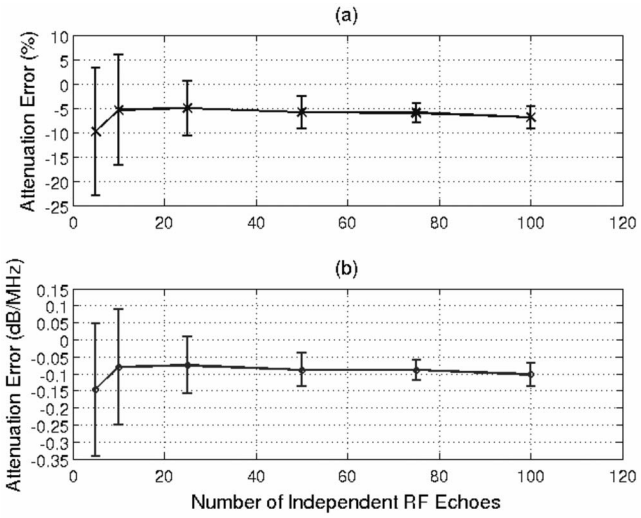


FIG. 3. Plot of mean and standard deviation of error in attenuation slope in (a) percent and in (b) dB/MHz vs the number of rf echoes used to obtain the estimates. The attenuation of the half-space for this simulation was 0.3 dB/cm-MHz, and a window length of 4.36 mm (i.e., 17λ) was used to gate the time domain waveforms.

independent rf echoes and a window length of 4.36 mm (i.e., 17λ).

B. Simulation results: Dependence on window length and number of rf echoes

Prior to assessing the algorithm dependence on attenuation, we needed to determine the optimal window length and the number of independent rf echoes needed to obtain reasonable estimates of attenuation. The algorithm assumes that the tissue is locally homogeneous in a small region about the focus. Therefore, the window length equates to the depth while the number of rf echoes along with the beam width at the focus relates to the width of the required homogeneous region. Ideally, this region should be as small as possible.

In order to assess the dependence on window length, the length of the rectangular window used to gate the time domain waveforms was varied from 9λ to 33λ (2.31–8.47 mm) in steps of 4λ, where λ corresponds to the wavelength in the tissue at 6 MHz, the smallest source frequency used in the simulations. The same rectangular window length (i.e., 2.31 mm for the smallest window) was used for all five sources even though the length corresponded to a larger number of wavelengths for the higher frequency sources. Therefore, the size of the homogeneous region along the beam axis was set by the wavelength of the lowest frequency source. In future studies, we anticipate this size to scale with frequency. The attenuation of the half-space for this simulation was 0.3 dB/cm-MHz, and 100 independent echoes were used when estimating the power spectrum.

Figure 2(a) shows the mean and standard deviation of the percentage error in the attenuation slope estimate, α_o, as a function of window length, while Fig. 2(b) shows the mean and standard deviation of the error in $\alpha_o z_T$ as a function of window length. The multiplication by z_T removes the influence of propagation depth when quantifying the error in the attenuation estimate. The error in the attenuation estimate is reduced as the window length is increased. However, the improvement is minimal for window lengths greater than 4.36 mm (i.e., 17λ). Therefore, a window length of 17λ was considered to be adequate and was used for the remainder of the simulations.

After determining an adequate window length, the next step was to find the number of independent rf echoes needed to obtain a reasonable estimate of attenuation. Therefore, the number of echoes used to obtain the estimates was set to 5, 10, 25, 50, 75, and 100. Once again, the attenuation of the half-space for this simulation was 0.3 dB/cm-MHz, and a window length of 4.36 mm was used to gate the time domain waveforms. The mean and standard deviation of the error in the attenuation estimate versus the number of echoes used to obtain the estimate is shown in Fig. 3.

Once again, the error is plotted as both a percentage error [Fig. 3(a)] as well as the error in $\alpha_o z_T$ [Fig. 3(b)]. Initially, as we go from five to ten echoes, both the mean and standard deviation of the error improve. For a larger number of echoes, the error in the mean remains relatively constant, but the precision of the estimates improves as the number of echoes increases. The improvement in precision is not significant when more than 25 echoes are used in the estimate. Therefore 25 echoes were considered adequate to obtain a reasonable estimate of total attenuation.

C. Simulation results: Dependence on half-space attenuation

After determining adequate values for window length and the number of rf echoes, the next step was to evaluate the dependence of the algorithm on half-space attenuation. Figure 4(a) shows the mean and standard deviation of the percentage error in the attenuation slope estimate, α_o, while Fig. 4(b) shows the mean and standard deviation of the error in $\alpha_o z_T$ for half-space attenuations of 0.1 to 0.9 dB/cm-MHz. The results were obtained using 25 independent rf echoes

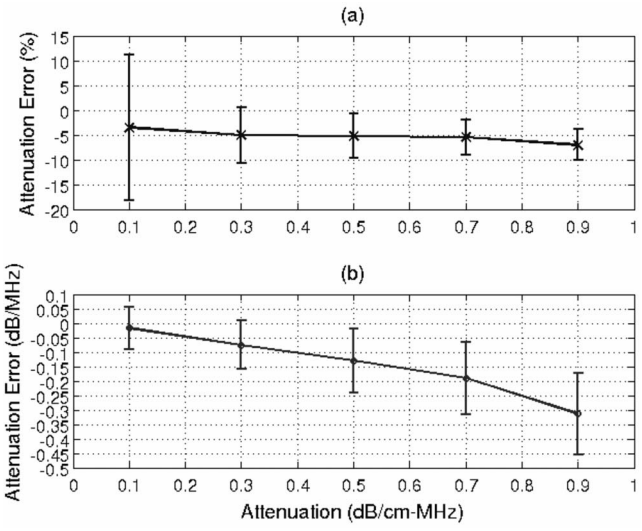


FIG. 4. Plot of mean and standard deviation of error in attenuation slope in (a) percent and in (b) dB/MHz vs the attenuation of the half-space. The window length used in this simulation was 4.36 mm to gate the time domain waveforms, and 25 independent echoes were used when estimating the power spectrum.

and a window length of 4.36 mm (i.e., 17λ). The percentage error [Fig. 4(a)] ranges from $-3.5\pm14.7\%$ for a tissue attenuation of 0.1 dB/cm-MHz to $-7.0\pm3.1\%$ for a tissue attenuation of 0.9 dB/cm-MHz, demonstrating that the attenuation along the propagation path could be accurately determined using the derived algorithm even for large attenuation values. The precision of the $\alpha_o z_T$ estimate is also very good, varying from 0.07 to 0.14 dB/MHz as the half-space attenuation increases [Fig. 4(b)]. The mean error of $\alpha_o z_T$ degrades as the attenuation of the half-space increases, but since the percentage error remains approximately −5%, the degradation is due to the ~−5% bias contributing more significantly at the higher attenuation values.

IV. DISCUSSION

In this paper, we demonstrated that the slope of the attenuation along the propagation path, α_o, can be determined by scanning the region of interest using multiple transducers. The algorithm only assumes that the attenuation along the propagation path has a linear dependence on frequency and that the spectrum transmitted by the ultrasound source, either array or spherically focused, is approximately Gaussian. The downshift in center frequency when comparing the backscattered spectrum to the echo from a reference spectrum was shown to have the form $\Delta f = m f_o + b$ after normalizing with respect the the source bandwidth where m depends on correlation length and b depends on α_o. Hence, the intercept of this line can be used to find α_o independent of the frequency dependence of the scattering. This algorithm could lead to significant advances in ultrasound tissue characterization and ultrasound therapy planning since attenuation along the propagation path is a critical parameter in both applications. This algorithm is different from previously proposed algorithms in that it can find the attenuation along the propagation path from echoes from a single region of interest rather than require attenuation estimates along the entire propagation path. The algorithm is also not strongly dependent on the scattering properties in the region of interest. Therefore, errors in the modeling of the sources of ultrasound scattering should not impact the attenuation estimates. The algorithm also does not assume that the tissue along the propagation path is homogeneous, allowing for clinical implementation using real tissues.

While the performance of the developed algorithm looks very promising, the performance could be significantly improved if the source of the ~−5% bias could be determined. One possibility that was considered is that the bias results from the assumption that $\sigma_\omega^2 \tilde{\sigma}_\omega^2 An/2$ is typically much smaller than $(\sigma_\omega^2+\tilde{\sigma}_\omega^2)/2$ from the derivation of Eq. (15). However, a more rigorous mathematical analysis revealed that a violation of this assumption would result in a slight overestimation of the attenuation, whereas the simulations demonstrate a slight underestimation. Therefore, the source of the underestimation is not clear at this time. Additional computer simulations and phantom experiments will provide more insights as the investigation is continued in the future.

While the ~−5% bias should be investigated further and possibly removed, the algorithm still has significant clinical utility even with this bias. Therefore, the algorithm is almost ready for clinical implementation. The main hurdle that remains is that the accuracy and precision of the algorithm should be verified experimentally. The experiments should include both phantom experiments as well as *ex vivo* tissue experiments. Also, when implementing the algorithm using array sources, calibrating the sources to determine the values of f_o using a planar boundary might be challenging and, other calibration techniques might need to be explored. Overall, the algorithm has great potential clinically and could greatly improve many different emerging clinical ultrasound applications.

ACKNOWLEDGMENTS

This project was supported by Grant No. R01 CA111289 from the National Institutes of Health as well as the University of North Dakota School of Engineering and Mines. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2004a**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Calibration measurements and phantom experiments," J. Acoust. Soc. Am. **116**, 594–602.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2004b**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Theoretical approximations and simulation analysis," J. Acoust. Soc. Am. **116**, 578–593.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2005a**). "Evaluation of the spectral fit algorithm as functions of frequency range and Dka_{eff}," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 2003–2010.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2005b**). "Signal processing strategies that improve performance and understanding of the quantitative ultrasound SPECTRAL FIT algorithm," J. Acoust. Soc. Am. **118**, 1808–1819.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2006**). "Impact of local attenuation approximations when estimating correlation length from backscattered ultrasound echoes," J. Acoust. Soc. Am. **120**, 546–553.

Bigelow, T. A., Oelze, M. L., and O'Brien, W. D., Jr. (**2005**). "Estimation of total attenuation and scatterer size from backscattered ultrasound waveforms," J. Acoust. Soc. Am. **117**, 1431–1439.

Christiansen, J. P., French, B. A., Klibanov, A. L., Kaul, S., and Lindner, J.

R. (**2003**). "Targeted tissue transfection with ultrasound destruction of plasmid-bearing cationic microbubbles," Ultrasound Med. Biol. **29**, 1759-1767.
Duvshani-Eshet, M., and Machluf, M. (**2005**). "Therapeutic ultrasound optimization for gene deliver: A key factor achieving nuclear DNA localization," J. Controlled Release **108**, 513–528.
Everbach, E. C., and Francis, C. W. (**2000**). "Cavitational mechanisms in ultrasound-accelerated thrombolysis at 1 MHz," Ultrasound Med. Biol. **26**, 1153–1160.
Fahey, B. J., Nightingale, K. R., Nelson, R. C., Palmeri, M. L., and Trahey, G. E. (**2005**). "Acoustic radiation force impulse imaging of the abdomen: demonstration of feasibility and utility," Ultrasound Med. Biol. **31**, 1185–1198.
Fahey, B. J., Nightingale, K. R., Stutz, D. L., and Trahey, G. E. (**2004**). "Acoustic radiation force impulse imaging of thermally- and chemically-induced lesions in soft tissues: Preliminary ex vivo results," Ultrasound Med. Biol. **30**, 321–328.
Feril, J. L. B., Ogawa, R., Kobayashi, H., Kikuchi, H., and Kondo, T. (**2005**). "Ultrasound enhances liposome-mediated gene transfection," Ultrason. Sonochem. **12**, 489–493.
Foley, J. L., Little, J. W., Starr, Ill, F. L., Frantz, C., and Vaezy, S. (**2004**). "Image-guided HIFU neurolysis of peripheral nerves to treat spasticity and pain," Ultrasound Med. Biol. **30**, 1199–1207.
Frenkel, P. A., Chen, S., Thai, T., Shohet, R. V., and Grayburn, P. A. (**2002**). "DNA-loaded albumin microbubbles enhance ultrasound-mediated transfection in vitro," Ultrasound Med. Biol. **28**, 817–822.
He, P., and Greenleaf, J. F. (**1986**). "Application of stochastic analysis to ultrasonic echoes—Estimation of attenuation and tissue heterogeneity from peaks of echo envelope," J. Acoust. Soc. Am. **79**, 526–534.
Hsu, S. J., Fahey, B. J., Dumont, D. M., Wolf, P. D., and Trahey, G. E. (**2007**). "Challenges and implementation of radiation-force imaging with an intracardiac ultrasound transducer," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 996–1009.
Hynynen, K. (**1997**). "Review of ultrasound therapy," Proc.-IEEE Ultrason. Symp. **2**, 1305–1313.
Insana, M. F., and Hall, T. J. (**1990**). "Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation," Ultrason. Imaging **12**, 245–267.
Insana, M. F., Wagner, R. F., Brown, D. G., and Hall, T. J. (**1990**). "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192.
Jongen, H. A. H., Thijssen, J. M., Aarssen, M. v. d., and Verhoef, W. A. (**1986**). "A general model for the absorption of ultrasound by biological tissues and experimental verification," J. Acoust. Soc. Am. **79**, 535–540.
Liang, H.-D., Lu, Q. L., Xue, S.-A., Halliwell, M., Kodama, T., Cosgrove, D. O., Stauss, H. J., Partridge, T. A., and Blomley, M. J. K. (**2004**). "Optimisation of ultrasound-mediated gene transfer (sonoporation) in skeletal muscle cells," Ultrasound Med. Biol. **30**, 1523–1529.
Lizzi, F. L., Astor, M., Liu, T., Deng, C., Coleman, D. J., and Silverman, R. H. (**1997**). "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," Int. J. Imaging Syst. Technol. **8**, 3–10.
Lizzi, F. L., Driller, J., Lunzer, B., Kalisz, A., and Coleman, D. J. (**1992**). "Computer model of ultrasonic hyperthermia and ablation for ocular tumors using b-mode data," Ultrasound Med. Biol. **18**, 59–73.
Lizzi, F. L., Greenebaum, M., Feleppa, E. J., Elbaum, M., and Coleman, D. J. (**1983**). "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373.
Lu, Z. F., Zagzebski, J. A., Madsen, E. L., and Dong, F. (**1995**). "A method for estimating an overlying layer correction in quantitative ultrasound imaging," Ultrason. Imaging **17**, 269–290.
Manome, Y., Nakayama, N., Nakayama, K., and Furuhata, H. (**2005**). "Insonation facilitates plasmid DNA transfection into the central nervous system and microbubbles enhance the effect," Ultrasound Med. Biol. **31**, 693–702.
Nightingale, K., Soo, M. S., Nightingale, R., and Trahey, G. (**2002**). "Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility," Ultrasound Med. Biol. **28**, 227–235.
Oelze, M. L., O'Brien, W. D., Jr., Blue, J. P., and Zachary, J. F. (**2004**). "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771.
Oelze, M. L., and Zachary, J. F. (**2006**). "Examination of cancer in mouse models using high-frequency quantitative ultrasound," Ultrasound Med. Biol. **32**, 1639–1648.
Ogawa, R., Kondo, T., Honda, H., Zhao, Q. L., Fukuda, S., and Riesz, P. (**2002**). "Effects of dissolved gases and an echo contrast agent on ultrasound mediated in vitro gene transfection," Ultrason. Sonochem. **9**, 197–203.
Otsuka, R., Fujikura, K., Hirata, K., Pulerwitz, T., Oe, Y., Suzuki, T., Sciacca, R., Marboe, C., Wang, J., Burkhoff, D., Muratore, R., Lizzi, F. L., and Homma, S. (**2005**). "In vitro ablation of cardiac valves using high-intensity focused ultrasound," Ultrasound Med. Biol. **31**, 109–114.
Parsons, J. E., Cain, C. A., Abrams, G. D., and Fowlkes, J. B. (**2006**). "Pulsed cavitational ultrasound therapy for controlled tissue homogenization," Ultrasound Med. Biol. **32**, 115–129.
Shortencarier, M. J., Dayton, P. A., Bloch, S. H., Schumann, P. A., Matsunaga, T. O., and Ferrara, K. W. (**2004**). "A method for radiation-force localized drug delivery using gas-filled lipospheres," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 822–831.
Sidney, D. A. (**1997**). "Three-dimensional ultrasound power deposition modeling, thermal field visualization, and clinical integration of hyperthermia therapy," Ph.D. thesis, Massachussetts Institute of Technology at Boston, MA.
Souchon, R., Rouviere, O., Gelet, A., Detti, V., Srinivasan, S., Ophir, J., and Chapelon, J.-Y. (**2003**). "Visualisation of HIFU lesions using elastography of the human prostate in vivo: Preliminary results," Ultrasound Med. Biol. **29**, 1007–1015.
Takegami, K., Kaneko, Y., Watanabe, T., Maruyama, T., Matsumoto, Y., and Nagawa, H. (**2005**). "Erythocytes, as well as microbubble contrast agents, are important factors in improving thermal and therapeutic effects of high-intensity focused ultrasound," Ultrasound Med. Biol. **31**, 385–390.
Tu, H., Zagzebski, J., and Chen, Q. (**2006**). "Attenuation estimations using envelope echo data: Analysis and simulations," Ultrasound Med. Biol. **32**, 377–386.
van Wamel, A., Bouakaz, A., Bernard, B., ten Cate, F., and de Jong, N. (**2004**). "Radionuclide turmour therapy with ultrasound contrast microbubbles," Ultrasonics **42**, 903–906.
Wei, W., Zhengzhong, B., Yongjie, W., Lefeng, Y., and Yalin, M. (**2004**). "A novel approach to quantitative ultrasonic naked gene delivery and its non-invasive assessment," Ultrasonics **43**, 69–77.
Wu, F., Wang, Z.-B., Lu, P., Xu, Z.-L., Chen, W.-Z., Zhu, H., and Jin, C.-B. (**2004**). "Activated anti-tumor immunity in cancer patients after high intensity focused ultrasound ablation," Ultrasound Med. Biol. **30**, 1217–1222.
Xu, Z., Fowlkes, J. B., Ludomirsky, A., and Cain, C. A. (**2004**). "Investigation of intensity thresholds for ultrasound tissue erosion," Ultrasound Med. Biol. **31**, 1673–1682.
Xu, Z., Ludomirsky, A., Eun, L. Y., Hall, T. L., Tran, B. C., Fowlkes, J. B., and Cain, C. A. (**2005**). "Controlled ultrasound tissue erosion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 726–736.
Zarnitsyn, V. G., and Prausnitz, M. R. (**2004**). "Physical parameters influencing optimization of ultrasound-mediated DNA transfection," Ultrasound Med. Biol. **30**, 527–538.