CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 126, No. 1 July 2009

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

NONLINEAR ACOUSTICS [25]

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PERCEPTION [71]

MUSIC AND MUSICAL INSTRUMENTS [75]

(Continued)

CONTENTS—*Continued from preceding page*

Closure duration analysis of incomplete stop consonants due to stop-stop interaction

Prasanta Kumar Ghosh and Shrikanth S. Narayanan
Department of Electrical Engineering, Signal Analysis and Interpretation Laboratory, University of Southern California, Los Angeles, California 90089
prasantg@usc.edu; shri@sipi.usc.edu

Abstract: An incomplete stop consonant is characterized either by an indistinguishable closure or a missing burst. If an incomplete stop happens due to a stop following another stop [stop-stop interaction (SSI)], its acoustics typically resemble that of a complete stop—one closure followed by a single burst. As a consequence, stop detectors would fail to distinguish an SSI from a complete stop. Analysis of the TIMIT corpus shows 35.04% incomplete stops (14.97% SSI). It is shown that by using automatically estimated (and hand-labeled) closure duration, complete stops can be distinguished from incomplete stops due to SSI with 69.66% (79.14%) accuracy.

PACS numbers: 43.72.Ar [DOS]
Date Received: February 21, 2009 **Date Accepted:** April 21, 2009

1. Introduction

Stop consonants in English speech have been a topic of research over the past few decades, particularly to better understand and analyze the dynamic and highly speaker- and context-dependent nature of these sounds. A stop consonant (/b/, /d/, /g/ [voiced] and /p/, /t/, /k/ [unvoiced]) is produced when there is complete closure of the articulators, stopping the airflow in the vocal tract, followed by a release or burst of air.[1] However, in conversational or even read speech, this acoustic signature of a stop is not always apparent due to intergestural overlap,[2] which sometimes results in the absence of a clear stop release or burst.[3,4] These short-dynamic acoustic variations make the state-of-the-art hidden Markov model based automatic speech recognizer (ASR) incapable of performing accurate fine phoneme distinctions for this class of sounds.[5]

To address this problem of ASR, researchers have proposed several alternative features and models to detect stop consonants. For example, to detect stop consonants, spectral and temporal features,[5,6] the optimal filter approach,[7] and the wavelet transform approach[8] have been used to capture a period of extremely low energy (corresponding to the period of closure) followed by a sharp, broadband signal (corresponding to the release). These features are in turn being used within novel automatic speech recognition frameworks such as those based on landmark detectors.[9] However, all these approaches for stop detectors implicitly assume that a stop should be a complete stop,[3] which is defined as one that should include an identifiable closure portion followed by a burst release. But corpus studies (in English) have shown that acoustic implementation of complete stops is only a fraction of the possibilities. For example, in a comprehensive study by Crystal and House,[3] complete stops accounted for only about 45% of the identified stops.

In this work, we investigate the complete and incomplete stops in the TIMIT database[10] with a specific focus on the robustness of the stop detectors in the presence of incomplete stops. In particular, we provide the detailed analysis of incomplete stop consonants[4] in the presence of stop-stop interaction (SSI). Spectro-temporally, these sounds share similar patterns with complete stops[1] motivating us to analyze their acoustic properties further. In a framework like distinctive feature landmark detection for speech recognition,[9] such an analysis would provide more insights into detecting stops more accurately. It should be noted that the formant transition is often used as an acoustic cue in stop detection,[6] which is not affected, in general, by the SSI.

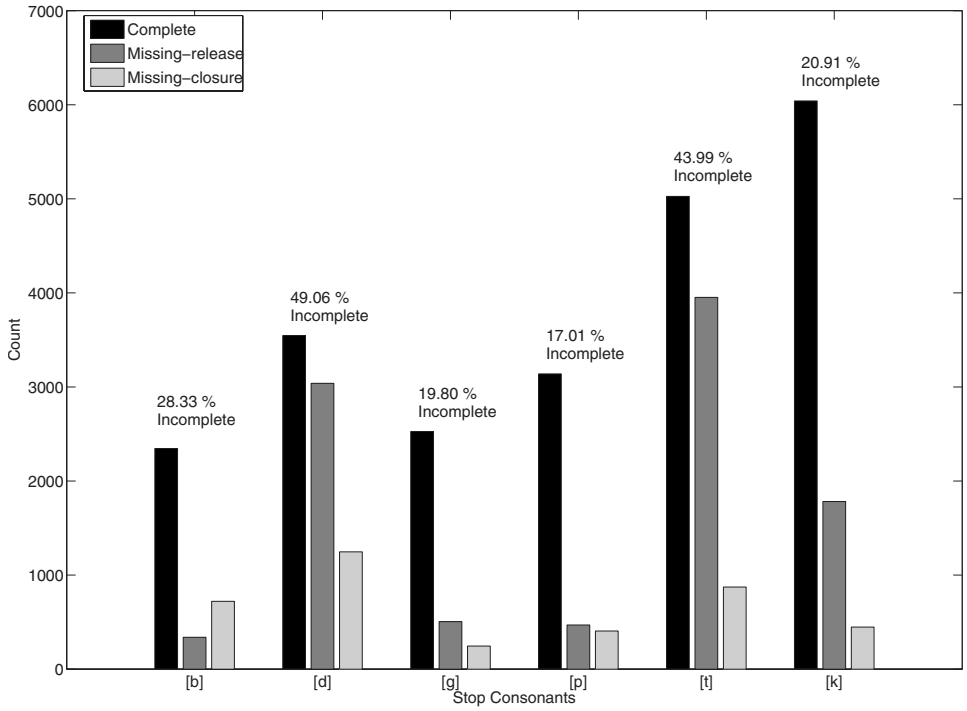


Fig. 1. Number of complete and different incomplete stops in TIMIT.

However, it is difficult to estimate formant transitions near a stop closure.[11] Hence both temporal and spectral properties of burst, closure duration, and voice onset time (VOT) are mainly used as acoustic features in stop detection.[5] But due to acoustic similarity between incomplete stops due to SSI and complete stop, stop detectors (which assume each stop is a complete stop) would miss one stop for every SSI. The analysis in this paper provides insight as to how the closure duration can be used to disambiguate SSIs from complete stops, leading to potential improvements to stop detectors. In our analysis, we found that with hand-labeled closure duration as a cue we can distinguish non-released stops due to SSI from complete stops with an accuracy of 79.14%. With automatically estimated closure durations, we can achieve a detection accuracy of 69.66%.

2. Complete and incomplete stops in TIMIT

We first study the relative frequencies of complete and incomplete stops of American English in TIMIT. The reason for choosing TIMIT is that it is a phonetically balanced database that has been well studied. In TIMIT, the phonetic transcriptions of the release of six stops are denoted by /b/, /d/, /g/, /p/, /t/, /k/ and their closures by /bcl/, /dcl/, /gcl/, /pcl/, /tcl/, /kcl/, respectively. A stop consonant [t] is counted as complete if /tcl/ is followed by /t/. On the other hand, an incomplete stop can be of two types—(1) if /tcl/ is followed by any phoneme other than /t/ (we call this a *missing-release* stop) and (2) if /t/ is preceded by any phoneme other than /tcl/ (we call this a *missing-closure* stop). This terminology applies to other stop consonants as well. *Carpet cleaners* (/kcl/k/p/ix/tcl/k/l/iy/n/..) is an example of a missing-release stop and *events* (/ix/v/eh/n/t/s/) is an example of a missing-closure stop (note the underlined portions). Figure 1 shows the number of complete, missing-release, and missing-closure stops for each of the six stop consonants in TIMIT. The percentages of the total incomplete (No. of missing-release + No. of missing-closure) stops are also shown in Fig. 1. Considering all the stop consonants together, there are 35.04% incomplete stops (missing-release, 28.96% and missing-closure, 11.31%) in the TIMIT database. It is clear from Fig. 1 that the percentage of incomplete [d] and [t] is high.

Table 1. Top 5 missing-release stops (for each stop closure); %SSI is the percentage of missing-release stops, due to SSI.

	Following phonemes					
Closures	First	Second	Third	Fourth	Fifth	%SSI
/bcl/	/d/	/t/	/jh/	/s/	/el/	28.02
/dcl/	/b/	/t/	/y/	/dh/	/z/	26.68
/gcl/	/l/	/z/	/n/	/ix/	/d/	20.19
/pcl/	/t/	/s/	/b/	/dh/	/m/	41.06
/tcl/	/s/	/dh/	/q/	/b/	/d/	16.41
/kcl/	/dh/	/m/	/t/	/s/	/d/	28.05

On average, for every two occurrences of [d] or [t], any acoustic feature based stop detector may fail to detect one of those occurrences.

Tables 1 and 2 show the top five phonemes that follow each of the stop closures and precede different stop releases in the case of missing-release and missing-closure stops, respectively. The underlined entries in both tables suggest that the release of a stop can follow a closure of another stop (we call this an SSI) and can contribute to an incomplete stop (23.54% for missing-release and 46.35% for missing-closure stops).

Browman and Goldstein[2] showed that gestural overlap can cause such deletion or assimilation leaving no acoustic evidence of the consonant burst, resulting in incomplete stops.

2.1 Incomplete stops due to SSI

From Tables 1 and 2, we see that when a stop consonant follows (interacts with) another stop consonant, it can lose its individual acoustic signature and manifest itself as an incomplete stop; we refer to them as incomplete stops due to SSI. A stop consonant interacts with another stop consonant either within a word (e.g., *subject* and *jumped*) or across words (e.g., *rapid car* and *sharp dresser*).[1] Incomplete stop due to SSI has one closure followed by a single burst and thus appears acoustically indistinguishable from complete stop.

%SSIs in Tables 1 and 2 refer to the percentages of different missing-release and missing-closure stops, which are due to SSI. We can see that the SSIs (particularly for missing-closure stops) cover a significant portion of the incomplete stops. This motivates us to investigate how an incomplete stop due to SSI can be distinguished from a complete stop using acoustic cues.

3. Closure duration of incomplete stop due to SSI vs complete stop

Although the acoustic pattern of an incomplete stop due to SSI appears similar to that of a complete stop, the mean duration of closure (as specified in the TIMIT transcription) for incom-

Table 2. Top 5 missing-closure stops (for each stop release); %SSI is the percentage of missing-closure stops, due to SSI.

	Preceding phonemes					
Release	First	Second	Third	Fourth	Fifth	%SSI
/b/	/dcl/	/tcl/	/pau/	/pcl/	/gcl/	56.31
/d/	/tcl/	/n/	/kcl/	/pau/	/bcl/	21.57
/g/	/tcl/	/ng/	/kcl/	/dcl/	/pau/	43.08
/p/	/tcl/	/dcl/	/kcl/	/gcl/	/bcl/	58.12
/t/	/kcl/	/dcl/	/pcl/	/pau/	/n/	71.93
/k/	/tcl/	/pau/	/dcl/	/pcl/	/bcl/	40.09

Table 3. Mean (SD) closure durations (in second) for different incomplete stops due to SSI, complete stops (bold entries correspond to complete stops) and stop-fricative, stop-nasal, stop-glides, stop-vowel interactions.

	Following phoneme categories									
	Stop release (SSI and complete stop)									
Closure	/b/	/d/	/g/	/p/	/t/	/k/	Fricative	Nasal	Glides	Vowel
/bcl/	**0.063**	0.099	0.133	0.111	0.098	0.094	0.063	0.056	0.049	0.059
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
/dcl/	0.086	**0.049**	0.088	0.096	0.072	0.093	0.043	0.05	0.056	0.057
	(0.02)	**(0.02)**	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
/gcl/	0.122	0.096	**0.047**	0.101	0.118	0.089	0.059	0.056	0.052	0.048
	(0.02)	(0.02)	**(0.02)**	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
/pcl/	0.109	0.122	0.125	**0.067**	0.091	0.1129	0.071	0.058	0.070	0.086
	(0.02)	(0.02)	(0.02)	**(0.02)**	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)
/tcl/	0.097	0.086	0.098	0.093	**0.048**	0.085	0.045	0.054	0.063	0.063
	(0.03)	(0.03)	(0.03)	(0.04)	**(0.02)**	(0.03)	(0.03)	(0.02)	(0.04)	(0.04)
/kcl/	0.103	0.104	0.112	0.120	0.094	**0.054**	0.050	0.057	0.073	0.089
	(0.03)	(0.02)	(0.05)	(0.02)	(0.03)	**(0.02)**	(0.02)	(0.02)	(0.04)	(0.03)

plete stops due to SSI is consistently higher than that of complete stop consonants. The mean closure durations [with standard deviation (SD)] for all incomplete stops due to SSI and complete stops are shown in Table 3. The bold entries in this table indicate the minimum mean closure duration among all SSIs in a row. It is clear that the minimum durations also correspond to the complete stops (bold entries). This observation supports many previous studies in the literature: Olive *et al.*[1] reported smaller closure duration for single [t] than that of a geminate; Homma[12] provided a similar observation from an experiment on a set of 24 words spoken by four speakers; Manuel *et al.*[13] also observed similar differences in nasal consonant durations in "in a" and "in the." However, to the best of our knowledge, there has not been a comprehensive analysis of stop closure durations of different SSIs on a large multitalker dataset.

The total number of complete stops in TIMIT is 22624 and that of incomplete stops due to SSI is 1826 (8.07%). The normalized histograms of the closure duration of these two classes are shown in Fig. 2(a). This figure clearly shows that incomplete stops due to SSI and complete stops can be distinguished to some extent based on their closure duration.

We also found that when there is an interaction (for *missing-release*) where any phoneme other than a stop follows a stop consonant, the closure duration of this stop is not necessarily higher than that of the complete stop. Table 3 supports this observation. To compute the mean closure duration in such incomplete stops due to non-SSI, we first categorize the other interacting phonemes into fricatives, nasals, glides (and liquids), and vowels. From Table 3, it is seen that the mean closure durations in these cases are similar to those of the complete stops (as seen in bold entries of Table 3).

Also for missing-closure stops, there are cases where a stop is preceded by a phoneme other than a stop consonant. In these cases the closure duration does not exist for the stop and these are, in general, difficult to detect by acoustic cues (they are 17.33% of all incomplete stops).

4. Automatic classification of complete stop consonants and incomplete stops due to SSI

We have already seen that the mean closure durations (as mentioned in the TIMIT transcription) of a complete stop and an incomplete stop due to SSI are different even though both exhibit a similar acoustic pattern of a single closure followed by a single air burst. We perform two classification experiments to investigate the discrimination power of using closure duration as a feature. First, we use the actual closure durations transcribed in the TIMIT for an "oracle test,"

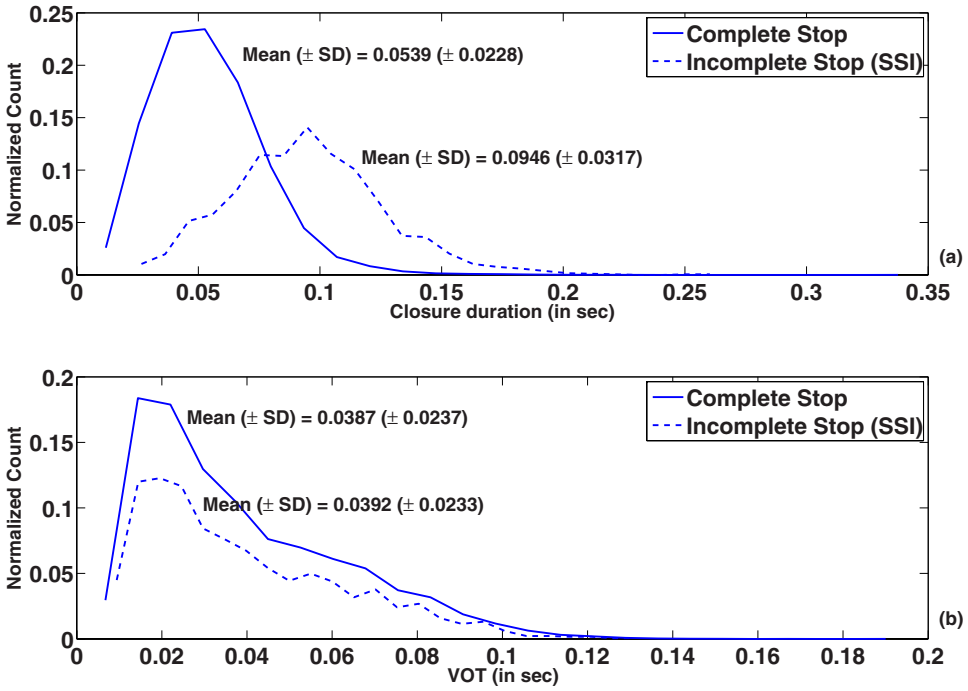


Fig. 2. (Color online) Normalized histogram (normalization is done so that the fractional counts in the histogram add up to 1) of (a) closure duration and (b) VOT.

that is, a classification experiment in which manually transcribed closure duration is used to discriminate between complete and incomplete stops due to SSI. Second, we perform the same detection experiments with automatically measured stop closure durations.

4.1 Oracle detection experiment

We randomly selected 500 complete stops and 500 incomplete stops due to SSI from TIMIT for testing and used the remaining segments for training. We trained two-component Gaussian mixture models (GMMs) for both classes using the EM algorithm (we tried other numbers of mixtures, but the best accuracy was obtained with two components). GMMs observed the closure durations that were provided in the TIMIT database. Since our test set was balanced and did not reflect the bias in the training set, we implemented maximum likelihood as opposed to maximum-a-posteriori classification scheme.

Classification accuracy was computed using 50-fold cross-validation, resulting in an average accuracy of 79.14% (SD=1.23%). Complete stops were classified with a mean accuracy of 80.08% (SD=1.92%), whereas SSIs with 78.2% (SD=1.73%). Thus class specific accuracies are not very different. We also tested the use of VOT as an additional feature, but it did not improve the accuracy significantly. Thus it can be concluded that VOT is not a useful cue for distinguishing SSI from complete stops. The histogram of VOT for these two classes supports this fact [see Fig. 2(b)].

4.2 Detection using automatically estimated closure duration

In this experiment, we designed a simple stop detector using the energy of the closure duration as a feature and consequently estimated the closure duration for a detected stop. Since stop releases are transient events, we used an analysis window of 1 ms length with no overlap. We computed the energy of the speech signal (normalized to ±1) in each analysis window. We used the TIMIT training dataset to learn the distributions of energies of the signal in the analysis

window for stop closures and for events other than stop closures. Stop closures were detected by thresholding the energy, using the equal error rate threshold (the threshold at which recall and precision rates are equal), which turned out to be 0.6974 for these data. If the energy of the signal in an analysis window was less than 0.6974, we declared that the frame belongs to a stop closure. However, in this approach, many spurious frames were detected as frames belonging to stop closure. To prevent this, we imposed another constraint. From the histogram of closure durations [Fig. 2(a)], we observe that the minimum closure duration is ~15 ms. Thus if at least 15 consecutive frame energies are below 0.6974, we declared that the respective sequence of frames corresponds to a stop closure. If N ($N \geq 15$) consecutive frame energies were consistently less than the threshold, we considered the estimated duration of the respective stop closure is N ms.

We used the TIMIT test dataset for evaluation. Using the above-mentioned simple stop closure detection algorithm, we could detect 92.39% of all stops (including both complete and SSI) in the test dataset. The estimated stop closure durations were then used to classify the stops into either complete stops or SSIs, using two-component GMMs trained on the transcribed closure durations of the training dataset. The mean classification accuracy was 69.66%. Complete stops were classified with a mean accuracy of 63.70%, whereas SSIs with 75.62%. The reduction in accuracy compared to that of the oracle test is due to the fact that the closure durations are estimated and not the actual ones as transcribed in the TIMIT database.

5. Conclusion

We showed that closure duration can be used as a feature to classify the incomplete stops due to SSI and the complete stops in read speech of the TIMIT with 69.66% (79.14%) accuracy for automatically estimated (reference) durations. We also found that the closure durations of the incomplete stops, when not due to SSI, are similar to those of complete stops.

Our analysis provides opportunities for improvement of stop consonant detectors, particularly for applications such as distinctive feature landmark detectors for speech recognition.[9] It is important to note that the problem of nonreleased stops is not speaker dependent as the data analyzed come from multiple talkers. However, a similar study on conversational speech remains to be done to fully understand the effect of variability in stop production on the performance of stop detectors.

Acknowledgments

Work supported in part by NIH and ONR-MURI.

References and links

[1]J. P. Olive, A. Greenwood, and J. Coleman, *Acoustics of American English Speech: A Dynamic Approach* (Springer-Verlag, Berlin, 1993), pp. 227–312.
[2]C. Browman and L. Goldstein, *Tiers in Articulatory Phonology, With Some Implications for Casual Speech*, Papers in Laboratory Phonology, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge, 1990), pp. 341–386.
[3]T. H. Crystal and A. S. House, "The duration of American-English stop consonants: An overview," J. Phonetics **16**, 285–294 (1988).
[4]T. Deelman and C. M. Connine, "Missing information in spoken word recognition: Nonreleased stop consonants," J. Exp. Psychol. Hum. Percept. Perform. **27**(3), 656–663 (2001).
[5]A. M. A. Ali, J. V. der Spiegel, and P. Mueller, "Acoustic-phonetic features for the automatic classification of stop consonants," IEEE Trans. Speech Audio Process. **9**, 833–841 (2001).
[6]M. F. Dorman, M. Studdert-Kennedy, and L. J. Raphael, "Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues," Percept. Psychophys. **22**(2), 109–122 (1977).
[7]P. Niyogi and M. M. Sondhi, "Detecting stop consonants in continuous speech," J. Acoust. Soc. Am. **111**, 1063–1076 (2002).
[8]F. Malbos, M. Baudry, and S. Montresor, "Detection of stop consonants with the wavelet transform," in Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (1994), pp. 612–615.
[9]A. Jansen and P. Niyogi, "Modeling the temporal dynamics of distinctive feature landmark detector for speech recognition," J. Acoust. Soc. Am. **124**, 1739–1758 (2008).
[10]J. S. Garofolo, "TIMIT acoustic-phonetic continuous speech corpus," LDC, Philadelphia (1993).

[11]Y. Zheng, "Acoustic modeling and feature selection for speech recognition," Ph.D. thesis, University of Illinois at Urbana-Champaign (2005).
[12]Y. Homma, "Durational relationship between japanese stops and vowels," J. Phonetics **9**, 273–281 (1981).
[13]S. J. Manuel, S. S. Hufnagel, M. Huffman, K. N. Stevens, R. Carlson, and S. Hunnicutt, "Studies of vowel and consonant reduction," in Proceedings of ICSLP (1992), pp. 943–946.

Construction of the temporal invariants of the time-reversal operator

Franck D. Philippe, Claire Prada, Dominique Clorennec, and Mathias Fink
Laboratoire Ondes et Acoustique, Université Denis Diderot Paris 7, UMRS CNRS 7587, ESPCI 10 rue Vauquelin, 75231 Paris Cedex 05, France
franck.philippe@espci.fr; claire.prada-julia@espci.fr; dominique.clorennec@espci.fr; mathias.fink@espci.fr

Thomas Folégot
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, Italy
folegot@nurc.nato.int

Abstract: This paper proposes a method to construct the temporal Green's function from a scatterer to an array of transducers in a waveguide using free-space back propagation of the eigenvectors of the time-reversal operator (TRO). The monostatic Green's function is obtained as an eigenvector of the TRO which is known with an arbitrary phase; thus the impulse response cannot be obtained by a simple inverse Fourier transform. Assuming that the monochromatic fields obtained by the back propagation of the eigenvectors are in phase at the focal point, the phase correction is determined. Simulations and laboratory experiments are presented.

PACS numbers: 43.60.Tj, 43.30.Vh, 43.30.Gv [JL]
Date Received: January 26, 2009 **Date Accepted:** April 22, 2009

1. Introduction

Time reversal is a self-adaptive technique that provides a mean to focus in inhomogeneous media. It has been studied intensively these past few years in various fields in acoustics[1–6] as well as electromagnetism.[7–9] The DORT method (French acronym for decomposition of the time-reversal operator) is a time-reversal based technique that uses the singular value decomposition (SVD) of the multistatic data matrix (MDM) to detect, locate, and separate two or more targets in an unknown medium. The initial version of DORT is essentially a monochromatic method and the singular vectors of the MDM are the monochromatic invariants of the time-reversal iterative process as demonstrated by Mordant *et al.*[10] Problems arise when time-domain signals, which we call temporal eigenvectors in the following, are needed (for ultrawide band imaging or telecommunication, for example). The principal difficulty comes from the fact that the singular vectors have an undetermined phase term; in other words, if $\mathbf{V}$ is a singular vector of the array response matrix $\mathbf{K}$, then for any phase φ, $e^{i\varphi}\mathbf{V}$ is also a singular vector. Thus, the vectors $\mathbf{V}_p(\omega)$ obtained by SVD and associated with the pth target are incoherent and the time-domain signals cannot be obtained by a simple inverse Fourier transform. This issue has been investigated in inhomogeneous media[11,12] with the introduction of a space-frequency MDM to extract a coherent frequency vector. It was also studied in the cases of waveguides by assuming that the phase of the eigenvectors is a continuous function of the frequency and using the symmetry of the MDM.[10,13] The continuity condition means this method is very sensitive to noise and requires extensive calculation with high frequency sampling rate; thus the method was not applied successfully in real scale experiments.

Here, a more robust and straightforward technique involving a free-space back propagation is proposed to reconstruct the impulse responses from the array to one of the scatterers. A simulation and a laboratory experiment illustrating this technique are presented.

2. Phase synchronization of the eigenvectors of TRO: Theory

For the DORT method, the SVD of the frequency-domain MDM is used to detect the number of isotropic point-like targets in the medium and to retrieve the Green function associated with each target. It was shown that the number of nonzero singular values (SVs) of the MDM provides the number of point-like targets and the corresponding singular vectors give the monochromatic Green function from the array to the targets associated.[10,13,14] Indeed at pulsation frequency ω, the singular vector $\mathbf{u}_p$ is related to the Green's function $\mathbf{G}_p$ connecting the array to the pth target's position $\mathbf{r}_p$ by

$$\mathbf{u}_p(\omega) = e^{i\Phi_{\mathrm{SVD}}(\omega)} \frac{\mathbf{G}_p(\mathbf{r}_p,\omega)}{\|\mathbf{G}_p(\mathbf{r}_p,\omega)\|}, \tag{1}$$

where $\mathbf{r}_p$ is the position of the pth target and Φ_{SVD} is the frequency dependent phase that comes from the SVD. Once the singular vectors are extracted, the position of the targets is found by construction of the pressure field p in the probed space at pulsation frequency ω with the formula

$$p(\mathbf{r},\omega) = \mathbf{u}_p{}^{\dagger}(\omega)\tilde{\mathbf{G}}(\mathbf{r},\omega), \tag{2}$$

where $\tilde{\mathbf{G}}(\mathbf{r},\omega)$ is the computed Green's function at position $\mathbf{r}$ and the dagger superscript denotes transpose conjugation. In a waveguide, an accurate computing the Green's function requires an extensive knowledge of the guide parameters that is rarely achieved in real scale experiments. As a consequence, we propose to use a free-space model to compute the Green's function. In this model, the image obtained by back propagation of the singular vector associated with one target in the waveguide shows several focal spots, as described by the well known method of images.[15] In this theory, the Green's function is decomposed into the free-space Green's function of the real source plus a sum of contributions from virtual sources that are the images of the real source symmetrically positioned about the interfaces. Thus the pth Green's function from the target to the transducer i can be written as

$$[G_p(\mathbf{r}_p,\omega)]_i = \frac{e^{ik|\mathbf{r}_i-\mathbf{r}_p|}}{|\mathbf{r}_i-\mathbf{r}_p|} + \sum_{n=1}^{N} \mathrm{sign}_n \frac{e^{ik|\mathbf{r}_i-\mathbf{r}_n|}}{|\mathbf{r}_i-\mathbf{r}_n|}, \tag{3}$$

where $\mathbf{r}_i$ is the position of the ith transducer of the array, $\mathbf{r}_n$ is the position of the nth image of the source and sign_n is the sign of the nth image.

Using Eqs. (1) and (3), Eq. (2) yields

$$\begin{aligned} p(\mathbf{r}_p,\omega) &= \sum_i e^{-i\Phi_{\mathrm{SVD}}(\omega)} \left[\frac{e^{-ik|\mathbf{r}_i-\mathbf{r}_p|}}{|\mathbf{r}_i-\mathbf{r}_p|} + \sum_{n=1}^{N} \mathrm{sign}_n \frac{e^{-ik|\mathbf{r}_i-\mathbf{r}_n|}}{|\mathbf{r}_i-\mathbf{r}_n|} \right] \frac{e^{ik|\mathbf{r}_i-\mathbf{r}_p|}}{|\mathbf{r}_i-\mathbf{r}_p|} \\ &= e^{-i\Phi_{\mathrm{SVD}}(\omega)} \left[\sum_i \frac{1}{|\mathbf{r}_i-\mathbf{r}_p|^2} + \sum_{n=1}^{N} \sum_i \mathrm{sign}_n \frac{e^{-ik|\mathbf{r}_i-\mathbf{r}_n|}}{|\mathbf{r}_i-\mathbf{r}_n|} \frac{e^{ik|\mathbf{r}_i-\mathbf{r}_p|}}{|\mathbf{r}_i-\mathbf{r}_p|} \right]. \end{aligned} \tag{4}$$

If the second term of the expansion is small, i.e., the sum of the scalar products between the free-space Green's function of the real source and the images is small, the phase at the focal spot gives $\Phi_{\mathrm{SVD}}(\omega)$. As the distance between the real source and the nth image grows with n, this condition is equivalent to an absence of overlap between the focal spots of the two first images and the focal spot of the direct path. Using the focal point of the real source is more robust than using one of the images' as the isophases at this point can be considered as planes in the small angle approximation. Thus, if the choice of location of the focal spot is not accurate enough, the error made is minimized. Once $\Phi_{\mathrm{SVD}}(\omega)$ is extracted from the numerical back propagation, the phase of the singular vectors is corrected and the time-domain eigenvector of the TRO is obtained by a simple inverse Fourier transform. This correction corresponds to a synchronization of the eigenvectors (SVP, for "synchronisation des vecteurs propres" in French) for all frequen-

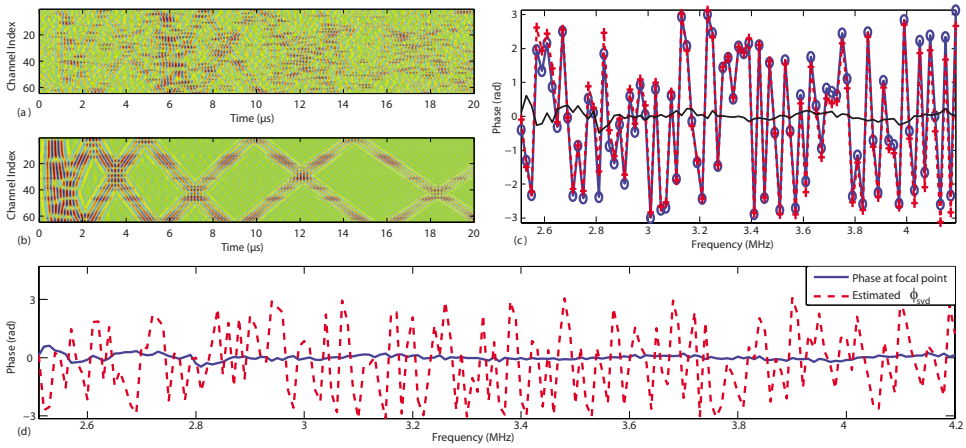


Fig. 1. (Color online) Inverse Fourier transform of the first singular vector of the simulated MDM (a) without phase correction and (b) with phase correction. (c) Phase as a function of frequency of the simulated Green's function from the 19th transducer to the target (blue line with circles) compared to the phase of the SVP eigenvector (red line with crosses). The black solid line is the difference between the two (d): (blue solid line) phase at focal point after back propagation of the corrected eigenvector in the RAM model and (red dotted line) phase at focal point after back propagation of the noncorrected eigenvector in the free-space model which is used as the $\Phi_{\mathrm{SVD}}(\omega)$ estimate.

cies. The temporal eigenvector is then back propagated in the real waveguide to ensure temporal focusing of the emitted signal at the target. We focus here on the reconstruction of time-domain invariants of the TRO for shallow water undersea acoustics, but this technique can be extended to various waveguide configurations. In Secs. 3 and 4, we present a broadband simulation made with RAM code[16] and experimental results from an ultrasound laboratory experiment. The simulation will provide quantification of the error made by using the phase of the back propagation in free space to estimate the phase $\Phi_{\mathrm{SVD}}(\omega)$ in Eq. (4) in order to synchronize the singular vectors.

3. Simulation

The simulation configuration is a Pekeris waveguide composed of a $d=30$ mm water layer above a Plexiglas-like medium. A target is placed at 400 mm distance and 12 mm depth. These parameters and the bandwidth (2.5–4.2 MHz) are chosen to match the experimental setup presented in Sec. 4. The temporal eigenvectors obtained before and after phase correction are displayed on Figs. 1(a) and 1(b). The first signal obtained by direct inverse Fourier transform is highly incoherent whereas the signal obtained from the eigenvector after phase compensation shows perfect coherence. To compare this last signal with the computed Green's function, Fig. 1(c) shows the frequency dependent phase of the Green's function from the target to the 19th transducer, the corresponding phase of the SVP eigenvector and their difference (continuous line). This figure shows that in this configuration, the error made by neglecting the phase of the second term in Eq. (4) is small enough ($<\pi/12$) to construct a near perfect temporal Green's function. Moreover, to confirm the phase coherence of the SVP eigenvector, it was back propagated in the simulated waveguide at each frequency. The phase of the resulting field at focal point is plotted on Fig. 1(d) and compared to the estimated $\Phi_{\mathrm{SVD}}(\omega)$ which is the phase of $p(\mathbf{r}_p,\omega)$ in Eq. (4). The error is again lower than $\pi/12$, which confirms that the second term in Eq. (4) only slightly modifies the SVP phase estimate.

4. Experiment description and results

In this section, results from an ultrasound water tank experiment are presented. Two wires considered as point-like targets are placed at a distance of 400 mm at 3 mm from each other. As mentioned before, the waveguide parameters are the same as in Sec. 3. The broadband impulse

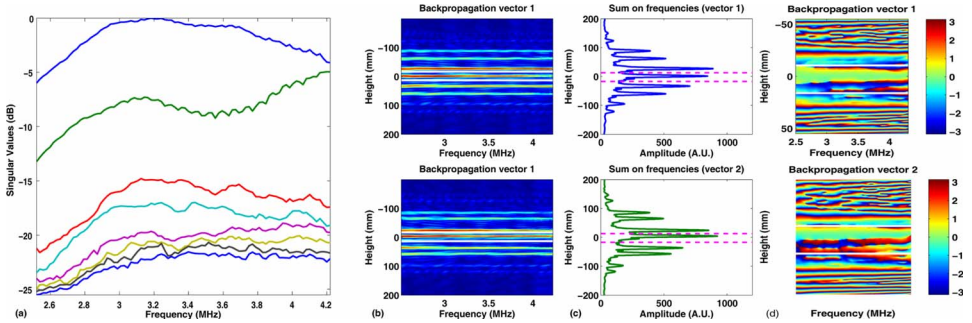


Fig. 2. (Color online) (a) Experimental SVs of the MDM as a function of frequency in logarithmic scale. As predicted, two SVs are above noise. (b) Back propagation in free space, of the two first eigenvectors as function of frequency. The first SV is associated with the lowest target and the second SV with the other target. (c) Amplitude averaged on all frequencies as a function of height in the waveguide. The two dotted lines (magenta) show the positions of the interfaces. (d) Phase of the pressure field at distance 400 mm after back propagation of the first two corrected eigenvectors as a function of frequency and height in the waveguide

response matrix is recorded with a 64 transducer array of central frequency 3.5 MHz. The MDM is obtained at each frequency by time windowing and Fourier transform. On Fig. 2(a), the SV of the MDM are plotted as a function of frequency. As is clearly seen, there are two distinct nonzero SV, each associated with one target. It is important to note that the reconstruction of the temporal singular vector assumes that the SV is associated with the same target in the whole bandwidth. This assertion is confirmed in Figs. 2(a) and 2(b) by plotting the focal spot at target distance as a function of frequency. Difficulties can arise if SV switching occurs, especially in the case of target resonances as presented by Philippe *et al.*[17] As predicted by the image theory, the focusing pattern is made of several focus spots for one target: one in the vertical extension corresponding to the guide and eight others outside the guide. The depth of the target in the guide is extracted from the directivity diagram in Fig. 2(b) and $\Phi_{SVD}(\omega)$ is obtained from the equivalent in phase of Fig. 2(b). Figure 2(d) shows the phase of the pressure field as a function of frequency and vertical range in free space after correction of the eigenvector and back propagation in the model. As expected, the phase at the focal point is zero. More interesting is the fact that the phase remains negligible in a large area surrounding the focal spot. It means that a certain amount of error in the choice of the point taken as reference for the phase correction is acceptable in the SVP method. The time-domain signals obtained from the inverse Fourier transform of the corrected eigenvectors are presented in Fig. 3. The signals are then back propa-

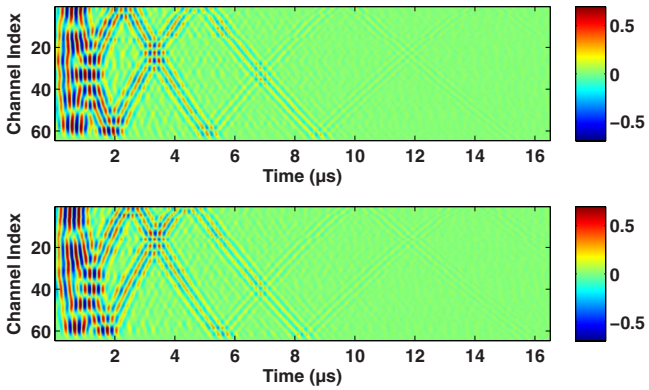


Fig. 3. (Color online) Reconstructed two first SVP temporal eigenvectors for the ultrasonic scale example.

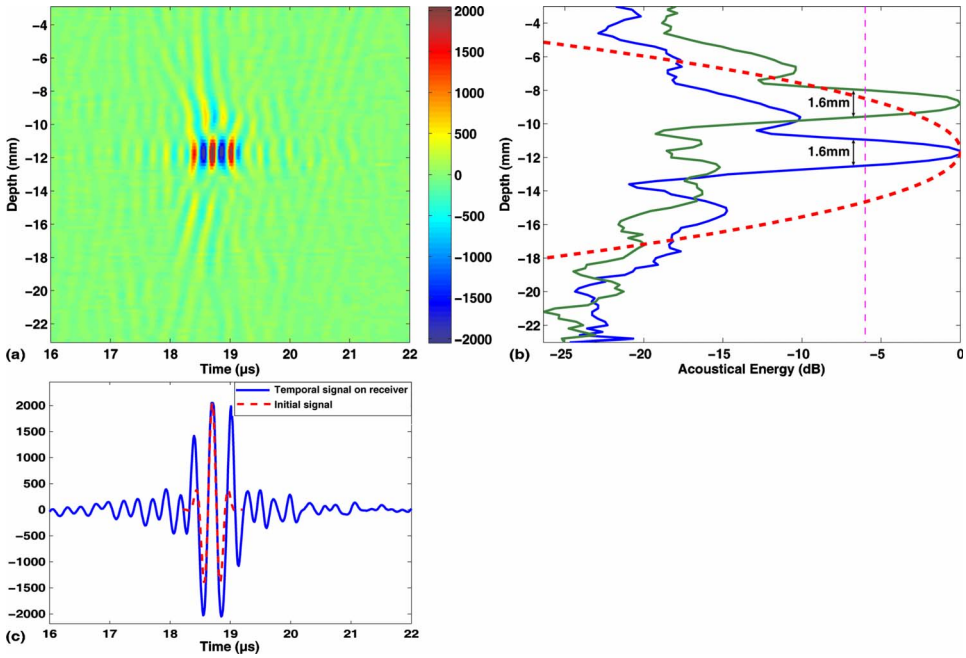


Fig. 4. (Color online) (a) Pressure field as a function of time at distance of 400 mm after emission of the first temporal eigenvector. (b) Sum of energy as a function of depth. The focal spot of the emission of the first eigenvector (blue) and second eigenvector (green) is six times smaller than the diffraction limited free-space focal spot (red dots). (c) The time signal at target position (blue) shows temporal focalization of the 13 μs long emission.

gated in the real waveguide and the pressure field is measured with a hydrophone over the depth at distance 400 mm. The results are plotted in Fig. 4.

As for the simulated data set, the transmission of the two first corrected singular vectors in the guide focuses at each target with higher resolution than in free space [see Fig. 4(b)]. In this case, the resolution is increased by a factor of 6 and the temporal eigenvectors show seven different significant paths. Moreover, and this is the foremost thing to see, Fig. 4(c) shows that the reconstruction of the temporal impulse response was effective as the 13 μs long emitted signal is compressed by the propagation in the medium into a signal matching the 1 μs long initial impulse.

5. Summary and discussion

This paper presents a simple method to compute the time-domain invariants of the time-reversal process in a waveguide from the DORT method. The SVP method is based on a monochromatic free-space imaging of the studied medium requiring little information compared to a much more complex model taking into account the waveguide properties. The free-space model allows for extraction of the frequency dependent phase added to the monochromatic singular vectors by the SVD. Once this phase is corrected, the time-domain invariants are built by a simple inverse Fourier transform of the set of monochromatic DORT singular vectors. This method is applied to simulation and laboratory data set, and emission of the time-domain signal obtained in the medium shows spatial and temporal focusing. The effectiveness of the SVP method on laboratory data with two targets shows how robust this simple method is. In this paper, emphasis was put on the underwater acoustics application of the method but it can be applied to any waveguide configuration provided a model allowing focalization of the direct path is available. Further investigations concerning the sensitivity of the method to environmental mismatch such as sound speed gradient are scheduled.

References and links

[1]M. Fink, C. Prada, F. Wu, and D. Cassereau, "Self focusing with time-reversal mirror in inhomogeneous media," Proc.-IEEE Ultrason. Symp. **2**, 681–686 (1989).

[2]M. Fink, "Time-reversal of the ultrasonics fields—Part I: Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).

[3]M. Pernot, J. F. Aubry, M. Tanter, A. L. Boch, F. Marquet, M. Kujas, D. Seilhean, and M. Fink, "In-vivo transcranial brain surgery with an ultrasonic time reversal mirror," J. Neurosurg. **106**, 1061–1066 (2007).

[4]J.-L. Thomas, F. Wu, and M. Fink, "Time reversal focusing applied to lithotripsy," Ultrason. Imaging **18**, 106–121 (1996).

[5]D. R. Jackson and D. R. Dowling, "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181 (1991).

[6]W. A. Kuperman, W. S. Hodgkiss, and H. C. Song, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).

[7]P. Kyritsi, G. Papanicolaou, P. Eggers, and A. Oprea, "MISO time reversal and delay-spread compression for FWA channels at 5 GHz," IEEE Antennas Wireless Propag. Lett. **3**, 96–99 (2004).

[8]B. E. Henty and D. D. Stancil, "Multipath-enabled super-resolution for RF and microwave communication using phase-conjugate arrays," Phys. Rev. Lett. **93**, 243904 (2004).

[9]G. Lerosey, J. de Rosny, A. Tourin, A. Derode, G. Montaldo, and M. Fink, "Time reversal of electromagnetic waves," Phys. Rev. Lett. **92**, 193904 (2004).

[10]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).

[11]M. E. Yavuz and F. L. Teixeira, "A numerical study of time reversed UWB electromagnetic waves in continuous random media," IEEE Antennas Wireless Propag. Lett. **4**, 43–46 (2005).

[12]M. E. Yavuz and F. L. Teixeira, "Space–frequency ultrawideband time-reversal imaging," IEEE Trans. Geosci. Remote Sens. **46**, 1115–1124 (2008).

[13]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).

[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**, 761–768 (2007).

[15]L. M. Brekhovskikh, *Fundamentals of Ocean Acoustics*, 3rd ed. (Springer, New York, 2003).

[16]M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

[17]F. D. Philippe, C. Prada, J. de Rosny, D. Clorennec, J.-G. Minonzio, and M. Fink, "Characterization of an elastic target in a shallow water waveguide by decomposition of the time-reversal operator," J. Acoust. Soc. Am. **124**, 779–787 (2008).

Time-domain Kirchhoff model for acoustic scattering from an impedance polygon facet

Keunhwa Lee and Woojae Seong
Department of Ocean Engineering, Seoul National University, Seoul 151-744, Korea
nasalkh2@snu.ac.kr, wseong@snu.ac.kr

Abstract: Kirchhoff formula for an impedance polygon facet is given in the time domain. The derived formula is expressed as a summation of the transient analytic functions and generalized functions and represents an impulse response of the impedance polygon facet. Current formula can be applied to transient scattering analysis of underwater objects such as fish and submarine, or rough surface in the geometrical scattering region.

PACS numbers: 43.30.Gv, 43.20.Fn, 43.55.Br [JL]
Date Received: December 4, 2008 **Date Accepted:** April 22, 2009

1. Introduction

The Kirchhoff method has been widely used to solve high frequency scattering problems in different applications, such as scattering from a sound diffuser,[1] ocean surface scattering,[2] and acoustic target scattering.[3] This method usually holds for the high frequency region $kL > 1$, where k is the wavenumber and L is the characteristic length.

For a general three-dimensional object, the scattered field obtained by Kirchhoff method is expressed as an integral over the surface of the object and numerically calculated by the discretization of this surface integral. If the scattered field from an arbitrary facet is known, the surface of the object can be divided into polygon facets and then the total scattered field will be obtained as a summation of contributions from all these facets.

Analytic evaluation of the scattered field in the time domain employing Kirchhoff method (or physical optics method in electromagnetism) was initially studied by Kennaugh and Cosgriff.[4] They derived the impulse response for a perfectly conducting simple body, equivalent to a soft body in acoustics. Lee *et al.*[5] derived an impulse response from a triangular facet of a perfectly conducting body. Their impulse response is expressed as a summation of amplified rectangular functions but the formula suffers from a singularity problem for some incident/scattered directions. In acoustics, Fawcett[6] also presented the time signal scattered from a triangular rigid facet by means of the Kirchhoff/diffraction method. Fawcett's solution is expressed as a summation of time-delayed and amplitude-modified form of the time integral of the source pulse. In fact, Fawcett's derivation can be regarded as an acoustic version of the work of Lee *et al.* and therefore still bears the problem of singularity for some incident/scattered directions.

In this paper, our goal is to derive a general acoustic impulse response for a polygon facet, such as a panel. Our derivation is more general and rigorous than those of previous authors, in that we treat an impedance polygon facet and have removed the singularity of the impulse response. This derived formula can be used for modeling the high frequency transient scattering from an underwater target or a rough surface, when they are discretized into multiple polygons.

2. Frequency domain solution for an impedance polygon facet

2.1 Helmholtz–Kirchhoff integral equation and Kirchhoff assumption

The Helmholtz–Kirchhoff integral equation of the acoustic pressure scattered from a body is given in the frequency domain by

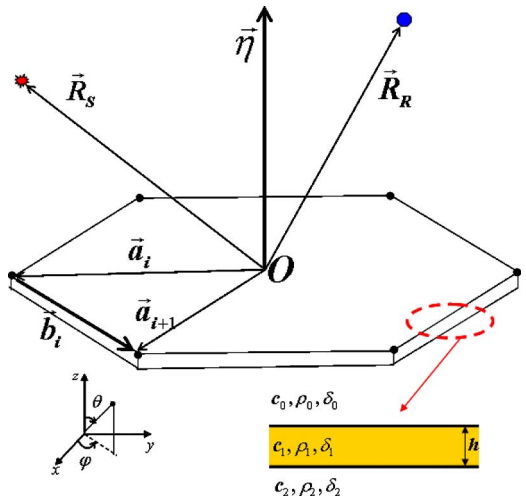


Fig. 1. (Color online) Definition of coordinate system and vectors used for the general impedance polygon facet.

$$p_s(\boldsymbol{r}|\boldsymbol{r}_0) = \int_{S_B} [p_t(\boldsymbol{r}'|\boldsymbol{r}_0)\,\nabla\, G(\boldsymbol{r}|\boldsymbol{r}') - G(\boldsymbol{r}|\boldsymbol{r}')\,\nabla\, p_t(\boldsymbol{r}'|\boldsymbol{r}_0)]d\boldsymbol{S}_B. \tag{1}$$

Here, $\boldsymbol{r}_0$ is the source position vector, $\boldsymbol{r}$ is the field position vector, and $\boldsymbol{r}'$ is the position vector on the body surface S_B. p_t is the total acoustic field, where $p_t = p_i + p_s$ with the incident field p_i given by $e^{jk_0|\boldsymbol{r}-\boldsymbol{r}_0|}/(4\pi|\boldsymbol{r}-\boldsymbol{r}_0|)$ and p_s is the scattered field. $G(\boldsymbol{r}|\boldsymbol{r}')$ is the free-field Green's function given by $e^{jk_0|\boldsymbol{r}-\boldsymbol{r}'|}/(4\pi|\boldsymbol{r}-\boldsymbol{r}'|)$, where k_0 is the medium wavenumber.

For an impedance facet, Kirchhoff assumption of the scattered field is obtained with the plane-wave reflection coefficient $\hat{R}(\omega)$, dependent on the angular frequency ω, given by

$$p_s = \hat{R}(\omega)p_i, \quad \frac{\partial p_s}{\partial n} = -\hat{R}(\omega)\frac{\partial p_i}{\partial n} \quad \text{on illuminated part},$$

$$p_s = 0, \quad \frac{\partial p_s}{\partial n} = 0 \quad \text{on shadowed part}, \tag{2}$$

where $\partial/\partial n$ is the directional derivative for outward normal direction from the body surface.

When Eq. (2) is inserted into Eq. (1), the scattered pressure is approximated and arranged as follows:

$$p_s(\boldsymbol{r}|\boldsymbol{r}_0) = \int_{S_B} \left[(1+\hat{R})p_i(\boldsymbol{r}'|\boldsymbol{r}_0)\frac{\partial G(\boldsymbol{r}|\boldsymbol{r}')}{\partial n} - (1-\hat{R})G(\boldsymbol{r}|\boldsymbol{r}')\frac{\partial p_i(\boldsymbol{r}'|\boldsymbol{r}_0)}{\partial n} \right] dS_B. \tag{3}$$

2.2 Frequency domain solution for an impedance polygon facet

It is well known that the surface integral of Eq. (3) can be reduced to a line integral.[7] For an opaque screen with polygon aperture, Gordon[8] showed that the line integral can be rearranged to an analytic closed expression using the far-field approximation. Based on Gordon's derivation, Eq. (3) can be rearranged into an analytic closed expression for a polygon facet.

For the case of non-specular incident/scattered angle, using the definition of vectors shown in Fig. 1, Eq. (3) becomes

$$p_s(\boldsymbol{r}|\boldsymbol{r}_0)=\frac{e^{jk_0(R_R+R_S)}}{(4\pi)^2R_SR_R}\frac{[(\boldsymbol{R}_n+\boldsymbol{\varsigma})+\hat{R}(\omega)(\boldsymbol{R}_n-\boldsymbol{\varsigma})]\cdot\boldsymbol{\eta}}{W^2}\sum_{i=1}^{n}(\boldsymbol{w}^*\cdot\boldsymbol{b}_i)\mathrm{sinc}\left(k_0\frac{\boldsymbol{w}\cdot\boldsymbol{b}_i}{2}\right)e^{jk_0\boldsymbol{w}/2\cdot(\boldsymbol{a}_{i+1}+\boldsymbol{a}_i)}, \tag{4}$$

where n is total number of vertices of the polygon facet, $\boldsymbol{R}_R$ is vector drawn from the origin of the facet to field point, $R_R=|\boldsymbol{R}_R|$, $\boldsymbol{R}_S$ is vector drawn from the origin of the facet to source point, $R_S=|\boldsymbol{R}_S|$, $\boldsymbol{R}_n=\boldsymbol{R}_R/|\boldsymbol{R}_R|$, $\boldsymbol{\zeta}=-\boldsymbol{R}_S/|\boldsymbol{R}_S|$, $\boldsymbol{\eta}$ is normal vector to the facet, $\boldsymbol{w}=\boldsymbol{\zeta}-\boldsymbol{R}_n$, $\boldsymbol{w}^*=\boldsymbol{w}\times\boldsymbol{\eta}$, W is projection length of $\boldsymbol{w}$ onto the facet, $\boldsymbol{a}_i$ is vector drawn from the origin of the facet to ith vertex, and $\boldsymbol{b}_i=\boldsymbol{a}_{i+1}-\boldsymbol{a}_i$ with $\boldsymbol{a}_{n+1}=\boldsymbol{a}_1$.

For the case of specular incident/scattered angle ($W=0$), Eq. (3) becomes

$$p_s(\boldsymbol{r}|\boldsymbol{r}_0)=-jk_0[(\boldsymbol{R}_n+\boldsymbol{\varsigma})+\hat{R}(\omega)(\boldsymbol{R}_n-\boldsymbol{\varsigma})]\cdot\boldsymbol{\eta}\frac{e^{jk_0(R_R+R_S)}}{(4\pi)^2R_SR_R}\cdot A, \tag{5}$$

where A is the area of the facet.

Equations (4) and (5) are given in equivalent but slightly different forms in Eqs. (4.4)–(4.6) of Gordon's paper. We assume a spherical incident field impinging on an impedance facet and use the surface integral of Eq. (3) including the incident scattered part, given as the first term on the right-hand side of Eqs. (4) and (5). Note that the incident scattered part disappears in the mono-static case, i.e., when $\boldsymbol{R}_R=\boldsymbol{R}_S$.

Here, the plane-wave reflection coefficient for an acoustic facet, illustrated schematically in the lower part of Fig. 1, can be written in a series form[9] as

$$\hat{R}(\omega)=R_{01}+(1-R_{01}^2)R_{12}\sum_{l=1}^{\infty}(-R_{01}R_{12})^{l-1}\exp(j2\gamma_1h_1l), \tag{6}$$

where $\gamma_i=\sqrt{k_i^2-k^2}$ is the vertical wavenumber of the ith medium (subscript i refers to the medium number, $i=0,1,2$), k_i is the medium wavenumber, and k is the horizontal wavenumber. The medium wavenumber is given by $k_i=(\omega/c_i)(1+j\delta_i)$, where the acoustic loss tangent δ_i is related to the attenuation α_i (dB/wavelength) by $\alpha_i=40\pi(\log_{10}e)\delta_i$ and c_i is the acoustic sound speed. Horizontal wavenumber $k=k_0\cos\theta_i$ when the incident grazing angle is θ_i. R_{01} and R_{12} are the plane-wave reflection coefficients and h_1 is the thickness of the facet.

We note that the frequency dependence of Eq. (6) is attributed to the phase of the exponential term and the conjugate symmetry of the local plane-wave reflection coefficients R_{01} and R_{12} for the frequency.

To reveal the frequency-dependency of Eq. (6) explicitly, it can be rearranged as follows:

$$\hat{R}(\omega)=\sum_{l=0}^{\infty}B_le^{-j\tau_l\,\mathrm{sgn}(\omega)}\exp\left[-\mathrm{Im}\left(\frac{2h_1l}{c_{z1}}\right)|\omega|\right]\exp\left[j\,\mathrm{Re}\left(\frac{2h_1l}{c_{z1}}\right)\omega\right], \tag{7}$$

where $c_{z1}=\omega/\gamma_1$, $B_0=|R_{01}|$, and $\tau_0=\arg(R_{01})$ when $l=0$; $B_l=|(1-R_{01}^2)R_{12}|(|R_{01}||R_{12}|)^{l-1}$ and $\tau_l=[\arg(R_{01})+\arg(R_{12})+\pi](l-1)+\arg[(1-R_{01}^2)R_{12}]$ when $l>0$.

In Eq. (7), τ_l and B_l are the frequency-independent parameters. The first exponential term $\exp[-j\tau_l\,\mathrm{sgn}(\omega)]$ results from total reflection of the incident wave and is slightly affected by the attenuation of the medium. To further simplify this relationship, the frequency-independent variables of the second and third exponential terms are set to

$$\alpha_l=\mathrm{Re}(2h_1l/c_{z1}),\quad \beta_l=\mathrm{Im}(2h_1l/c_{z1}). \tag{8}$$

Then, Eq. (7) is expressed as

$$\hat{R}(\omega) = \sum_{l=0}^{\infty} B_l e^{-j\tau_l \,\mathrm{sgn}(\omega)} e^{-\beta_l|\omega|} e^{j\alpha_l\omega}. \tag{9}$$

Equations (4) and (5) along with the plane-wave reflection coefficient of Eq. (9) are the frequency domain solution of the scattered pressure from an impedance polygon facet, which will be transformed into the time domain in Sec. 3.

3. Impulse function for an impedance polygon facet

The impulse function of the scattered pressure is obtained by using the Fourier transform of the frequency domain solutions of Eq. (4) or Eq. (5) as follows:

$$H_s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} p_s(\omega) e^{-j\omega t} d\omega. \tag{10}$$

Then, the time scattered signal $P(t)$ with the source function $S(t)$ can be calculated by the convolution of $S(t)$ with the impulse function as

$$P(t) = \int_{-\infty}^{\infty} S(\nu) H_s(t-\nu) d\nu. \tag{11}$$

3.1 Non-specular incident/scattered angle region

Observing Eq. (10) with Eqs. (4) and (9), the basis function of impulse function for non-specular incident/scattered angle region can be expressed as

$$h_l^i(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (\boldsymbol{w}^* \cdot \boldsymbol{b}_i) B_l e^{-j\tau_l \,\mathrm{sgn}(\omega)} \,\mathrm{sinc}\left(\omega \frac{\boldsymbol{w}\cdot\boldsymbol{b}_i}{2c_0}\right) e^{-\beta_l|\omega|} \exp\left[-j\omega\left(t - \frac{1}{c_0}\left(R_R + R_S + \frac{\boldsymbol{w}}{2}\cdot(\boldsymbol{a}_{i+1}+\boldsymbol{a}_i)\right) - \alpha_l\right)\right] d\omega. \tag{12}$$

A little manipulation of Eq. (12) yields

$$h_l^i(t) = \frac{B_l \cos\tau_l}{2\pi}\int_{-\infty}^{\infty} (\boldsymbol{w}^* \cdot \boldsymbol{b}_i) \mathrm{sinc}\left(\omega \frac{\boldsymbol{w}\cdot\boldsymbol{b}_i}{2c_0}\right) e^{-\beta_l|\omega|} \exp\left[-j\omega\left(t - \frac{1}{c_0}\left(R_R + R_S + \frac{\boldsymbol{w}}{2}\cdot(\boldsymbol{a}_{i+1}+\boldsymbol{a}_i)\right) - \alpha_l\right)\right] d\omega + \frac{B_l \sin\tau_l}{\pi}\int_0^{\infty} (\boldsymbol{w}^* \cdot \boldsymbol{b}_i) \mathrm{sinc}\left(\omega \frac{\boldsymbol{w}\cdot\boldsymbol{b}_i}{2c_0}\right) e^{-\beta_l\omega} \sin\left[\omega\left(t - \frac{1}{c_0}\left(R_R + R_S + \frac{\boldsymbol{w}}{2}\cdot(\boldsymbol{a}_{i+1}+\boldsymbol{a}_i)\right) - \alpha_l\right)\right] d\omega. \tag{13}$$

The second term of Eq. (13), the so-called refracted precursor, is significant mainly when total reflection of the incident wave happens.[2] Since the total reflection does not occur near the vertical incidence region where the Kirchhoff approximation is valid, we can neglect the second term of Eq. (13).

Analytic integration of the first term of Eq. (13) gives us the basis function of the impulse response as

$$h_l^i(t) = \frac{c_0 B_l \cos\tau_l}{\pi} \frac{(\boldsymbol{w}^* \cdot \boldsymbol{b}_i)}{(\boldsymbol{w} \cdot \boldsymbol{b}_i)} \left[\arctan\left(\frac{t - \frac{R_R + R_S + \boldsymbol{w} \cdot \boldsymbol{a}_i}{c_0} - \alpha_l}{\beta_l} \right) - \arctan\left(\frac{t - \frac{R_R + R_S + \boldsymbol{w} \cdot \boldsymbol{a}_{i+1}}{c_0} - \alpha_l}{\beta_l} \right) \right], \tag{14}$$

where $-\pi/2 \leq \arg(\arctan) \leq \pi/2$.

Equation (14) is a general impulse function, which implies the expression of Lee *et al.*[5] We can easily see that when β_l goes to zero, depending on the arctangent function values, Eq. (14) becomes either a rectangular function or zero.

Although the arctangent function of Eq. (14) is regular, Eq. (14) has a singularity when $\boldsymbol{w} \cdot \boldsymbol{b}_i = 0$. $\boldsymbol{w} \cdot \boldsymbol{b}_i = 0$ means that line vector of the polygon facet is vertical to $(\boldsymbol{\zeta} - \boldsymbol{R}_n)$. In this case, the first term of Eq. (13) is rearranged as

$$h_l^i(t) = \frac{B_l \cos\tau_l}{2\pi} \int_{-\infty}^{\infty} (\boldsymbol{w}^* \cdot \boldsymbol{b}_i) e^{-\beta_l|\omega|} \exp\left[-j\omega \left(t - \frac{1}{c_0} \left(R_R + R_S + \frac{\boldsymbol{w}}{2} \cdot (\boldsymbol{a}_{i+1} + \boldsymbol{a}_i) \right) - \alpha_l \right) \right] d\omega. \tag{15}$$

Then, the impulse function is obtained as follows:

$$h_l^i(t) = \frac{\beta_l B_l \cos\tau_l (\boldsymbol{w}^* \cdot \boldsymbol{b}_i)}{\pi \left[\left(t - \frac{R_R + R_S + \boldsymbol{w} \cdot (\boldsymbol{a}_{i+1} + \boldsymbol{a}_i)/2}{c_0} - \alpha_l \right)^2 + \beta_l^2 \right]}. \tag{16}$$

Note that the impulse function [Eq. (16)] is regular when $\beta_l \neq 0$. In case when $\beta_l = 0$, Eq. (15) is transformed into

$$h_l^i(t) = B_l \cos\tau_l (\boldsymbol{w}^* \cdot \boldsymbol{b}_i) \delta\left(t - \frac{R_R + R_S + \boldsymbol{w} \cdot (\boldsymbol{a}_{i+1} + \boldsymbol{a}_i)/2}{c_0} - \alpha_l \right). \tag{17}$$

Equation (17) is the basis function when $\boldsymbol{w} \cdot \boldsymbol{b}_i = 0$ and $\beta_l = 0$. For the general purpose, the delta function of Eq. (17) can be evaluated by $\sin(\omega_{\max} t)/\pi t$, with the maximum angular frequency $\omega_{\max}$ higher than the half bandwidth of the source signal.[10]

Using the basis functions of Eqs. (14), (16), and (17), the impulse response of a three-layered polygon facet is arranged as

$$H_s(t) = \frac{1}{(4\pi)^2 W^2 R_S R_R} \left[(\boldsymbol{R}_n + \boldsymbol{\varsigma}) \cdot \boldsymbol{\eta} \sum_{i=1}^{n} h_{-1}^i(t) + (\boldsymbol{R}_n - \boldsymbol{\varsigma}) \cdot \boldsymbol{\eta} \sum_{l=0}^{\infty} \sum_{i=1}^{n} h_l^i(t) \right]. \tag{18}$$

Here, $h_{-1}^i(t)$ represents the basis function with $B_{-1} = 1$, $\tau_{-1} = 0$, $\alpha_{-1} = 0$, and $\beta_{-1} = 0$.

3.2 Specular incident/scattered angle region

For specular incident/scattered angle region, using Eqs. (5), (9), and (10), the basis function can be defined by

$$h_l^0(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-j) \frac{B_l}{c_0} \omega e^{-j\tau_l \operatorname{sgn}(\omega)} e^{-\beta_l|\omega|} \exp\left[-j\omega \left(t - \frac{R_R + R_S}{c_0} - \alpha_l \right) \right] d\omega. \tag{19}$$

Neglecting the refracted precursor term from Eq. (19), the result is

$$h_l^0(t) = -\frac{B_l \cos\tau_l}{\pi c_0}\int_0^\infty \omega e^{-\beta_l\omega}\sin\left[\omega\left(t-\frac{R_R+R_S}{c_0}-\alpha_l\right)\right]d\omega. \tag{20}$$

Equation (20) is analytically integrated, when $\beta_l \neq 0$, as

$$h_l^0(t) = -\frac{2B_l \cos\tau_l}{\pi c_0}\frac{\beta_l\left(t-\frac{R_R+R_S}{c_0}-\alpha_l\right)}{\left[\beta_l^2+\left(t-\frac{R_R+R_S}{c_0}-\alpha_l\right)^2\right]^2}. \tag{21}$$

Also, in case when β_l=0, Eq. (20) is calculated in the sense of the generalized limit as

$$h_l^0(t) = \frac{B_l \cos\tau_l}{c_0}\frac{d\delta\left(t-\frac{R_R+R_S}{c_0}-\alpha_l\right)}{dt}. \tag{22}$$

The time derivative of delta function in Eq. (22), $d\delta(t)/dt$, can be evaluated equivalently by $d(\sin(\omega_{\max}t)/\pi t)/dt$, which is regular for all time arguments.

The impulse response function for specular incident/scattered angle region is arranged with the basis functions of Eqs. (21) and (22) as follows:

$$H_s(t) = \frac{\mathrm{A}}{(4\pi)^2 R_S R_R}\left[((\boldsymbol{R}_n+\boldsymbol{\varsigma})\cdot\boldsymbol{\eta})h_{-1}^0(t) + ((\boldsymbol{R}_n-\boldsymbol{\varsigma})\cdot\boldsymbol{\eta})\sum_{l=0}^{\infty}h_l^0(t)\right]. \tag{23}$$

Here, $h_{-1}^0(t)$ represent the basis function where B_{-1}=1, τ_{-1}=0, α_{-1}=0, and β_{-1}=0.

Finally, we mention the causality of the basis functions. Equations (14), (16), and (21) seem to violate the time causality. In fact, these basis functions have some finite value for all times. When the total reflection does not occur, it is because of the frequency-independent attenuation assumed for a polygon facet. This loss term, causing the frequency-independent sound speed and the attenuation to increase linearly with frequency, does not satisfy the requirement of causality. Upon total reflection, the continuous re-radiation of the attenuating wave within a polygon facet, considered to be an infinite panel by the Kirchhoff assumption, is mainly attributed to the non-causality of the time signal. However, since the amount of the pre-arrivals is much smaller than the true arrival, these basis functions can be regarded approximately as functions satisfying the requirement of causality.

4. Numerical results

In this section, the direct time-domain solution of Eq. (11) along with Eqs. (18) and (23) is compared with the transformed solution of Eq. (10). To show the impedance facet effect without unnecessary complex multiple arrivals, a thin penetrable acoustic panel with moderate reflection coefficient is chosen. A square panel 15×15 m^2, representative of a facet with discernible scattering contributions from each four side, with a 20 cm thickness is used for the numerical calculation. The panel has a sound speed of 2000 m/s, a density of 2000 $\mathrm{kg/m}^3$, and an attenuation of 0.55 dB/λ and is submerged in the water having a sound speed of 1500 m/s and a density of 1000 $\mathrm{kg/m}^3$. The distance between the center of the panel and the source/receiver location is 1 km. Each line of the panel is parallel to x-axis or y-axis of the rectangular coordinate. A 5 cycle sine wave of 1 kHz is used as the source pulse. Since the panel thickness is smaller than the pulse wavelength, the scattered arrivals from the front and back face are overlapped. The time-domain solution converges very fast and using l=3 is found to be enough.

Figure 2 shows the numerical results at the source/receiver location with the vertical angle of 0.1 rad and the azimuthal angle of 0.0 rad. The transformed solution is calculated by fast Fourier transform with 4096 samples at the sampling frequency of 15 kHz. For consistency, the $\omega_{\max}$ of Eqs. (17) and (22) is chosen as 7.5 kHz, a half of the angular sampling frequency. As

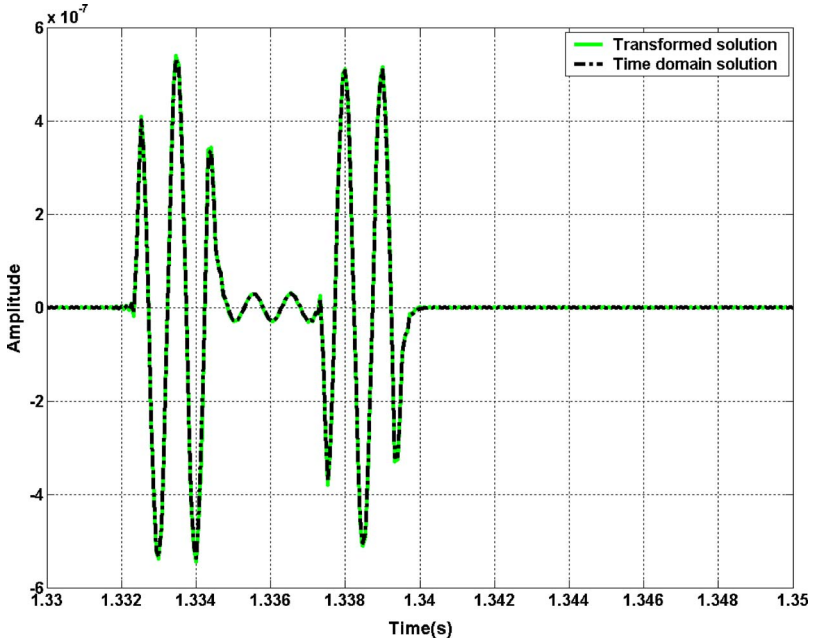


Fig. 2. (Color online) Comparison of the time-domain solution and the solution transformed from the frequency domain at the source/receiver location with the vertical angle $\theta=0.1$ rad and the azimuthal angle $\varphi=0.0$ rad.

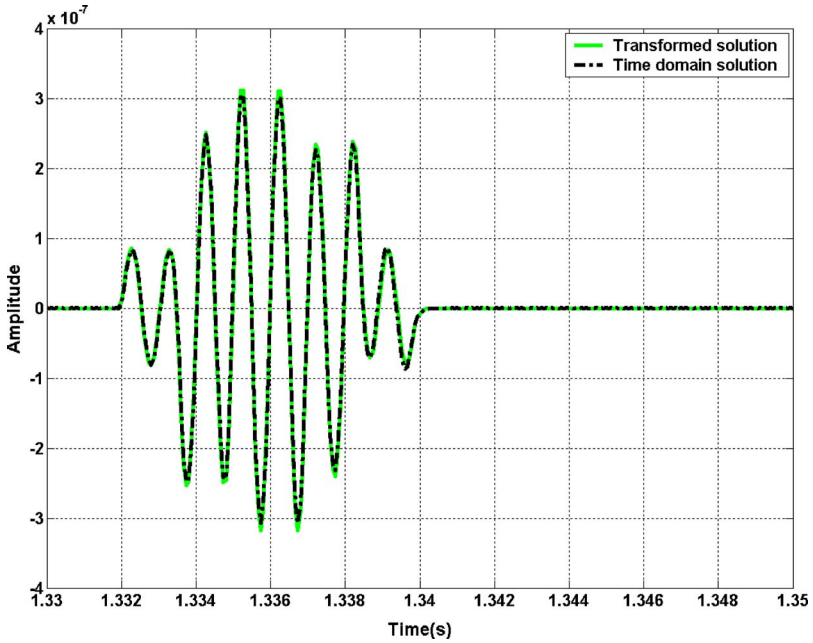


Fig. 3. (Color online) Comparison of the time-domain solution and the solution transformed from the frequency domain at the source/receiver location with the vertical angle $\theta=0.1$ rad and the azimuthal angle $\varphi=\pi/4$ rad.

shown in Fig. 2, the time-domain solution and the transformed solution shows very good agreement. In this example, the scattering occurs mainly at the leading and trailing edges parallel to y-axis. From Eq. (17), since the $\boldsymbol{b}_i$'s for this two edges are in opposite direction, they are out of phase, as can be seen in Fig. 2.

Figure 3 shows the variation of the scattered pulse for the location between source/receiver and target. Aside from the differing azimuthal angle of $\pi/4$ rad, other conditions are the same as those considered in Fig. 2. In Fig. 3, the scattering occurs at all edges but its contribution is smaller than that of Fig. 2 because the viewing angle is farther from the specular direction from the edges. Also, the received signal becomes smoother and longer than the original pulse because of a dispersion resulting from the layering and the finite size of the panel. We observe that the time-domain solution matches the transformed solution.

5. Conclusions

An impulse response function for an acoustic scattering from an impedance polygon facet is derived using the Kirchhoff method. The derived function is expressed as a summation of regular functions and shows good agreement with the transformed solution of the frequency domain. This formula is useful for modeling the broadband scattering signal and benchmarking the transformed solution for an arbitrarily shaped scatterer, based on Kirchhoff assumption.

References and links

[1]T. J. Cox and Y. W. Lam, "Prediction and evaluation of the scattering from quadratic residue diffusers," J. Acoust. Soc. Am. **95**, 297–305 (1994).
[2]I. Tolstoy and C. S. Clay, *Ocean Acoustics* (McGraw-Hill, New York, 1966).
[3]C. S. Clay and J. K. Horne, "Acoustic models of fish: The Atlantic cod (*Gadus morhua*)," J. Acoust. Soc. Am. **96**, 1661–1668 (1994).
[4]E. M. Kennaugh and R. L. Cosgriff, "The use of impulse response in electromagnetic scattering problems," in IRE International Convention Record (1958), pp. 72–77.
[5]S. W. Lee, S. K. Jeng, C. L. Yu, C. S. Liang, and R. A. Shepherd, "Physical optics impulse response from faceted targets," in Antennas and Propagation Society International Symposium (1992), pp. 1456–1459.
[6]J. A. Fawcett, "Modeling of high-frequency scattering from objects using a hybrid Kirchhoff/diffraction approach," J. Acoust. Soc. Am. **109**, 1312–1319 (2001).
[7]M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, Cambridge, 1999).
[8]W. B. Gordon, "Far-field approximations to the Kirchhoff-Helmholtz representations of scattered fields," IEEE Trans. Antennas Propag. **23**, 590–592 (1975).
[9]L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).
[10]A. Papoulis, *The Fourier Integral and Its Applications* (McGraw-Hill, New York, 1962).

Scaled model experiment of long-range across-slope pulse propagation in a penetrable wedge

Alexios Korakas and Frédéric Sturm
LMFA UMR CNRS 5509, Ecole Centrale de Lyon, 36 avenue Guy de Collongue, 69134 Ecully Cedex, France
alexis.korakas@ec-lyon.fr, frederic.sturm@ec-lyon.fr

Jean-Pierre Sessarego and Didier Ferrand
LMA UPR CNRS 7051, 31 chemin Joseph Aiguier, 13402 Marseille, France
sessarego@lma.cnrs-mrs.fr, ferrand@lma.cnrs-mrs.fr

Abstract: In this paper, laboratory scale measurements of long-range across-slope propagation of broadband pulses in a shallow-water wedge-shaped environment with a sandy bottom are reported. The scaled model was designed to study the three-dimensional (3D) acoustic field in the presence of only a few propagating modes. The recorded time series exhibit prominent 3D effects such as mode shadow zones and multiple mode arrivals. Inspection of the spectral content of the time signals gives evidence of intra-mode interference and frequency dependence of the mode cut-off range in the across-slope direction.

PACS numbers: 43.30.Zk, 43.30.Bp, 43.30.Gv [JL]
Date Received: December 2, 2008 **Date Accepted:** April 22, 2009

1. Introduction

It is well known that in realistic three-dimensional (3D) oceanic environments, the acoustic propagation may be affected by the horizontal refraction of the sound energy. The effects of horizontal refraction, commonly referred to as 3D effects, require fully 3D modeling to be accounted for. During the past decades, the 3D wedge-shaped oceanic waveguide has received considerable attention as it approximates realistic oceanic environments such as the continental slope. In this specific environment, the 3D effects are perceived in the across-slope direction rather than in the up-/down-slope direction. Most of the modeling efforts to identify and quantify the 3D effects for the wedge problem focused on single-frequency considerations (see Ref. 1 for a detailed bibliography). Although computationally more intensive, the broadband approach provides a more complete and straightforward picture for the analysis of 3D effects. The broadband modeling of the 3D wedge was implemented using either a ray method[2] or a parabolic equation (PE) approach,[3] both considering a penetrable bottom. On the other hand, a number of laboratory scale experiments were conducted to investigate acoustic propagation in this particular environment and compare theoretical predictions to experimental data. Part of these experimental studies intended to examine up-slope[4] or down-slope[5] propagation alone, while others also investigated the 3D aspect of across-slope propagation.[6,7] More specifically, Ref. 6 reported results of continuous wave (cw) and pulse propagation over a perfectly reflecting bottom, and Ref. 7 investigated cw propagation considering a penetrable bottom. Both concluded to good agreement with single-frequency numerical predictions.

In this paper, we present laboratory scale measurements of long-range across-slope propagation of broadband pulses in a shallow-water wedge-shaped waveguide with a penetrable bottom. This work is part of a research program for the investigation of long-range propagation of acoustic waves in well-defined oceanic waveguides. The campaign reported here was preceded by a calibration phase that was achieved in a Pekeris-like configuration.[8] The bathymetry was subsequently modified to simulate a wedge-shaped oceanic waveguide. Preliminary tests of across-slope pulse propagation exhibited evident 3D effects and comparisons of the experimental results with numerical predictions by a fully 3D PE based code turned out to be very

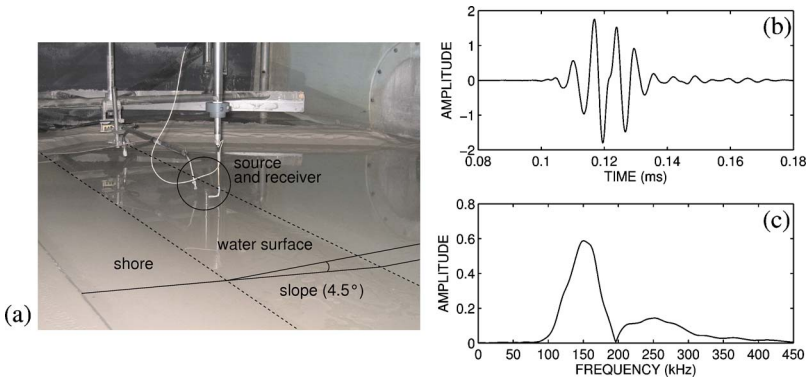


Fig. 1. (Color online) (a) Model experiment of the wedge-shaped oceanic waveguide (with a slope of approximately 4.5°), (b) source signal recorded in free-field, and (c) its frequency spectrum.

encouraging.[9] In this work, additional series of experimental measurements in the wedge-like environment are reported. The signals were recorded on a fine spatial grid within a vertical plane along the across-slope direction. To facilitate the analysis of the 3D effects, the experiment was designed to keep the number of propagating modes in the waveguide relatively low. The recorded time series exhibit strong 3D effects that are identical to those described in the literature.[1–3,6,7,10,11] The frequency dependence of the mode cut-off range is illustrated by inspecting the spectral content of the time signals recorded at several ranges. The data sets obtained during this measurement campaign are intended for 3D model-comparisons. In Sec. 2, a short description of the experimental set-up is given. Then, the experimental results are interpreted according to theoretical predictions. Results and future directions are discussed in Sec. 4.

2. Experimental set-up

The measurement campaign was conducted at the indoor tank facilities of the LMA-CNRS laboratory in Marseille (France). The shallow-water tank is 10-m-long and 3-m-wide, thus allowing for long-range propagation measurements. It consists of a thin layer of water over a thick layer of calibrated sand simulating a bottom half-space. The grain size of the sand is considerably smaller than a wavelength at the operational frequencies. The objective was to set-up a wedge-like configuration and measure the across-slope propagation up to long ranges. For this reason the sandy bottom was tilted with the wedge apex aligned along the longer side of the tank, and was made as flat as possible, see Fig. 1(a). The slope angle was approximately 4.5°. The source and receiver were cylindrical piezoelectric transducers both with diameters of approximately 6.0 mm. They can be seen in Fig. 1(a). The source could be positioned at any depth and the receiver was allowed to move in the three directions. For a detailed description of the experimental facilities and procedures, we refer to Refs. 8 and 9.

3. Experimental results

Prior to the measurements in the shallow-water tank, the source signal was analyzed in a deep-water tank. The signal recorded at a distance of 68 mm from the source appeared well separated in time from its surface echo. It is shown in Fig. 1(b). It is a 5-cycle Gaussian pulse of 40-μs duration with a weak tail of about the same duration most likely due to the mechanical response of the transducer. Its frequency spectrum presents a main lobe, carrying most of the acoustical energy, centered at approximately 150 kHz with a 100-kHz bandwidth, and a secondary lobe with a maximum at 250 kHz. The water depth at the source was 48 mm (± 1 mm) and the water sound speed was 1488.9 m/s (± 0.3 m/s). The bottom compressional wave speed and attenuation were 1700 m/s (± 50 m/s) and 0.5 dB/wavelength (± 0.1 dB/wavelength) respectively, and the bottom density was 1.99 g/cm^3 (± 0.01 g/cm^3). A preliminary simulation with a two-dimensional normal mode code showed the existence of four trapped modes at 150 kHz corre-

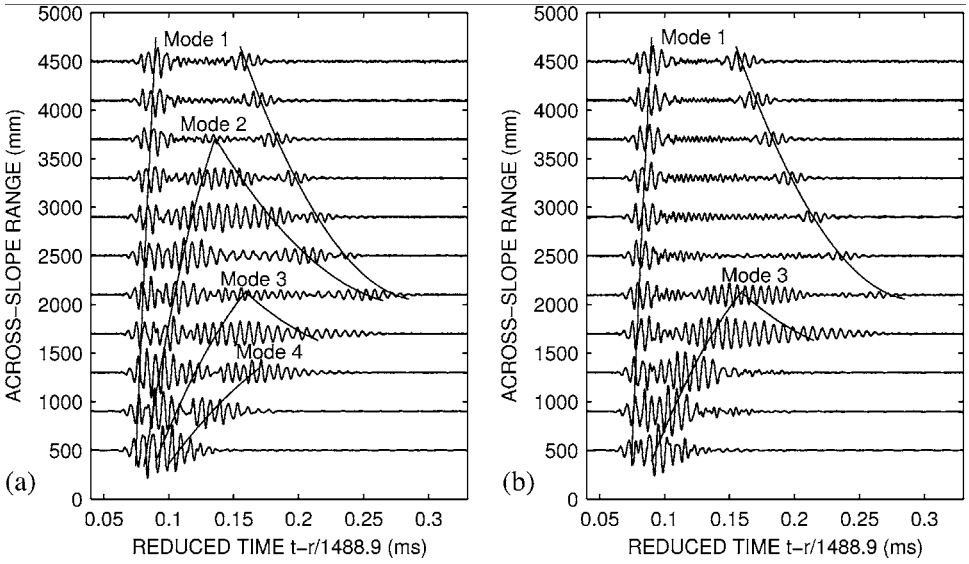


Fig. 2. Stacked time series vs source/receiver range along the across-slope direction. The source depth is 10 mm and the receiver depth is (a) 20 mm where no modal nulls occur and (b) 26 mm where modes 2 and 4 almost vanish. The lines superimposed on both panels indicate the evolution of the mode arrivals with range.

sponding to the center frequency of the main lobe. During the measurements the source was fixed at a depth of 10 mm (±0.5 mm). The received signal was recorded along the across-slope direction at several source/receiver distances from 100 to 5000 mm (±2 mm) with increments of 100 mm, and, at each range step, it was recorded at depths between 4 and 45 mm (±0.5 mm) with a depth increment of 1 mm. The results are presented in Figs. 2, 3(a), and 3(b).

Figure 2 shows two range stacks of the time signals recorded in the across-slope direction at receiver depths of 20 mm [Fig. 2(a)] and 26 mm [Fig. 2(b)]. Note that the experimental data were scaled appropriately to compensate for cylindrical spreading. The curved lines superimposed on the time series were drawn approximately to indicate the modal separation of the initial pulse with range. Four distinct modes are overall identified by simple inspection of the nulls in the depth stacks at each specific range (except at short ranges). For instance, the upper panels of Fig. 3 show the depth stacks at two distinct ranges: one at 2100 mm with modes 1–3 identified [Fig. 3(a)], and one at 2900 mm with only modes 1 and 2 identified [Fig. 3(b)]. Beyond the range of 1000 mm (i.e., ≈100 wavelengths) the time series in Fig. 2 exhibit evident 3D effects. Each mode in Fig. 2(a), with the exception of mode 4, presents two distinguishable time arrivals at some range. The relative time delay between these two arrivals decreases with increasing range. At a certain range, the two arrivals merge together and form what appears to be a more dispersed modal wave packet. Farther in range across-slope, the "merged" modal wave packet shortens and weakens, before entering into a shadow zone. Note that this shadow zone occurs at shorter ranges for higher modes. As can be seen in Fig. 2(b), modes 2 and 4 almost vanish at a receiver depth of 26 mm, and modes 1 and 3 are now separate even at short ranges. Note lastly that the noise-like signals observed between the two arrivals of mode 1 beyond the range of 4000 mm in Fig. 2(a), mainly originate from the secondary lobe of the spectrum displayed in Fig. 1(c). However, their low signal-to-noise ratio did not permit us to perform a more detailed analysis.

Mode shadow zones and multiple mode arrivals over a sloping bottom can be straightforwardly explained by means of ray/mode analogies.[10,11] Let us first consider the problem at a fixed frequency. A given mode propagates as a ray along hyperbolic paths in the horizontal plane, being gradually refracted toward regions of deeper water. As shown in Fig. 3(c), a modal-ray launched obliquely toward the wedge apex travels up-slope, and then turns back down-slope

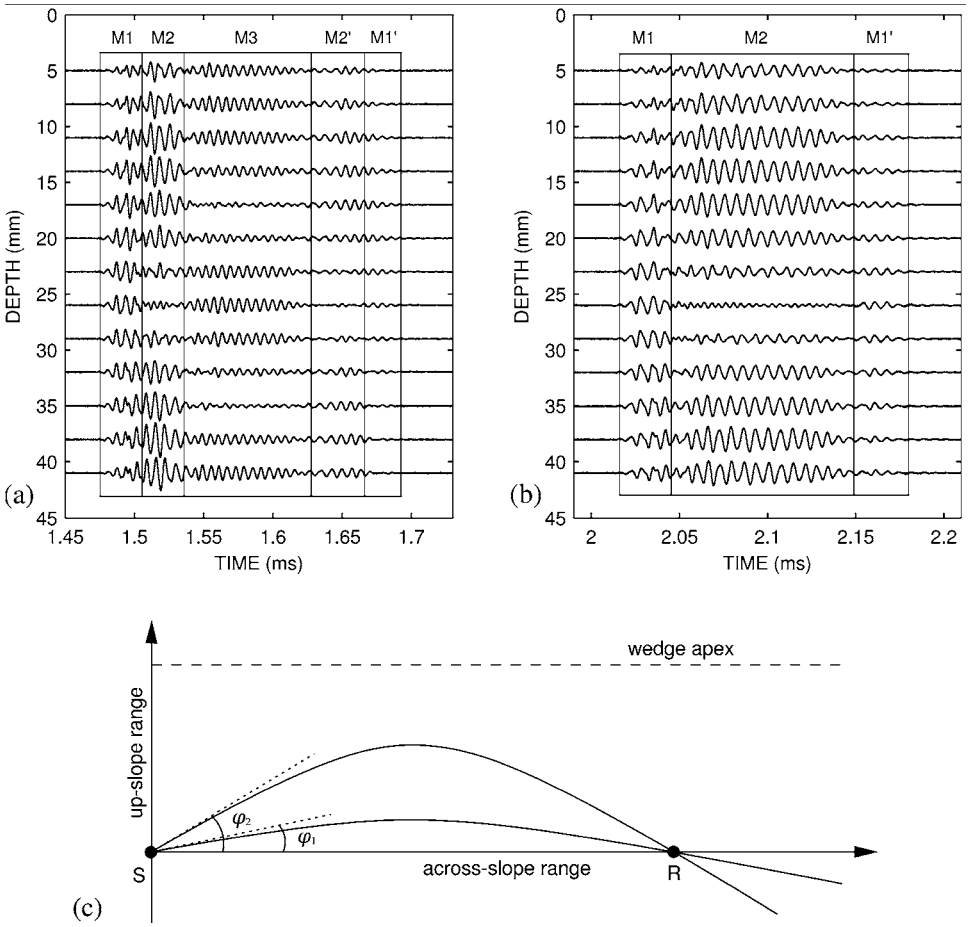


Fig. 3. Stacked time series vs receiver depth: (a) the receiver is at an across-slope range of 2100 mm. From left to right in (a), we identify the first arrival of modes 1 (M1) and 2 (M2), two merged arrivals of mode 3 (M3), a second arrival of mode 2 (M2′), and a very weak second arrival of mode 1 (M1′). (b) The receiver is at an across-slope range of 2900 mm. From left to right in (b), we identify a first arrival of mode 1 (M1), two merged arrivals of mode 2 (M2), and a now more pronounced second arrival of mode 1 (M1′). (c) Horizontal projection of modal-ray paths in the wedge environment. The receiver R positioned across-slope may see two distinct arrivals of the same mode, originating from modal-rays launched from the source S at different angles $\varphi_2 > \varphi_1$.

to intersect the across-slope direction at some range. As the horizontal launch angle, φ, with respect to the across-slope direction increases, the curvature of the modal-ray path increases, and the intersection with the across-slope occurs at shorter ranges. As a result, a receiver R positioned across-slope may see two time arrivals of the same mode: a first mode arrival launched at a low angle φ_1 and a second mode arrival launched at a higher angle φ_2. The modal-ray corresponding to the second arrival passes through regions of shallower depths leaking more energy into the bottom. Hence, the second arrival is weaker than the first arrival. Beyond a critical launch angle the modal-ray passes through its mode cut-off depth, the respective mode being thus transmitted into the bottom.[11] This explains why multiple arrivals of modes 1–3 are not observed at short ranges in Fig. 2(a), while mode 4 does not exhibit any second arrival. On the other hand, the modal-ray is subjected to less bottom loss when launched at a lower angle. This, in turn, explains the amplitude increase in the second arrival of mode 1 with increasing range, in Fig. 2. Continuing this analysis, a high-angle mode arrival has traveled a longer distance than a low-angle mode arrival. As we move out in range across-slope, the difference in the

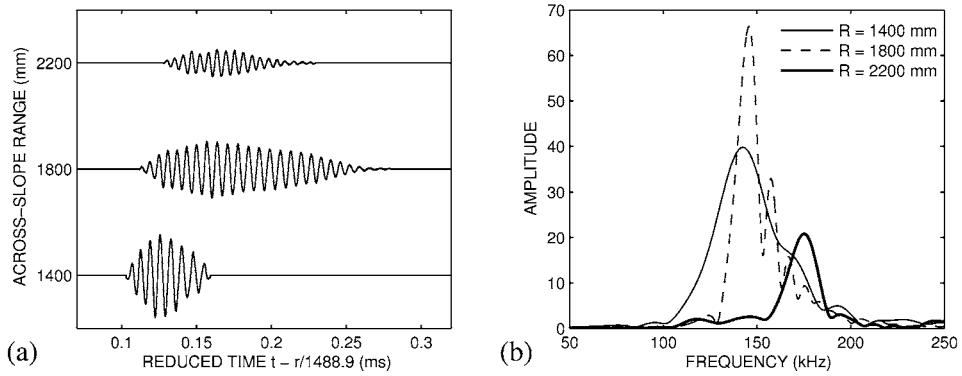


Fig. 4. (a) Mode-3 wave packet at three different ranges: at the range of 1400 mm, one arrival occurs, while two merged arrivals are observed at 1800 and 2200 mm. (b) Spectra of mode 3 at ranges of 1400, 1800, and 2200 mm, giving evidence of the range dependence of the mode cut-on frequency.

distance traveled by each of the two arrivals decreases. Accordingly, the relative time delay between these two arrivals decreases. Furthermore, as mode energy is continuously refracted down-slope, each mode reaches eventually a cut-off range beyond which its shadow zone extends. Since higher-order modes have larger incident angles to the bottom, they are refracted more abruptly. It turns out that, with increasing mode number, mode cut-off occurs at a shorter range. The experimental results presented in Fig. 2 are in complete qualitative agreement with these observations.

Let us now address the frequency dependence of the sound field. Several publications showed that the cut-off range of a given mode is shifted out in range with increasing frequency.[2,6,7,10,11] We thus expect the lower frequency part of the energy of each mode to be progressively removed due to down-slope refraction as we move out across-slope. As a consequence, the extinction of a given modal wave packet is expected to take place in an extended region along the across-slope direction. Figure 4 shows the wave packets associated to mode 3 [Fig. 4(a)] and their frequency spectra [Fig. 4(b)] at three distinct ranges: 1400, 1800, and 2200 mm. Note that mode 3 was extracted from the results of Fig. 2(b) and weighted with a Hanning window to smooth the edges. The frequency spectra were obtained as Fourier transforms of the windowed signals. As observed in Fig. 4(a), at the range of 1400 mm one single arrival of mode 3 occurs, whereas two merged arrivals are observed at the subsequent ranges (1800 and 2200 mm). In Fig. 4(b), the range dependence of the cut-on frequency of mode 3 is evident. More precisely, at the range of 1400 mm, no cut-off has taken place yet, and the cut-on frequency of the mode is approximately 100 kHz. Then, at the range of 1800 mm its cut-on frequency has moved to approximately 130 kHz, leaving the part of the spectrum up to that frequency in the shadow zone. Finally, at the range of 2200 mm, the cut-on frequency is approximately 160 kHz. Moreover, by comparing the spectra at 1400 and 1800 mm in Fig. 4(b), the peak amplitude is seen to almost double, giving evidence of an additional arrival of mode 3 at 1800 mm. Note also the interference patterns present in the spectrum of mode 3 at 1800 mm [dashed curve in Fig. 4(b)]. They are attributed to an effect known in the literature as intra-mode interference,[11] i.e., the mutual interference between arrivals of the same mode occurring at different times. These interference patterns depend on the relative time delay of the two arrivals. At the range of 2200 mm, these arrivals are almost simultaneous and intra-mode interference is weakly observed [bold curve in Fig. 4(b)]. Similar observations hold for other modes (results not shown here).

4. Conclusion and discussion

Results of scaled laboratory experiments of long-range across-slope propagation of broadband pulses in a 3D penetrable wedge were reported. The number of modes contributing to the acous-

tic field was intentionally kept relatively low. Prominent 3D effects such as modal shadow zones and multiple mode arrivals were observed experimentally in agreement with theoretical predictions. Furthermore, intra-mode interference and the frequency dependence of mode cut-off range (or, equivalently, the range dependence of mode cut-on frequency) were put in evidence by examining the spectral content of mode 3 along the across-slope direction.

To conclude, we note that, in contrast with at-sea experiments, high quality data can be collected in laboratory conditions that are suitable for comparisons with numerical propagation models. In this perspective, the experiment was designed to finely sample the sound field in the across-slope direction in both range and depth. The recorded time series can now be appropriately transformed to provide frequency-domain data at several frequencies, e.g., to obtain transmission loss vs across-slope range or depth curves. The data sets obtained during the measurement campaign turn out to be promising for future use as a real-data benchmark for 3D model-comparison. Current work focuses onto detailed comparisons of the experimental data with a fully 3D PE code.

References and links

[1]A. Tolstoy, "3-D propagation issues and models," J. Comput. Acoust. **4**, 243–271 (1996).
[2]E. K. Westwood, "Broadband modeling of the three-dimensional penetrable wedge," J. Acoust. Soc. Am. **92**, 2212–2222 (1992).
[3]F. Sturm, "Numerical study of broadband sound pulse propagation in three-dimensional oceanic waveguides," J. Acoust. Soc. Am. **117**, 1058–1079 (2005).
[4]H. Hobaek and E. K. Westwood, "Measurements of upslope wave-front curvature in a sand-bottom wedge," J. Acoust. Soc. Am. **84**, 1787–1790 (1988).
[5]C. T. Tindle, H. Hobaek, and T. G. Muir, "Normal mode filtering for downslope propagation in a shallow water wedge," J. Acoust. Soc. Am. **81**, 287–294 (1987).
[6]S. A. L. Glegg and J. R. Yoon, "Experimental measurements of three-dimensional propagation in a wedge-shaped ocean with pressure-release boundary conditions," J. Acoust. Soc. Am. **87**, 101–105 (1990).
[7]S. A. L. Glegg, G. B. Deane, and I. G. House, "Comparison between theory and model scale measurements of three-dimensional sound propagation in a shear supporting penetrable wedge," J. Acoust. Soc. Am. **94**, 2334–2342 (1993).
[8]P. Papadakis, M. Taroudakis, F. Sturm, P. Sanchez, and J.-P. Sessarego, "Scaled laboratory experiments of shallow water acoustic propagation: Calibration phase," Acta. Acust. Acust. **94**, 676–684 (2008).
[9]F. Sturm, J.-P. Sessarego, and D. Ferrand, "Laboratory scale measurements of across-slope sound propagation over a wedge-shaped bottom," in Proceedings of the Second International Conference & Exhibition on Underwater Acoustic Measurements, Greece, 2007, pp. 1151–1156.
[10]C. H. Harrison, "Acoustic shadow zones in the horizontal plane," J. Acoust. Soc. Am. **65**, 56–61 (1979).
[11]M. J. Buckingham, "Theory of three-dimensional acoustic propagation in a wedgelike ocean with a penetrable bottom," J. Acoust. Soc. Am. **82**, 198–210 (1987).

High frequency measurements of sound speed and attenuation in water-saturated glass-beads of varying size

Keunhwa Lee and Eungkyu Park
Department of Ocean Engineering, Seoul National University, Seoul 151-744, Korea
nasalkh2@snu.ac.kr, eungga37@snu.ac.kr

Woojae Seong
Department of Ocean Engineering, Seoul National University, Seoul 151-744, Korea and Research Institute of Marine Systems Engineering, Seoul National University, Seoul 151-744, Korea
wseong@snu.ac.kr

Abstract: Acoustic measurements of *p*-wave speed and attenuation were made for water-saturated granular medium, consisting of six kinds of glass-beads with mean grain size ranging from 90 to 875 μm, at frequency range between 400 kHz and 1.1 MHz. Sound speed and attenuation were obtained using the inter-receiver broadband estimation technique. The measured data exhibit various frequency dependencies for the different mean grain sizes, consistent with earlier measurements from other researches. These results reveal that the trend of dispersion relation for the sound speed and attenuation, in the high frequency region, is strongly dependent on the range of Rayleigh parameter *kd*.

PACS numbers: 43.30.Ma, 43.30.Pc [GD]
Date Received: April 9, 2009 **Date Accepted:** May 14, 2009

1. Introduction

Many acoustic measurements[1–3] have been performed in water-saturated granular medium in order to observe the dispersion characteristic and to confirm the prediction model. However, few studies[4,5] have been made around the frequency range of a few hundred kilohertz and/or with a wide selection of granular media categorized by its mean grain size.

Acoustic measurements were performed in the frequency range between 400 kHz and 1.1 MHz for six sizes of water-saturated glass-beads, corresponding in granular size ranging from fine sand to coarse sand. Although a similar experiment by Salin and Schön (SS) (Ref. 5) is found in literature, their objective was to confirm the validity of Biot theory using their measurements. In this paper, the authors focus in analyzing the dispersion relation of the measured data as a function of frequency and mean grain size, and comparing with earlier measurements obtained, respectively, by Nolle, Hoyer, Mifsud, Runyan, and Ward (NHMRW),[4] SS,[5] Schwartz and Plona (SP),[6] and Lee, Humphrey, Kim, and Yoon (LHKY).[7]

Section 2 describes the experimental setup including the sediment preparation along with measurement techniques. In Sec. 3, the measured results are presented and compared with previous measurements obtained by other researchers. Discussion of some disparity between measurements is also given. Finally, Sec. 4 contains the conclusion of the paper.

2. Experiments

2.1 Experimental setup and procedure

Acoustic measurements were performed in a water-saturated granular medium for the purpose of obtaining the *p*-wave speed and attenuation. Glass-beads of six different sizes were used as the granular media. All of them are soda lime silica glass-beads manufactured by Sigmund-Lindner, Germany. The mean grain sizes are 90 μm (S1), 150 μm (S2), 375 μm (S3),

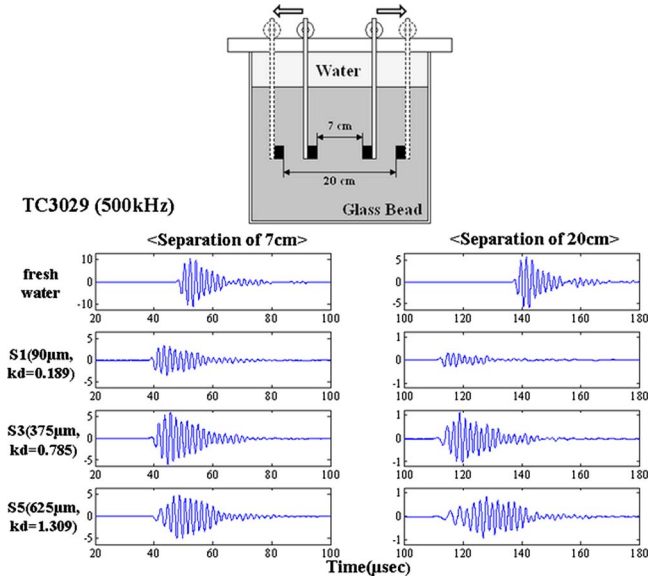


Fig. 1. (Color online) Experimental setup and examples of received signal, showing the typical pulse elongation; in water, S1 sediment, S3 sediment, and S5 sediment.

500 μm (S4), 625 μm (S5), and 875 μm (S6). Before performing the acoustic measurements, the sediment was strained through a coarse sieve to remove any pollutants and then washed in fresh water several times. According to the manufacturer's specification, the roundness of the glass-beads is better than 90% and the density is 2550 kg/m^3. The porosity measurements were performed for S2, S3, and S4, and they all proved to be 0.38. Although no porosity measurements were made for the remaining samples, the authors presume that the porosity of glass-beads is also 0.38.

The sediment was first boiled in order to remove trapped air and then replaced in a small cylindrical vessel (50 cm high and 40 cm diameter) for acoustic measurements. The thickness of the sediment is approximately 42 cm with a water layer of 8 cm thickness above it. The cooling system was run and the temperature of the sediment was kept between 17 and 18 °C. Prior to measurements, the mixture was stirred by a stainless bar and the source and receiver were buried at a depth of 22 cm within the mixture. Then, the mixture was packed by internal vibrator making the glass-bead structure dense. For the fine samples (S1 and S2), in particular, the observed change in the sediment volume was noticeable before and after packing.

Two pairs of transducers [TC3029 (500 kHz) and TC3027 (1 MHz)] made by Reson were used to cover a broad frequency range. In order to employ the phase-delay timing method,[8] signal measurements were obtained at two source-receiver separations of 7 and 20 cm for each transducer pair (Fig. 1). For calibration reason, same measurements were made in water without the sediment. The entire procedure was repeated twice for each of the six sediments.

The source pulse was chosen as a one-cycle sine pulse generated by a HP function generator, with center frequencies at 450, 500, and 550 kHz for TC3029 and at 900 kHz, 1 MHz, and 1.1 MHz for TC3027. The received pulse was displayed on an oscilloscope and saved through a general purpose interface bus (GPIB) port to a PC. For each event, the experiment was repeated five times and the average received pulse was used for data processing.

2.2 Sound speed and attenuation measurement

From the phase-delay timing method[8] the wrapped phase delay $\beta(\omega)$ is given by $\tan^{-1}[\mathrm{Im}(P_{g\_20\ \mathrm{cm}}/P_{g\_7\ \mathrm{cm}})/\mathrm{Re}(P_{g\_20\ \mathrm{cm}}/P_{g\_7\ \mathrm{cm}})]$, where $P_{g\_7\ \mathrm{cm}}(\omega)$ and $P_{g\_20\ \mathrm{cm}}(\omega)$ are the Fourier transforms of near/far (7 cm/20 cm) measurements and ω is the angular frequency.

Since the inverse tangent function yields principal values between $-\pi$ and π, the phase unwrapping has to be performed by adding $2\pi n$ to the wrapped phase delay, where n is an integer increasing by 1 at frequencies where the phase discontinuity appears.

Then, the p-wave sound speed is obtained from the unwrapped phase delay as follows:

$$c_p(\omega) = \frac{\omega d_w}{2\pi n + \beta(\omega)}, \tag{1}$$

where d_w is the separation distance obtained from the water column acoustic calibration.

The p-wave attenuation was determined by the ratio of spectral amplitudes at two receivers in the sediment and in the water, as presented in Ref. 8. The attenuation coefficient (dB/m) is obtained as follows:

$$\alpha_p(\omega) = \frac{1}{2d_w}\ln\left(\frac{A_w(\omega)}{A_g(\omega)}\right), \tag{2}$$

where $A_w(\omega)=|P_{w\_20\ \mathrm{cm}}(\omega)/P_{w\_7\ \mathrm{cm}}(\omega)|^2$ with two spectral amplitudes $P_{w\_20\ \mathrm{cm}}(\omega)$ and $P_{w\_7\ \mathrm{cm}}(\omega)$ in the water and $A_g(\omega)=|P_{g\_20\ \mathrm{cm}}(\omega)/P_{g\_7\ \mathrm{cm}}(\omega)|^2$ with two spectral amplitudes $P_{g\_20\ \mathrm{cm}}(\omega)$ and $P_{g\_7\ \mathrm{cm}}(\omega)$ in the sediment.

The usable frequency range was determined by the characteristics of transducers used, the source spectrum shape and the signal to noise ratio (SNR) of the received signal. According to the manufacturer's specification an effective frequency band for TC3029 is from 370 to 590 kHz while that of TC3027 is from 570 kHz to 1.1 MHz, having an overlap region from 570 to 590 kHz. Considering the spectrum of the source signal used, the two pairs of transducers employed can cover from 400 kHz to 1.1 MHz. These frequency bandwidths will become narrower for the high attenuating sediments since the SNR becomes lower. During the experiment, a noticeable pulse elongation and distortion was observed for coarse samples (S4, S5, and S6) as shown in Fig. 1. In this case, the received pulse was windowed for a sufficiently long time for all of the received pulse energy to be recorded and thus correctly transformed to the spectral domain.

Following the experimental procedure described in Sec. 2.1, four pairs of sound speed and attenuation can be calculated from the four raw measurements; two pairs of transducer measurements at 7 and 20 cm separations, respectively. Since the authors use three frequencies for a pair of transducers, total of 12 pairs of data are, respectively, obtained for TC3029 and TC3027. The sound speed and attenuation values are obtained by averaging of all collected data.

3. Results and data-data comparisons

3.1 Measurements

Figure 2(a) shows the sound speed measurements between 400 kHz and 1,1 MHz for six water-saturated glass-beads. Because of the large variations of the measurements, the sound speeds of samples S1 (90 μm) and S2 (150 μm) are plotted by shaded bands (maximum deviation of ±22 m/s is observed). Conclusion as to whether the sound speeds of S1 and S2 exhibit any frequency dependency cannot be made; the bend at frequency range between 500 and 700 kHz is where the effective frequency bandwidths of TC3029 and TC3027 overlap. Such variation of the sound speed in a fine water-saturated glass-bead is also reported in an earlier measurement[9] for a water-saturated glass-bead with a mean grain size similar to S2. The authors suppose that the fine sediment could be susceptible to the grain packing. However, coarser samples (S3, S4, S5, and S6) have maximum sound speed uncertainties of ±7 m/s and they are shown by their average values, with agreement at the overlapping frequency band of the two transducers. As shown in Fig. 2(a), the sound speed exhibits the negative dispersion for the frequency. As the grain size increases, the slope of the dispersion curve becomes steeper.

Figure 2(b) shows the attenuation measurements. The attenuation uncertainty is ±5.7 dB/m on average. Although the sound speed measurements of fine samples (S1 and S2) showed

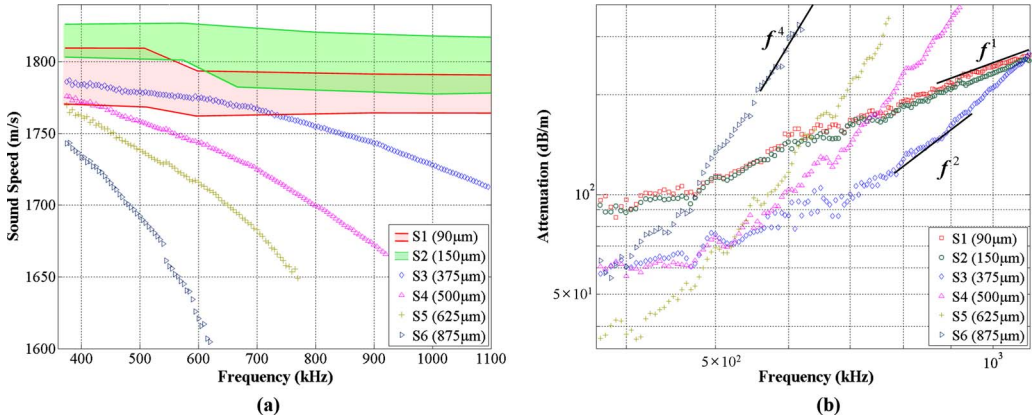


Fig. 2. (Color online) Sound speed and attenuation measurements as a function of frequency. (a) Sound speed measurements and (b) attenuation measurements.

increased uncertainties, their attenuation measurements show variations similar to that of other samples. So, all the attenuation curves of six samples are plotted by the averaged data points. As shown in Fig. 2(b), the attenuation curves of samples S1 and S2 vary as nearly f^1. The attenuation of S3 has f^1 dependence near the frequency of 500 kHz and reaches f^2 dependence starting around 800 kHz. In contrast, samples S4, S5, and S6 start to follow f^4 dependency already at the frequency of 600 kHz. These changes of the frequency dependency are strongly related to the mean grain size. As the mean grain size increases, the frequency dependency varies from f^1 to f^4, as depicted in Fig. 2(b).

3.2 Data-data comparison

The measurements are compared with earlier measurements obtained by four research groups[4–7] to see how they match.

The laboratory measurements by NHMRW (Ref. 4) were made at three frequencies (200 kHz, 500 kHz, and 1 MHz) for four water-saturated refined compacted sands of similar porosity (~0.36) with mean grain sizes of 80, 160, 320, and 640 μm, respectively. According to their paper, the sound speed and attenuation were determined by time peak-peaking method using a narrow band pulse. The measured *p*-wave speed was 1740 m/s [Fig. 3(a)], independent

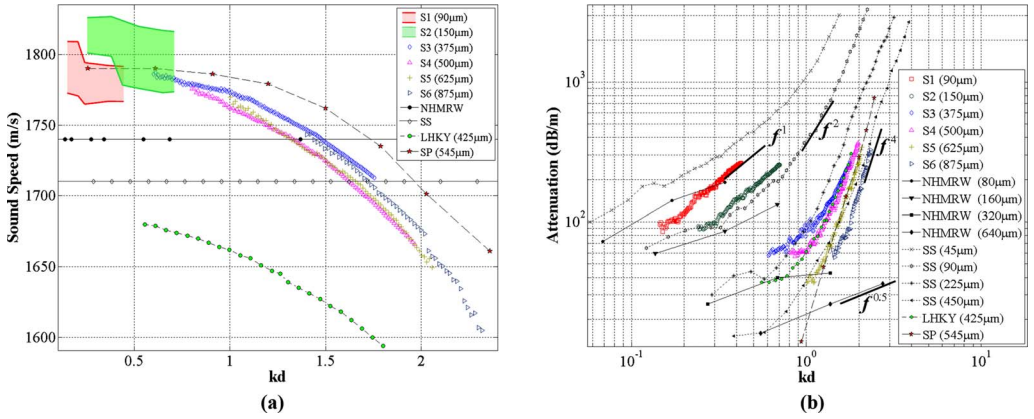


Fig. 3. (Color online) Comparison of (a) sound speed and (b) attenuation measurements as a function of Rayleigh parameter *kd* obtained from present study along with others reported.

of the mean grain size and frequency. The *p*-wave attenuation was found to decrease with increasing mean grain size and follow $f^{0.5}$ dependency as shown in Fig. 3(b). Two reasons for these disparities supposedly are from their lack of measurement data points and from the difference of estimation technique. In the high frequency region where the distortion of received pulse is noticeable, the time domain measurement technique can result in possible estimation errors.

SS (Ref. 5) performed acoustic measurements in the frequency band between 200 kHz and 10 MHz for four water-saturated glass-beads of mean grain sizes 45, 90, 225, and 450 μm, with the grain density of 2900 $\mathrm{kg/m^3}$ and porosity of 0.40 ± 0.02. They used three pairs of transducers with center frequencies of 500 kHz, 2.25 MHz, and 5 MHz, respectively, to cover the broadband frequency range and estimate the sound speeds and attenuations by means of a pulse echo technique. As shown in Fig. 3, the sound speed is measured as 1710 ± 30 m/s, irrespective of the frequency and mean grain size, whereas the measured attenuations show frequency dependency and grain size dependency similar to ours. The difference for the sound speed measurement between our results and those of SS (Ref. 5) cannot be explained since their measurement technique is not described precisely.

SP (Ref. 6) measured the sound speed and attenuation for a water-saturated glass-bead with mean grain size of 545 μm at the frequency range between 300 kHz and 2 MHz. Sound speed and attenuation were determined by a comparison of the phase spectra and magnitude of two received pulses. In their study, the sound speed exhibits a negative dispersion [SP in Fig. 3(a)] and the attenuation follows f^4 dependency [Fig. 3(b)], suggesting Rayleigh scattering. LHKY (Ref. 7) performed a similar experiment in a water-saturated refined sand with mean grain size of 425 μm at the frequency range between 300 kHz and 1 MHz. They also observed the negative sound speed dispersion [LHKY in Fig. 3(a)] and the frequency dependence of the attenuation growing from f^1 to f^4. The estimation procedure of both SP and LHKY is similar to ours; their measurement results also show similar behavior to our results.

In Fig. 3, the sound speed and attenuation measurements are shown as a function of kd, where k is the reference wavenumber (using the water sound speed of 1470 m/s) and d is the mean grain size. Looking at Fig. 3(a), the negative dispersion is seen to start when $kd \approx 0.5$ and tends to get stronger as kd increases, suggesting that the dispersion is associated with Rayleigh scattering. The negative dispersion measurements of SP (Ref. 6) and LHKY (Ref. 7) are in accordance with our findings.

Figure 3(b) shows the attenuation curves plotted as a function of kd. It is interesting to note that the attenuation curves are arranged in order from upper lines of fine sediment to lower lines of coarse sediment when plotted as a function of kd. In case when $kd > 1$, all attenuation curves [excluding the results of NHMRW (Ref. 4)] have a frequency dependence greater than f^1. The frequency dependency of S3 varies from f^1 to $f^{2.5}$ in this range while other samples follow the frequency dependency between $f^{3.5}$ and f^4. In case of $kd < 1$, the frequency dependence becomes much more complicated. The four samples of NHMRW (Ref. 4) follow $f^{0.5}$. SS (45 μm) of SS (Ref. 5) shows the frequency dependency varying from $f^{0.5}$ to f^2. The rest follow nearly f^1 frequency dependence. Such variability of the dependence of attenuation on frequency when $kd < 1$ is supposedly from the competition of scattering mechanism vs other attenuation mechanisms, such as viscous loss, grain shearing loss, etc., strongly related to the distribution of the grain size. In the lower frequency limit region (beginning part) of the attenuation curves for samples SS (450 μm), LKHY, and S4 the frequency dependence lower than $f^{0.5}$ are observed. The authors believe this to be measurement error resulting from low SNR. Overall, in the high kd region, it is clear that the attenuation increases drastically as kd increases and the slope of the attenuation curve is dependent on kd. Additionally, the absolute level of attenuation at the same kd value is always lower for coarser sediment.

4. Conclusion

This works focuses on examining the effect of mean grain size and frequency on the acoustic properties, a granular medium in the high frequency range between 400 kHz and 1.1 MHz. Six sizes of glass-beads were used as the granular medium. The inter-receiver broadband technique

is applied for the estimations of the *p*-wave speed and attenuation. In the high frequency range, measurements of the sound speed and attenuation show strong *kd* dependency. Starting from $kd \approx 0.5$ the negative sound speed dispersion is observed, with negative slope increasing with larger *kd* value. The frequency dependence of attenuation curve is almost linear when $kd < 1.0$; and rises up to f^4 as *kd* becomes larger than 1, suggesting a strong scattering. Present experimental results are compared with earlier laboratory measurements by several authors and they are seen to be in good agreement.

Acknowledgments

The authors thank Mike Buckingham for his valuable insights and comments. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOE-HRD, Basic Research Promotion Fund) (Grant No. KRF-2008-313-D00937).

References and links

[1]K. L. Williams, D. R. Jackson, E. I. Thorsos, D. Tang, and S. G. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng. **27**, 413–428 (2002).

[2]N. P. Chotiros and M. J. Isakson, "A broadband model of sandy ocean sediments: Biot-Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am. **116**, 2011–2022 (2004).

[3]B. T. Hefner and K. L. Williams, "Sound speed and attenuation measurements in unconsolidated glass-bead sediments saturated with viscous pore fluids," J. Acoust. Soc. Am. **120**, 2538–2549 (2006).

[4]A. W. Nolle, W. A. Hover, J. F. Mifsud, W. R. Runyan, and M. B. Ward, "Acoustical properties of water-filled sands," J. Acoust. Soc. Am. **35**, 1394–1408 (1963).

[5]D. Salin and W. Schön, "Acoustics of water saturated packed glass spheres," J. Phys. (France) Lett. **42**, 477–480 (1981).

[6]L. Schwartz and T. J. Plona, "Ultrasonic propagation in close-packed disordered suspensions," J. Appl. Phys. **55**, 3971–3977 (1984).

[7]K. I. Lee, V. F. Humphrey, B. N. Kim, and S. W. Yoon, "Frequency dependencies of phase velocity and attenuation coefficient in a water-saturated sandy sediment from 0.3 to 1.0 MHz," J. Acoust. Soc. Am. **121**, 2553–2558 (2007).

[8]M. J. Buckingham and M. D. Richardson, "On tone-burst measurements of sound speed and attenuation in sandy marine sediments," IEEE J. Ocean. Eng. **27**, 429–453 (2002).

[9]R. D. Costley and A. Bedford, "An experimental study of acoustic waves in saturated glass beads," J. Acoust. Soc. Am. **83**, 2165–2174 (1988).

Natural frequency of a gas bubble in a tube: Experimental and simulation results

Neo W. Jang
Mechanical Engineering, University of Rochester, Rochester, New York 14627

Sheryl M. Gracewski
Mechanical Engineering, Biomedical Engineering, and Rochester Center for Biomedical Ultrasound, University of Rochester, Rochester, New York 14627
grace@me.rochester.edu

Ben Abrahamsen, Travis Buttaccio, and Robert Halm
Mechanical Engineering, University of Rochester, Rochester, New York 14627

Diane Dalecki
Biomedical Engineering and Rochester Center for Biomedical Ultrasound, University of Rochester, Rochester, New York 14627

Abstract: Use of ultrasonically excited microbubbles within blood vessels has been proposed for a variety of clinical applications. In this paper, an axisymmetric coupled boundary element and finite element code and experiments have been used to investigate the effects of a surrounding tube on a bubble's response to acoustic excitation. A balloon model allowed measurement of spherical gas bubble response. Resonance frequencies match one-dimensional cylindrical model predictions for a bubble well within a rigid tube but deviate for a bubble near the tube end. Simulations also predict bubble translation along the tube axis and aspherical oscillations at higher amplitudes.

PACS numbers: 43.80.Gx, 43.35.Ei, 43.35.Wa, 43.25.Yw [CC]
Date Received: March 31, 2009 **Date Accepted:** May 17, 2009

1. Introduction

Accurate determination of the natural frequency of an oscillating gas body, suspended in a liquid medium, surrounded by a compliant vessel has become a topic of importance in diagnostic and therapeutic ultrasound due to increased interest in medical applications of ultrasonically excited bubbles (Dayton and Rychak, 2007; de Jong *et al.*, 2000; Ferrara *et al.*, 2007; Klibanov, 1999; Sassaroli and Hynynen, 2005). One-dimensional (1D) linear models of a cylindrical bubble in a rigid tube have been studied theoretically and experimentally (Geng *et al.*, 1999; Oguz and Prosperetti, 1998; Sassaroli and Hynynen, 2005; Leighton, 1995). Qin and Ferrara (2006) developed a model for a bubble within a compliant vessel and tissue layer using COMSOL MULTIPHYSICS 3.2 that predicts a bubble natural frequency in agreement with the one-dimensional cylindrical bubble model but predicts frequency increases as the tube stiffness decreases (Qin and Ferrara, 2007). The goal of the present study is to further investigate the effect of a surrounding compliant vessel on the natural frequency of a gas body, both experimentally and with a coupled boundary element method and finite element method (BEM-FEM) model.

2. Methods

2.1 Theory and simulation

Simulations of the three phase system, consisting of the gas bubble, surrounding liquid, and solid elastic tube, were done using a coupled BEM-FEM model developed in Miao and

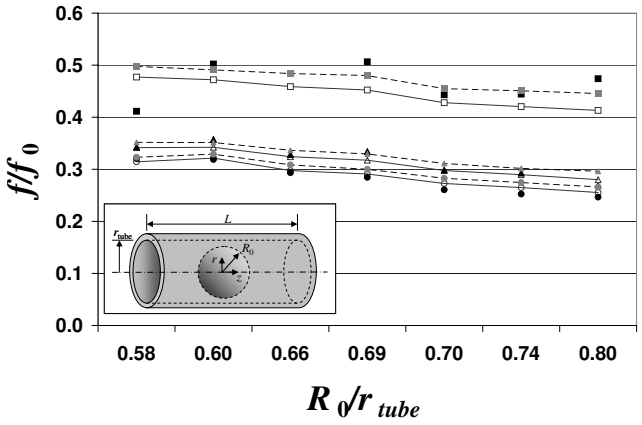


Fig. 1. Bubble normalized natural frequency versus ratio of bubble radius to tube radius. Experimental results (solid markers) are compared to 1D model predictions (open markers) and simulation results (gray markers). Marker shape indicates whether the bubble is at the middle (circles), intermediate (triangles), or end (squares) of the tube. Solid and dashed lines connecting the data points are added so that the 1D model predictions and simulation results, respectively, can be readily identified. Experimental value for $R_0/r_{\text{tube}}=0.66$ at tube end was not obtained. Inset shows a spherical gas bubble inside a circular tube immersed in liquid.

Gracewski, 2008. The model geometry is axisymmetric with an initially spherical gas bubble, located with its center on the axis of symmetry and a circular cylindrical tube with its generator along the axis of symmetry, as shown schematically in the inset of Fig. 1. The acoustic excitation is modeled as a pressure applied at infinity in the surrounding liquid similar to that used in the Rayleigh–Plesset and Gilmore models for spherically symmetric bubble dynamics (Leighton, 1994; Young, 1989).

The gas inside the bubble is assumed to be spatially uniform and to obey the polytropic gas law (Prosperetti, 1991). A pressure jump is applied at the gas-liquid boundary equal to the surface tension times the local curvature (Miao and Gracewski, 2008; Hartland, 2004). The liquid is assumed to be incompressible, irrotational, and inviscid, and therefore modeled with the potential flow equations using the BEM. A linear elastic FEM was employed to solve the dynamic equations in the solid structure domain. Tractions and the normal velocity are assumed continuous across the fluid-solid boundary to couple the finite element and boundary element domains. To obtain the natural frequency of a bubble in a tube using the BEM-FEM model, an initial tensile pulse of half cycle sinusoid is applied at infinity to trigger the bubble's harmonic oscillation. The free vibration period is determined from the equivalent radius (radius of a spherical bubble with equal volume) versus time plot after the excitation pulse is over. Except where indicated, bubbles in the simulations were excited by a half pulse of amplitude 1 kPa and frequency 1 kHz.

The results from the simulations and experiments were compared to the model for the resonance frequency of large bubbles developed in Oguz and Prosperetti, 1998, which is accurate when the initial radius, R_0, of the bubble is greater than ~0.2 times the tube radius, r_{tube}. This model replaces the spherical gas bubble with a cylindrical one of the same volume, occupying the cross-section of the tube, as shown in the inset of Fig. 2. This simplifies the system to one-dimensional motion, where the gas bubble provides the effective stiffness, and the two liquid columns to either side provide the effective inertia, from which the natural frequency of the spring-mass system can readily be determined. The bubble position is specified by L_1 and L_2, the distances from the bubble center to the left and right ends of the tube, respectively, in a tube of length, L. The effective length of the liquid column on either side is obtained by subtracting half of the thickness h from L_1 and L_2, and adding a correction factor $\Delta L=0.62r_{\text{tube}}$ that accounts for the inertia of the liquid outside the tube (Levine and Schwinger, 1948). For this model, the natural frequency of the bubble is determined as

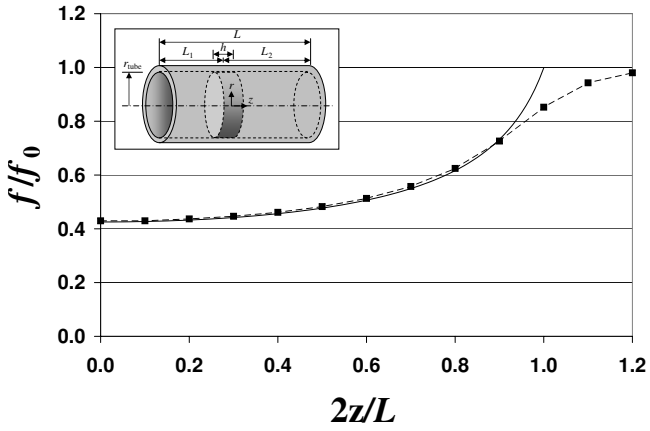


Fig. 2. Bubble normalized natural frequency versus its normalized position. The bubble radius is 1 cm, tube radius is 2 cm, and tube length is 20 cm. Solid line is the 1D model with large bubble assumption, and dotted line is the simulation result with surface tension of 20 N/m in an essentially rigid tube (E=2 GPa). The simulation without surface tension (not shown) predicted frequencies about 0.5% lower than the 1D model, whereas the simulation with surface tension predicted frequencies on average 1.0% higher than the 1D model up to $2z/L \approx 0.8$. Inset shows the spherical gas bubble replaced by a cylindrical bubble of equal volume, with radius equal to the tube radius.

$$\left(\frac{f}{f_0}\right)^2 = \frac{r_{\text{tube}}^2}{4R_0}\left(\frac{1}{L_1 - h/2 + \Delta L} + \frac{1}{L_2 - h/2 + \Delta L}\right), \tag{1}$$

where f_o is the open volume resonance frequency of the bubble.

2.2 Experimental methods

Resonance frequency was measured for a range of bubble sizes, bubble axial positions, tube sizes, and tube materials. In the proposed biomedical applications, gas bubbles with a diameter of a few microns will be injected into the blood stream and pass through capillaries in the human body with diameters several times larger than that of the gas bubbles. The system of interest is scaled up for the laboratory testing, where a balloon of approximately 1.0–2.0 cm in diameter is placed within a tube of 2.5 cm diameter.

Tubes with a range of stiffnesses were used, including Plexiglas, high-density polyethylene (HDPE), and polyvinyl chloride (PVC). The tube material properties and dimensions are given in the Table 1. The cross-section of the PVC tube was slightly oval and the radius of the minor axis was used for the simulation and model calculations. If the material tensile properties could not be readily obtained from the manufacturer, they were determined by following ASTM D638-03 procedure—Standard Test Method for Tensile Properties of Plastics. The samples were prepared in the form of standard dumbbell-shaped type I specimens and pulled with an MTS Alliance RT/50. The elastic modulus values in Table 1 were obtained using a strain rate of 0.001 in./s in the strain range from 0% to 2%. Poisson's ratio for each material was determined

Table 1. Material properties, dimensions of tubes used in the experiments, and corresponding balloons.

Material	Length (cm)	Radius (mm)	Thickness (mm)	Density (g/cm^3)	Elastic modulus (MPa)	R_0/r_{tube}
Plexiglas	19.1	12.5	2.6	1.19	2200	0.60
HDPE	20.5	12.5	3.0	0.91	170	0.58, 0.66, 0.69
PVC	21.0	11.3	2.6	1.21	16	0.70, 0.74, 0.80

to be approximately 0.499 from the elastic modulus values and longitudinal wave speed measurements across the thickness of samples, and this value was used for the simulation.

A spherical air-filled finger-cot (balloon) was used to experimentally model a gas bubble. Each balloon was inflated using a syringe to a gauge pressure of 60 cm of water. The resulting balloon radii ranged between 7.0 and 9.0 mm, corresponding to a membrane tension range of 20–26 N/m. For this range, the nondimensionalized membrane tension $\sigma/(p_0R_0)$ is an order of magnitude smaller than for the surface tension of a 3 μm diameter bubble in water and the presence of the membrane increases the resonance frequency by less than 5% (Young, 2006). Therefore, the balloon was considered to be an adequate representation of a bubble. The measured balloon radii were compared to the values calculated from the balloon's measured open volume resonance frequency without the tube using the Minnaert equation with surface tension σ (Minnaert, 1933),

$$f_0 = \frac{1}{2\pi R_0}\left(\frac{3\Gamma p_0}{\rho}\left(1+\frac{2\sigma}{p_0R_0}\right)-\frac{2\sigma}{\rho R_0}\right)^{1/2}. \quad (2)$$

where R_0 is the equilibrium bubble radius, Γ is the polytropic exponent, p_0 is the ambient liquid pressure, and ρ is the liquid density. The values for these constants used in the calculations are $\Gamma=1.4$, $p_0=101\,230$ Pa, and $\rho=1000$ kg/m^3. Due mainly to folding of the balloon's membrane where it was tied off, there were minor differences ($<6\%$) between the balloon's measured and calculated radii. The last column of Table 1 summarizes the balloon sizes used for each tube. The middle, intermediate, and end positions were approximately 2.0 cm, 6.0 cm, and 9.7 cm, respectively, from tube end.

A stainless steel cylindrical exposure chamber with a shaker (Labworks Inc., model ET-140) on the bottom was used to measure the bubble resonance frequency. The chamber, 25.5 cm in diameter and 35.5 cm in height, was filled with degassed, de-ionized water at room temperature. Each balloon was held in position with strings tied to a vertical flexible tube or the supporting fixture and located using a three-way positioner such that the balloon center was 10 cm below the water surface and centered in the exposure chamber. The shaker was excited by a digital signal generator (Hewlett-Packard HP33120A) and power amplifier (Labworks Inc. PA141). Frequency was swept over a specified range, from 80 to 500 Hz, well below the lowest tank resonance frequency of approximately 950 Hz. A hydrophone (B&K 8103) was used to measure the pressure near a balloon and the resonance frequency of a balloon was identified by a peak in the pressure versus frequency plot.

3. Results and Discussion

Measured and predicted values of the normalized frequency ratio, f/f_0, are presented in Fig. 1 for all of the balloon radii and positions used experimentally. The results in Fig. 1 are ordered by the ratio of balloon radius to the tube radius. A bubble's natural frequency in a stiff tube decreases as the bubble's position is moved from tube end to tube center, as well as when the bubble size increases. There is a good agreement between the results from the experiments, 1D model predictions, and the BEM-FEM simulations, especially when the bubble is located near the tube center.

The experimental variability was higher for balloons positioned near the tube's end than at intermediate positions or in the middle of the tube, due to the higher sensitivity of frequency on axial position. In Fig. 2, the frequency ratio versus bubble position along the tube is plotted to illustrate this higher sensitivity. This plot also shows that the 1D model results are close ($\sim$1% difference) to the BEM-FEM simulation results when the bubble is near the middle of the tube. Near the tube ends, Eq. (1) becomes less accurate because the correction factor ΔL dominates either L_1 or L_2. In comparison, the BEM-FEM simulation predicts the bubble frequency to asymptotically converge to that for a bubble in an open volume, as the bubble moves away from the tube. Simulations predict that the effect of the tube presence is still noticeable ($f/f_0=0.93$) for a distance of one bubble radius away from the tube end.

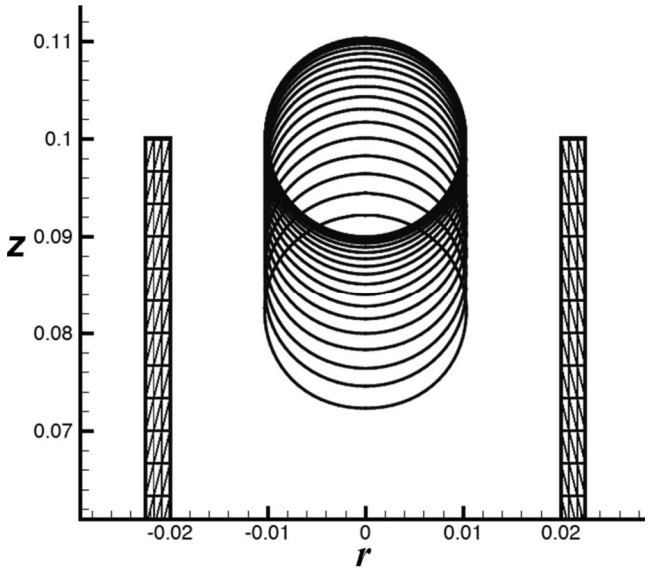


Fig. 3. Bubble position at successive peaks of the oscillation versus time curve corresponding to Mm. 1, showing that the bubble translates into the tube as it oscillates. The bubble radius is 1 cm, tube radius is 2 cm, tube length is 20 cm, and the tube is essentially rigid (E=2 GPa). Excitation is a tensile half pulse with frequency 500 Hz and amplitude 10 kPa.

The animation in Mm. 1 discussed in Fig. 3 shows that a bubble initially positioned at the end of the tube is drawn into the tube as the bubble oscillates. The bubble with 1 cm radius is initially positioned such that one-half is inside the tube with 2 cm radius and 20 cm length. The elastic modulus was chosen to be 2 GPa, so that the tube is essentially rigid. The bubble is excited by a half pulse with frequency 500 Hz and amplitude 10 kPa. The bubble frequency decreases as the bubble moves into the tube, consistent with the resonance frequency trend shown in Fig. 2. Though similar behavior is observed for a bubble initially at an intermediate position along the tube, the magnitude of translational motion is less pronounced, because the translational velocity decreases as the initial bubble position approaches the middle of the tube. Slight asphericity can also be observed in the shape of bubble oscillation, even with this small amplitude.

[Mm. 1. Animation of a 1 cm radius bubble initially at the tube opening, showing the bubble translating into the tube as it oscillates. This is a file of type "avi" (4.11 Mbytes).]

The asphericity of the bubble oscillations increases with oscillation amplitude, as shown in the animation in Mm. 2 discussed in Fig. 4 for a bubble that experiences a 45% change in volume upon expansion. Simulation parameters are the same as in Fig. 3, except that the excitation pulse amplitude is 100 kPa and the bubble is positioned at the middle of the tube, so it does not translate. The aspect ratios (z dimension/r dimension) shown in Fig. 4 are 1.07 and 0.69 for the first expansion and collapse, respectively. The displacements of the bubble top and bottom surfaces (at r=0) along the axis of the tube are 73% larger than the radial displacements of the bubble sides (at z=0). In the animation in Mm. 2, the bubble shape alternates between more aspherical and nearly spherical for each successive oscillation. For example, the asphericity decreases during the second expansion and collapse, with aspect ratios equal to 1.01 and 1.00, respectively, but increases again for the third oscillation. The particular behavior of the bubble oscillations depends on the bubble to tube size ratio, the oscillation amplitude, and the tube stiffness. The asphericity also increases as R_0/r_{tube} increases (results not shown).

[Mm. 2. Animation of a 1 cm radius bubble centered in a tube, showing aspherical bubble oscillations. This is a file of type "avi" (3.5 Mbytes).]

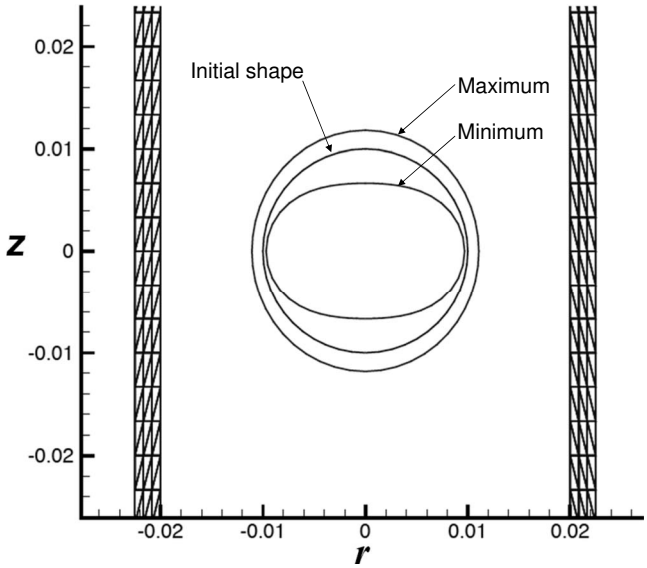


Fig. 4. Bubble shape at time $t=0$ and at its first maximum and minimum volume corresponding to Mm. 2, showing aspherical bubble oscillations. The bubble radius is 1 cm, tube radius is 2 cm, tube length is 20 cm, and the tube is essentially rigid ($E=2$ GPa). Excitation is a tensile half pulse with frequency 500 Hz and amplitude 100 kPa.

4. Summary

In this paper, experiments and simulations were used to investigate the effect of a surrounding tube on the resonance frequency and free vibration response of a bubble. To the authors' knowledge, this is the first published experimental measurement of resonance frequency of a *spherical* gas body in a tube. Simulations were obtained with a coupled BEM-FEM code that was developed specifically to investigate the response of acoustically excited bubbles near deformable structures. Simulated and experimentally measured resonance frequencies decreased as the bubble moved into the tube or increased in size. For a bubble near the center of the tube, the resonance frequency was consistent with a 1D cylindrical bubble model for a large bubble in a rigid tube. For a bubble near the end of the tube, the experimental variation in resonance frequency increased due to inaccuracies in bubble location and the higher sensitivity of frequency on position near the end of the tube. The BEM-FEM simulation results predict that the resonance frequency asymptotically approaches the open volume value as the bubble moves away from the tube end into the open volume. The effect of tube elastic modulus on bubble frequency is the subject of further investigation; however, varying the magnitude of the tube elastic modulus from a few kPa to few GPa in the BEM-FEM simulations had a negligible effect on the predicted bubble frequency (results not shown). Simulation of the free vibration response of the bubble also predicts that a bubble will translate toward the center of the tube due to aspherical oscillations. This is similar to the bubble translation observed in simulations by Ory *et al.* (2000) for the much larger expansion and collapse of a vapor bubble in a narrow tube. The asphericity increases with vibration amplitude, tending to be more elongated in the axial direction on expansion and in the radial direction on collapse.

Acknowledgments

The coupled boundary element and finite element code used in this work was written by Hongyu (Jacky) Miao. The authors gratefully acknowledge Sally Child's assistance with all the experiments. This work was supported by NIH and NSF. Robert Halm was supported by NSF REU.

References and links

Dayton, P. A., and Rychak, J. J. (**2007**). "Molecular ultrasound imaging using microbubble contrast agents," Front. Biosci. **12**, 5124–5142.

de Jong, N., Bouakaz, A., and Frinking, P. (**2000**). "Harmonic imaging for ultrasound contrast agents," Proc.-IEEE Ultrason. Symp. **2**, 1869–1876.

Ferrara, K. W., Pollard, R., and Borden, M. (**2007**). "Ultrasound microbubble contrast agents: Fundamentals and application to gene and drug delivery," Annu. Rev. Biomed. Eng. **9**, 415–447.

Geng, X., Yuan, H., Oguz, H. N., and Prosperetti, A. (**1999**). "The oscillation of gas bubbles in tubes: Experimental results," J. Acoust. Soc. Am. **106**, 674–681.

Hartland, S. (**2004**). *Surface and Interfacial Tension: Measurement, Theory, and Applications* (Dekker, New York).

Klibanov, A. L. (**1999**). "Targeted delivery of gas-filled microspheres, contrast agents for ultrasound imaging," Adv. Drug Delivery Rev. **37**, 139–157.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, London).

Leighton, T. G. (**1995**). "Applications of one-dimensional bubbles to lithotripsy, and to diver response to low frequency sound," Acta Acust. **3**, 517–529.

Levine, H., and Schwinger, J. (**1948**). "On the radiation of sound from an unflanged circular pipe," Phys. Rev. **73**, 383–405.

Miao, H., and Gracewski, S. M. (**2008**). "Coupled FEM and BEM code for simulating acoustically excited bubbles near deformable structures," Comput. Mech. **42**, 95–106.

Minnaert, M. (**1933**). "On musical air-bubbles and sounds of running water," Philos. Mag. **16**, 235–248.

Oguz, H. N., and Prosperetti, A. (**1998**). "The natural frequency of oscillation of gas bubbles in tubes," J. Acoust. Soc. Am. **103**, 3301–3308.

Ory, E., Yuan, H., Prosperetti, A., Popinet, S., and Zaleski, S. (**2000**). "Growth and collapse of a vapor bubble in a narrow tube," Phys. Fluids **12**, 1268–1277.

Prosperetti, A. (**1991**). "The thermal behavior of oscillating gas bubble," J. Fluid Mech. **222**, 587–616.

Qin, S. P., and Ferrara, K. W. (**2006**). "Acoustic response of compliable microvessels containing ultrasound contrast agents," Phys. Med. Biol. **51**, 5065–5088.

Qin, S. P., and Ferrara, K. W. (**2007**). "The natural frequency of nonlinear oscillation of ultrasound contrast agents in microvessels," Ultrasound Med. Biol. **33**, 1140–1148.

Sassaroli, E., and Hynynen, K. (**2005**). "Resonance frequency of microbubble in small blood vessels: A numerical study," Phys. Med. Biol. **50**, 5293–5305.

Young, F. R. (**1989**). *Cavitation* (McGraw-Hill, New York).

Young, J. (**2006**). "The relation between lung damage induced by acoustic excitation and the subharmonic response of bubbles," MS thesis, University of Rochester, Rochester, NY.

Temporal sound field fluctuations in the presence of internal solitary waves in shallow water

Boris G. Katsnelson and Valery Grigorev
Voronezh University, 1 Universitetskaya Square, Voronezh 394006, Russia
katz@phys.vsu.ru, grig@box.vsi.ru

Mohsen Badiey
College of Marine and Earth Studies, University of Delaware, Newark, Delaware 19716
badiey@udel.edu

James F. Lynch
Woods Hole Oceanographic Institution, 98 Water Street, MS No. 12, Woods Hole, Massachusetts 02543
jlynch@whoi.edu

Abstract: Temporal variations of intensity fluctuations are presented from the SWARM95 experiment. It is hypothesized that specific features of these fluctuations can be explained by mode coupling due to the presence of an internal soliton moving approximately along the acoustic track. Estimates are presented in conjunction with theoretical consideration of the shallow water waveguide.

PACS numbers: 43.30.Bp, 43.30.Dr, 43.30.Es, 43.30.Zk [WS]
Date Received: July 17, 2007 **Date Accepted:** May 14, 2009

1. Introduction

Sound intensity fluctuations are observed in the SWARM95 experiment[1] when an internal soliton (IS), or a train of IS, propagates approximately along an acoustic track with wave crest making an angle of approximately 40° (Fig. 1). During a part of this experiment, broadband pulses (in the 30–200 Hz band) were radiated every minute from an airgun and received on two vertical line arrays (called NRL-VLA and WHOI-VLA) at a distance of 14–18 km from the sound source. Figure 1 shows a schematic of the source and receiver arrays. The authors reported analysis of the WHOI-VLA results previously[2] and established that the mechanism dictating the temporal behavior of arriving sound signals is governed by horizontal refraction. Other numerical calculations involved mode coupling.[3] However, the same mechanism is not adequately depicting the temporal variations observed in the NRL-VLA data, for example, Ref. 3. In this paper, the authors present theoretical estimates, supported by experimental data, stating that specific features of mode coupling are responsible for the intensity fluctuations.

The observed temporal variations during a 1 h long pulse transmission (1 pulse/min) on the NRL-VLA have the characteristics of (1) predominant frequency in the fluctuations' spectrum below 20 cph for different sound frequencies [shown in Fig. 2(b)], (2) specific arrival time spreading for different modes on the time-frequency (TF) diagram [shown in Fig. 3(b)], and (3) approximately constant correlation time in a wide range of sound frequency [shown in Fig. 4(b)].

To complement our previous results on the WHOI-VLA having distinctly different of acoustic track geometry in relation to the IS direction,[1,2] and where the temporal variations were due to horizontal refraction, the variations observed on the NRL-VLA can be explained by specific features of the mode coupling.[3] This mechanism can uniquely describe the behavior of all the pulses that have passed through the IS and were received by the NRL-VLA. A pulse radiated from the airgun is a sum of separate normal modes propagating with separate group velocities. After interaction with the IS located at some distance R from the source, this sum changes and in turn the sound field at the receiver changes (in comparison with unperturbed

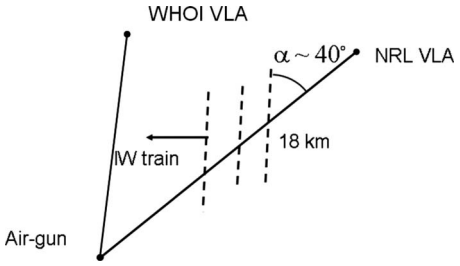


Fig. 1. Experimental schematic of the SWARM95 during Aug. 14, 1995. The directions of acoustic track formed by UD source and NRL and WHOI vertical line arrays are shown with respect to the predominant direction of the internal wave shown by dotted lines; α is the horizontal angle between these directions.

waveguide). For other positions of the moving IS relative to the source/receiver track (or for another geotime), there will be other combinations of modes created, and a different sound field at the receiver is observed. For fixed source-receiver geometry, the authors perceive this variability as temporal fluctuations. Typical frequencies of the fluctuations are about $\sim 1-10$ cph.

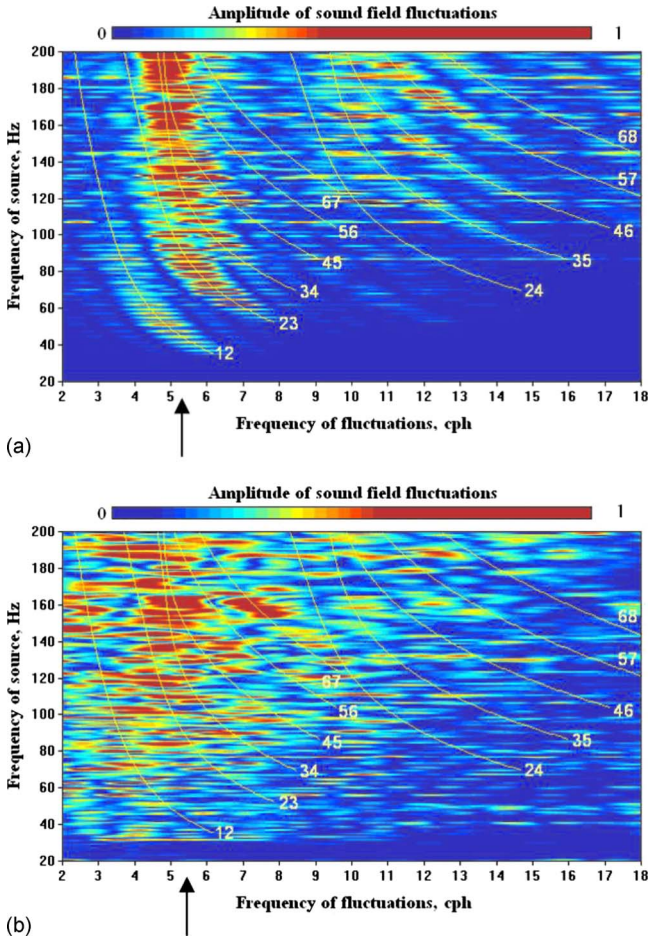


Fig. 2. (Color online) A frequency-frequency (fF) diagram for SWARM95 showing the spectral intensity G of received signal as function of frequency: (a) theory and (b) experiment.

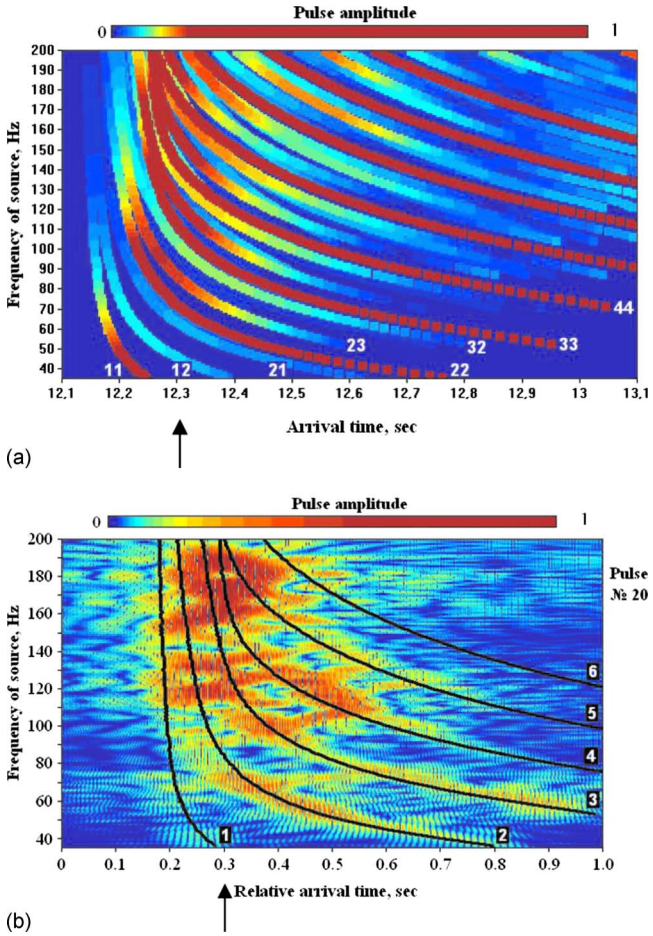


Fig. 3. (Color online) Frequency-time diagram for arrival times. Color scale gives amplitudes of separate modes according to scale above the figures: (a) theory and (b) experiment. Lines 11, 22, etc., in (a) denote positions of modes without coupling (or without a soliton); in (b), these lines are denoted by black curves with numbers 1, 2, 3, etc.

The specific waveguide features essential for understanding the nature and properties of these temporal fluctuations are (1) a narrow thermocline layer (10–35 m deep, within a water column of ~ 90 m total depth), (2) approximately constant soliton shapes and velocities during a few hours ($v_s \sim 0.5-1$ m/s), and (3) length ΔR of the IS less than the acoustic track L ($\Delta R \sim 300-1000$ m $\ll L \sim 15-20$ km). Here, preliminary results in support of the above statement are presented to explain the received temporal variability at the NRL-VLA array.

2. Waveguide model

A point source with spectrum $S(\omega)$ was placed at depth z_s in a shallow water waveguide. The authors assume the unperturbed waveguide has a constant depth H, bottom parameters c_1 and ρ_1, and unperturbed sound speed profile $c(z)$. Bottom attenuation is neglected. At the receiver point (r,z) the Fourier decomposition of the sound field pressure is

$$P(r,z,t) = 2\int_0^\infty P_\omega(r,z)e^{-i\omega t}d\omega. \quad (1)$$

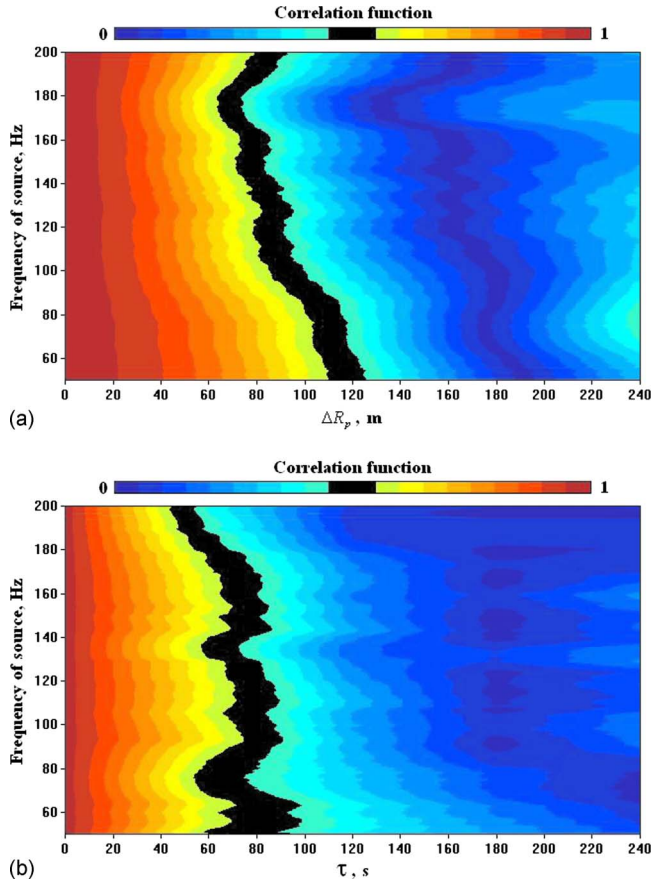


Fig. 4. (Color online) Correlation function for SWARM'95 conditions in color gradation of amplitude (scale is shown); black denotes half level: (a) theory and (b) experiment.

In the absence of perturbations or in the area $r<R$, the amplitude $P_\omega(r,z)$ is the following modal decomposition over waveguide modes ψ_l (where q_l is eigenvalue):

$$P_\omega(r,z) = iS(\omega)\sum_l \frac{\psi_l(z_s)\psi_l(z)}{\sqrt{8\pi i q_l r}}\exp(iq_l r). \tag{2}$$

In the presence of the IS moving with constant velocity v (if there is some angle α between the direction of the acoustic track and the wave front of the IS, the actual velocity of the IS is $v_s=v\sin\alpha$; see Fig. 1), there is a perturbation of the sound field. At a given moment of time T, this perturbation occupies an area at the range $R<r<R+\Delta R$, where $R=vT$, and in this area the authors have a small addition to the sound speed profile, $\delta c(r,z,T)$, which depends on the "slow" time T. This moving sound speed perturbation produces fluctuations of the sound field also depending on T. After acoustic interaction with the soliton, the authors have another modal decomposition for the sound field for $r>R+\Delta R$. The authors describe this decomposition using S-matrix formalism

$$P_\omega(r,z;R) = \sum_m \left[iS(\omega)\sum_l \frac{\psi_l(z_s)\psi_m(z)}{\sqrt{8\pi i q_m r}} S_{ml}(R+\Delta R,\omega)\exp(i\Delta q_{ml}R)\right]\exp(q_m r), \tag{3}$$

where $\Delta q_{lm}=q_l-q_m$, the matrix $\mathbf{S}(r)$ satisfies the equation with "initial" condition

$$\frac{d\mathbf{S}}{dr} = \mathbf{WS}, \quad \mathbf{S}(R) = \mathbf{I}, \tag{4}$$

where $\mathbf{I}$ is the unit matrix, and the coupling coefficients in simple perturbation theory are

$$W_{ml}(r) = i\frac{k^2 \exp[i(q_l - q_m)r]}{\sqrt{q_m q_l}} \int_0^H \frac{\delta c(r,z)}{c} \psi_m(z)\psi_l(z)dz. \tag{5}$$

Due to $T=R/v$, the authors will simply write $P_\omega(T)$ (omitting receiver coordinates). Expression in square brackets in Eq. (3) is effectively the amplitude of the mth mode. The sound field at the receiver equation (3) has several different scales of periodicity as a function of the IS's position R, and thus on the time T. Since the spatial scales of acoustic variability are determined by the parameter $\Lambda_{lm}=2\pi/|\Delta q_{lm}|$, the authors have correspondingly frequencies of temporal variability $\Omega_{lm}=v|\Delta q_{lm}|$. Typically the value of Ω_{lm} is $\sim 10^{-4}-2\times 10^{-3}$ Hz ($\sim 1-15$ cph). Assuming that the strongest coupling interaction is between adjacent modes (close coupling), the authors can write from Eq. (3) a form for the normalized (with respect to the source spectrum) amplitude of mode m created from mode $m+1$: $A_{m,m+1}=|\psi_{m+1}(z_s)\psi_m(z)S_{m,m+1}/\sqrt{8\pi i q_m}|$. This quantity depends on frequency, and for each pair of coupling modes, the coefficients $A_{m,m+1}$ have their own "optimal" frequency, denoted by $\omega_{\text{opt}}^m = \omega_{\text{opt}}^{m,m+1} \sim \omega_{\text{opt}}^{m+1,m}$, where $A_{m,m+1}$ has a maximum. This maximum corresponds to the situation where the two coupling modes have turning point positions within (or close to) the thermocline layer. In this case, pairs of coupling modes have approximately the same scale of interference beating (or mode cycle distance) $D_{\text{opt}} = 2\pi/|\Delta q_{m+1,m}|_{\omega=\omega_{\text{opt}}^m} \approx \text{const}$.

The authors apply this model to examine our hypothesis with the experimental data.

3. Intensity temporal fluctuations

Let us consider the variability of the sound intensity at a fixed frequency (spectral intensity) $I_\omega(T)=|P_\omega(T)|^2/2\rho c$ during an interval ΔT (i.e., $0<T<\Delta T$, where ΔT can be on the order of a few hours, when the waveguide parameters and IS field's characteristics are supposed to be constant). The authors thus consider the spectrum of intensity fluctuations, determined by the equation

$$G(\omega,\Omega) = \int_0^{\Delta T} \delta I_\omega(T) e^{i\Omega T} dT, \tag{6}$$

where

$$\delta I_\omega(T) = I_\omega(T)/\langle I_\omega\rangle - 1, \quad \langle I_\omega\rangle = \frac{1}{\Delta T}\int_0^{\Delta T} I_\omega(T)dT. \tag{7}$$

The function $|G(\omega,\Omega)|$ can be shown in a plot with the vertical axis denoting sound frequency ω and water column fluctuation frequency Ω on the horizontal axis. Values of $|G|$ can be shown in a color scale on this plane. The authors denote this picture as a "frequency-frequency (fF)-diagram."

Figure 2(b) is from the SWARM95 experimental data[1] showing the fluctuation spectrum versus the sound frequency spectrum. In this case ΔT is 1 h (i.e., 19:00–20:00 GMT on August 4, 1995), and the discrete set T_i corresponds to geotimes for 60 broadband pulses, which were emitted every minute ($i=1,2,\ldots,60$). In this diagram there exists a predominating frequency of fluctuations in the range $\sim 5\pm 2$ cph for the entire sound frequency band [see arrow

in Fig. 2(b)]. To explain this feature, the authors use the modal decomposition shown in Eqs. (2) and (3) as follows:

$$G(\omega,\Omega)=\sum_{m,l,n,k} G_{mlkn}(\omega,\Omega), \tag{8}$$

where each member of the sum is

$$G_{mlkn}(\Omega,\omega)=\frac{\Delta T}{r\sqrt{q_m q_n}} S_{ml} S^*_{nk} \psi_l(z_s)\psi^*_k(z_s)\psi_m(z)\psi^*_n(z)\frac{\sin\theta}{\theta}\exp[i(r\Delta q_{mn}+\theta)]. \tag{9}$$

Here, $\theta=\theta_{mlkn}=0.5\Delta T[\Omega-(\Omega_{ml}-\Omega_{nk})]$ and $\Omega_{ml}=v|\Delta q_{lm}|$.

Equation (9), as a function of frequency (ω), has maxima at the points $\Omega-(\Omega_{ml}-\Omega_{nk})=0$.

In the spectrum of intensity oscillations, the authors observe many spectral lines corresponding to the aforementioned differences. The positions of these spectral lines will depend on sound frequency, due to the frequency dependence of the modal eigenvalues. These differences are determined by the dispersion curves $q_l(\omega)$ of the separate modes. Their vertical shape depends on the sound speed profile, the overall depth, and the sound speed in the bottom (i.e., the general waveguide properties). The amplitude (or intensity) of each spectral line also depends on frequency and mode number. In particular, in accord with the fact that the main contribution is provided by adjacent modes (the values $A_{m,m+1}$ are maximal in the sum), the spectral line with maximal intensity occurs at the frequency $\Omega=\Omega_{m,m+1}$ or $\Omega=v|\Delta q_{m,m+1}|$. In Fig. 2(a), the authors show the result of a calculation of the spectrum of intensity fluctuations $G(\omega,\Omega)$ for different sound frequencies, using a color scale (normalized here and in all figures from 0 to 1).

For these calculations, the authors take waveguide parameters corresponding to the experiment in Ref. 1. Specifically, the depth of the source is $\sim$12 m, the water depth at the receiver array is $\sim$85 m, and the velocity of the perturbation along the acoustic track is $v\sim1.1$ m/s. This value of v provides the best correspondence between the theory in Fig. 2(a) and the experimental data in Fig. 2(b). One can see that the maxima in these spectra are found on the curves $v|\Delta q_{m,m+1}|=$const, shown by thin yellow lines in both figures. Every slice of these curves by a horizontal straight line gives the spectrum of the fluctuations for a given sound frequency. The authors see that in these spectra there are many maxima, with the most significant one corresponding approximately to the fluctuation frequency $\Omega_{\text{opt}}=v|\Delta q_{m,m+1}|_{\omega=\omega^m_{\text{opt}}}$ [see arrow in Fig. 2(a)]. However, for different sound frequencies, this predominant frequency of oscillation is provided by different pairs of modes. For example, near $\omega=\omega^1_{\text{opt}}\sim50$ Hz, $\Omega_{\text{opt}}=\Omega_{12}$, whereas near $\omega=\omega^2_{\text{opt}}\sim100$ Hz, $\Omega_{\text{opt}}=\Omega_{23}$, etc.

The value of the dominant fluctuation frequency is $\Omega_{\text{opt}}\sim13\times10^{-4}$ Hz (i.e., $\sim$5–6 cph). Existence of this dominant frequency of intensity fluctuations is the result of the previously mentioned constant length scale for the interference beating of pairs of modes at optimal frequencies. Rough estimates from ray theory for the cycle distance of a ray tangent to the thermocline layer give $D_{\text{opt}}\sim700-800$ m, corresponding to the above frequency, $\sim$5 cph. The authors also note that the estimated velocity of the soliton, $v_s\sim0.7$ m/s, is in accord with observations in this area.[1,3]

4. Arrival time fluctuations

The group velocity of an individual mode is $v^{\text{gr}}_l=(dq_l/d\omega)^{-1}$. If the length of the acoustic track is L, then arrival time of a non-coupling mode is $t_l=L/v^{\text{gr}}_l$. Because group velocities of individual modes depend on frequency, arrival time is also function of frequency, $t_l=t_l(\omega)$. A set of these functions (or curves in the TF-plane) for different mode numbers is classically referred to as a "TF-diagram" and is often used in pulse propagation analysis. If a propagating mode l meets a soliton at distance R, then after their interaction, a newly created set of modes is generated. The "additional" modes with index m propagate after leaving the soliton with their own group velocities. Their arrival time t_{lm} can be estimated as

$$t_{lm} = \frac{R}{v_l^{\mathrm{gr}}} + \frac{L-R}{v_m^{\mathrm{gr}}}. \tag{10}$$

Thus for the TF-diagram the authors have additional curves $t_{lm}(\omega)$. For uniqueness of notation, the authors denote the arrival time of the uncoupled mode as t_{ll}. The frequency and modal dependencies of arrival time can be different, and this is determined by the specific waveguide parameters. For typical conditions (including those of the SWARM95 experiment), for a given frequency $v_1^{\mathrm{gr}} > v_2^{\mathrm{gr}} > \ldots$, and correspondingly $t_{11} < t_{22} < \ldots$. For the newly created modes, the authors have somewhat more complex relationships between the arrival times and the positions of the additional curves on the TF-diagram. A simpler situation results if the authors take into account the interaction between adjacent modes only. In this case $t_{11} < t_{12}$, $t_{21} < t_{22} < \ldots$. The relationship between t_{12} and t_{21} (as well as between t_{lm} and t_{ml}) depends on the position of the soliton (on R) and on the relationship between the group velocities v_1^{gr} and v_2^{gr}.

It is noted that the amplitude of the additional pulse, as per Eq. (3), depends on frequency and distance R. It fluctuates as a function of position of the soliton R or as a function of time T. This situation can also be described via the TF-diagram, if the authors introduce information about the amplitude of a mode at a given frequency using color gradation. These diagrams are shown in Fig. 3. In Fig. 3(a), the results of theoretical calculations for our model of the SWARM95 waveguide are shown, where the authors assumed a definite position of the soliton. Amplitudes of the modes are denoted using color gradation. The authors can see that, in accordance with our theory, the authors have the most significant amplitude of the additional modes near the optimal sound frequencies, as introduced earlier. Due to the existence of an optimal ray cycle distance D_{opt}, the authors can introduce an optimal group velocity $v_{\mathrm{opt}}^{\mathrm{gr}}$ for each mode, having maximal coupling with its neighbors. Arrival times for these optimal frequencies are also almost the same, $t_{\mathrm{opt}} \sim L/v_{\mathrm{opt}}^{\mathrm{gr}}$ (in our case $v_{\mathrm{opt}}^{\mathrm{gr}} \sim 1463$ m/s), which is pointed out by the arrow in Figs. 3(a) and 3(b). In Fig. 3(b) the authors show results of the processing of one pulse received in the SWARM95 experiment. The authors also see, in these experimental data, a similar location for the arrival times of modal pulses, which undergo the most significant interaction with each other, and give the corresponding point in the TF-diagram.

5. Correlation function

The temporal correlation function for intensity at fixed frequency is the subject of many studies (see, for example, Ref. 4). It can be introduced as

$$\Gamma_\omega(\tau) = \int_0^{\Delta T} I_\omega(T) I_\omega(T-\tau) dT. \tag{11}$$

If the temporal variability is a result of IS motion with a constant velocity, $T = R/v$, then the authors can connect correlation range and correlation time via $\tau = \Delta R_p/v$, and therefore the authors can work with the correlation functions $\Gamma_\omega(\Delta R_p)$. The range dependent correlation function is often created theoretically due to lack of experimental data that is usually not readily available. However, the temporal correlation function can be constructed more simply from a single point measurement of temporal sequence of received pulses. The temporal correlation length and spatial correlation length are related with each other by the speed of perturbations in the waveguide. One can be obtained from another by using the speed of the IS.

The frequency dependent range correlation function, calculated within the framework of our model, is shown in Fig. 4(a). It shows the correlation between two different positions of the IS. The connection between the range and time correlations can be established if the authors know the velocity of the IS. However, because the authors do not know the velocity experimentally, the authors will construct this function in the model using range. The correlation length denoted in this figure is determined using the half maximum value shown in black. The authors note a comparatively weak decrease in the correlation length with increasing frequency. The value ΔR_p is estimated at about 80–110 m over a rather broad frequency band. The reason for

this weak frequency dependence of the correlation length is that only the modes turning tangent to the thermocline give a significant contribution to the range dependent fluctuations. They have close "ray cycles," and a moving soliton creates interaction mostly between these modes. Thus, in range fluctuations, there is a predominant length scale (the ray cycle distance of the above mentioned rays), which does not depend (or depends weakly) on frequency.

If the authors now consider the temporal correlation function, then given a constant velocity of the moving perturbation, the authors have a weak frequency dependence of the correlation time and also a connection between the correlation length and the correlation time. The correlation length depends mainly on the basic waveguide parameters. Thus the correlation time is sensitive only to the waveguide parameters and to the soliton velocity.

The authors can estimate the velocity of the perturbation, which produces the mode coupling. If the authors take the correlation length from theory [Fig. 4(a)] and the correlation time from experiment [Fig. 4(b)], $v=\Delta R_p/\tau\sim 1.1-1.2$ m/s. As was mentioned above, the relevant velocity of the soliton is the projection of its intrinsic velocity on the direction of propagation of solitons (thus the authors multiply by $\sin\alpha\sim 0.64$), giving $v_s\sim 0.7-0.8$ m/s. This result is in good agreement with experimental data[1] and with the results of Sec. 3.

6. Conclusion

The mechanism governing the intensity temporal fluctuations for large angle ($\sim 40°$ in SWARM95 data) between the acoustic track and IS crest can be explained by specific features of mode coupling initiated by IS in shallow water. This is in contrary to the case of small angle ($\sim 5°$), which was shown to be governed by horizontal refraction.[3]

In shallow water the sound speed perturbations (or IS) are usually concentrated in the thermocline region. Therefore, most significant mode coupling takes place between adjacent modes, at some optimal frequencies ω^m_{opt} (Fig. 2). Here, the acoustic ray turning points are found within (or near) the thermocline region. A pair of modes at the optimal sound frequency gives the most significant contribution to the temporal fluctuations. For different frequency bands, there are other pairs giving the predominant contribution to the fluctuations. However, they have approximately the same interference beat length (or ray cycle) D_{opt}, and provide the predominant frequency of intensity fluctuations Ω_{opt} (for the SWARM95 experiment $\Omega_{\text{opt}}\sim 5$ cph).

Strongly interacting modes have close group velocities, approximately $v^{\text{gr}}_{\text{opt}}$ (each pair at different frequencies) and correspondingly close arrival times t_{opt} (Fig. 3). In particular, good mode separation for narrowband pulses is possible only away from this arrival time.

The optimal parameters including the correlation time (or equivalently the correlation length) are determined mostly by the properties of the unperturbed waveguide.

Acknowledgments

The authors are grateful for the support provided by RFBR Grant No. 06-05-64853 and ONR Grant Nos. N00014-01-1-0114 and N00014-04-10146 for this research.

References and links

[1]M. Badiey, Y. J. Mu, J. F. Lynch, R. Apel, and S. N. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, 117–129 (2002).
[2]M. Badiey, B. Katsnelson, J. Lynch, S. Pereselkov, and W. Siegmann, "Measurement and modeling of 3-D sound intensity variations due to shallow water internal waves," J. Acoust. Soc. Am. **117**, 613–625 (2005).
[3]S. D. Frank, M. Badiey, J. Lynch, and W. L. Siegmann, "Analysis and modeling of broadband airgun data influenced by nonlinear internal waves," J. Acoust. Soc. Am. **116**, 3404–3422 (2004).
[4]T. F. Duda, "Temporal and cross-range coherence of sound traveling through shallow-water nonlinear internal wave packets," J. Acoust. Soc. Am. **119**, 3717–3725 (2006).

LETTERS TO THE EDITOR

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

A method for time-varying annoyance rating of aircraft noise (L)

Crispin Dickson[a)]
Marcus Wallenberg Laboratory for Sound and Vibration Research, KTH Aeronautical and Vehicle Engineering, SE-100 44 Stockholm, Sweden

(Received 27 June 2008; revised 11 March 2009; accepted 8 May 2009)

The method of continuous judgment by category is used and evaluated to measure time-varying attributes in aircraft flyover sounds. The results are also used to estimate preference between the different experimental sounds. Jurors were asked to rate perceived annoyance on a Borg CR 100 scale continuously during the playback of 11 flyover sequences and the results showed differences in perception in the time segment where the sound had been modified. The method can be used to evaluate maximum perceived annoyance, threshold levels, duration of perceptual presence temporal integration in perception, and perceptual mixtures over time.
 [DOI: 10.1121/1.3147485]

PACS number(s): 43.50.Rq, 43.66.Yw [BSF] Pages: 1–3

I. INTRODUCTION

This work is a continuation of European project Sound Engineering for Aircraft (SEFA) with aim to develop technology design criteria to make aircraft more acceptable for airport communities from noise signature standpoint. During the SEFA project, several evaluations using the paired comparison method[1] (PCM) and semantic differential method[2] (SDM) were used to investigate the perception and preference of different aircraft sounds.

Using PCM and SDM to evaluate aircraft flyover sounds the outcome gives an overall judgment on the flyover sequence of an aircraft sound but no information about temporal judgments in time segments. Due to the complexity of aircraft sounds (duration and multiple time-varying components), the PCM is tedious since the length of the sound makes it hard for the subject to remember the prior sound while the SDM evaluations become difficult since the sounds vary and may contain two extremes of a measured attribute [tonality, high and low frequency noises (HFN and LFN), etc.]. The duration of the sound also becomes a limiting factor for the accuracy since the juror has to remember the whole sound.

The objective for this work is to find a method for aircraft sound quality evaluations that overcomes the difficulties described above and that gives information about the temporal differences of perceived annoyance. This is done by a continuous annoyance rating, or method of continuous judgment by category.[3]

[a)]Electronic mail: crispin@kth.se

II. METHOD

Some modifications are made to the method as described by Kuwano.[4] Instead of rating loudness on a scale by typing keys with seven discrete values from "very soft" to "very loud," the juror is asked to use a turnable knob to rate annoyance on a continuous Borg CR 100 scale.[5] The idea of using a turnable knob rather than a slider[6] or keys was that a twisting motion would help the juror to refer the perceived annoyance to a physical exertion and would therefore work as a reference.

A. Measurement equipment

The hardware consisted of a handheld device with a turnable knob connected to a computer. The computer registered and recorded the position of the knob every 100 ms and displayed the response by the juror on a Borg CR 100 scale on the computer monitor.

B. Experimental sounds

An original recording of an aircraft flyover sound was modified[7] and the buzz saw, tonal, LFN (<315 Hz), and

TABLE I. Attribute and stimuli name.

Attribute	Stimuli name 1	Stimuli name 2
Low frequency	LFN +4 dB	LFN −4 dB
High frequency	HFN +4 dB	HFN −4 dB
Buzz saw	BSC +5 dB	BSC −5 dB
Sound level	Orig +3 dB	Orig −3 dB
Tonality	Tones +5 dB	Tones −5 dB

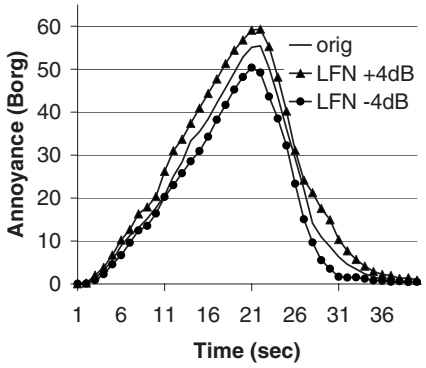


FIG. 1. Effects of modification of LFN.

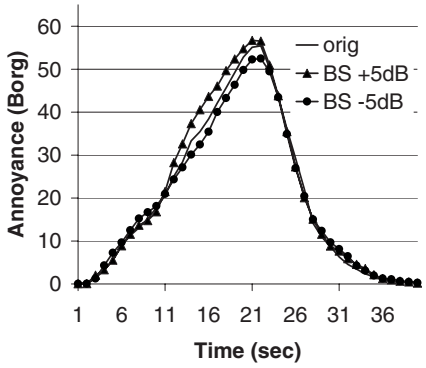


FIG. 3. Effects of modification of BSCs.

HFN (>315 Hz) components, all prominent in different parts of the sound, were modified and used to create new synthesized aircraft flyover sounds. The overall sound level was also modified. Amplifying and reducing each attribute led to 11 experimental sounds (including the unmodified sound) to evaluate (Table I). The sounds were played in a random order for each test subject at 74 dB(A) L_{max} except for the sounds with modified overall sound levels. The playlist was repeated twice, the first time for training the subject and the second for data recording. For the playback a pair of AKG K501 high fidelity headphones were used for ultimate acoustic reproduction.

C. Jurors

Nine jurors, six males and three females, aged 20–50 years performed the evaluation. The jury consisted of students and staff from the laboratory.

III. RESULTS

The results for each recording was averaged over the subjects and used for analysis.

The result for LFN is shown in Fig. 1. The analysis showed that the subjects rated the sound with reduced LFN as less annoying throughout the whole flyover sequence, while amplification led to an increase in annoyance.

The effects of modification of the HFN gave a difference in judgments in the center part of the flyover sequence, see Fig. 2. This is expected since this is the part where the HFN is prominent, as the approaching and departing sound is mainly dominated by LFN.

Buzz saw components (BSCs) are saw tooth shaped pressure waves created at the fan tips of the engine at high load, causing a very specific metallic sound[8] and can be heard at the approaching phase of an aircraft flyover. As Fig. 3 shows, the effects of modification can also be seen in this region.

The modification of the overall sound level gave an effect of the perceived annoyance during the whole flyover sequence (Fig. 4).

The tonal components, often heard in the approaching phase of a flyover, have also an effect on the annoyance. In this case a decrease did not show any effect. This could be because the tonal components were not prominent in the original recording (Fig. 5).

IV. FURTHER ANALYSIS

It is obvious from the results above that in most cases an increase in a sound component also led to an increased annoyance. The peak level of the sound was reached at 20 s, the annoyance rating peaked at 22 s. This indicates that the jurors' response time was on average 2 s delayed, which correlates with previous results by Kuwano and Namba.[3]

There could be an interest to quantify the judgments into a preference scale to conclude what sound is preferable. For this two different methods have been compared with the merit values obtained from an earlier PCM evaluation[9] (Table II). The PCM merits were calculated using the Gulliksen procedure.[10]

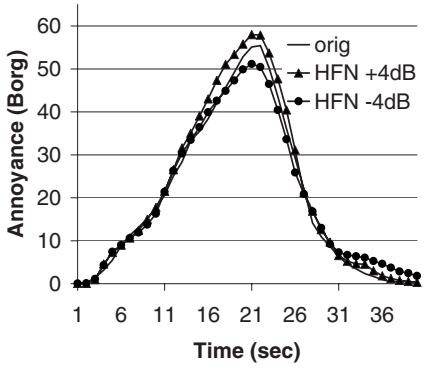


FIG. 2. Effects of modification of HFN.

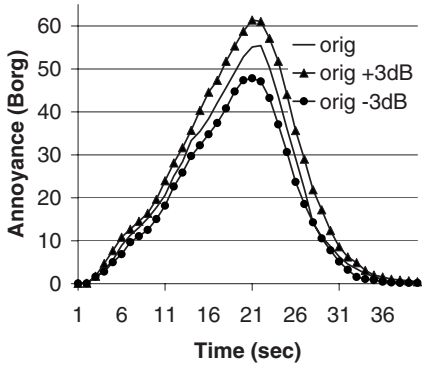


FIG. 4. Effects of modification of sound level.

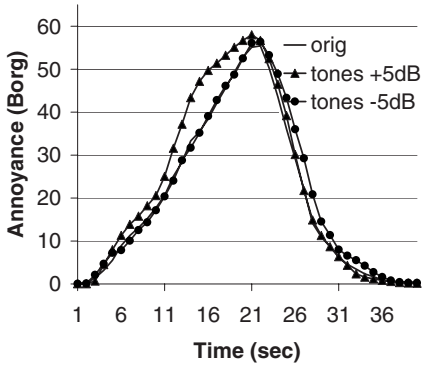


FIG. 5. Effects of modification of tonal components.

A. Sum of values above 45

The PCM merit values were compared with the sum of given responses over a certain threshold. Using 45 as a threshold value, where the perceived annoyance is "strong" according to the Borg CR 100 scale, the analysis showed a correlation coefficient of 0.85 with PCM merits.

B. Peak value

Another hypothesis was that the preference between two experimental sounds could be predicted by comparing the maximum received (peak) values. To test this, the averaged maximum annoyance values for each sound were calculated and compared to the results obtained with the PCM (Table II). An analysis showed a correlation coefficient of 0.86.

V. DISCUSSION

The method of continuous judgment by category together with Borg CR 100 scale showed results that were consistent with PCM.

The proposed method can be used to quantify annoyance as a function of maximum perceived annoyance, threshold levels, duration of perceptual presence temporal integration in perception, and perceptual mixtures over time. The evaluation with the two repetitions lasted only 20 min for each subject. Performing the evaluation using PCM would take about 2 h. By using PCM with sorting algorithms[8] the evaluation would take about 1 h. The correlation between the proposed method and the much more time consuming PCM is acceptable. For more detailed results, the test should be repeated with larger numbers of subject and more repetitions to help customize the subjects to the evaluation method.

TABLE II. The calculated values according to the different methods.

Sound	SV[a]	PV[b]	PCM[c] merit
Orig −3 dB	142.27	47.82	0.00
LFN −4 dB	192.99	50.36	0.34
HFN −4 dB	245.27	51.15	0.26
BS −5 dB	250.44	52.49	0.41
Orig	308.56	55.44	0.37
Tones −5 dB	362.09	56.41	0.32
BS +5 dB	367.15	56.69	0.45
HFN +4 dB	424.89	57.98	0.64
Tones +5 dB	527.06	58.01	0.57
LFN +4 dB	432.10	59.33	0.47
Orig +3 dB	444.62	61.40	0.64

[a]Sum of values over 45.
[b]Peak value.
[b]Paired comparison.

ACKNOWLEDGMENTS

The author wishes to thank European project SEFA for providing sound stimuli and valuable feedback and AAA foundation for financial support.

[1]L. L. Thurstone, "A law of comparative judgement," Psychol. Rev. **34**, 278–286 (1927).
[2]C. E. Osgood, G. Suci, and P. Tannenbaum, *The Measurement of Meaning* (University of Illinois Press, Urbana, IL, 1957).
[3]S. Kuwano and S. Namba, "Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness," Psychol. Res. **47**, 27–37 (1985).
[4]S. Kuwano, "Temporal aspects in the evaluation of environmental noise," in *Inter Noise 2000*, edited by D. Cassereau (Noise Control Foundation, Poughkeepsie, NY, 2000), pp. 109–119.
[5]E. Borg and G. Borg, "A comparison between AME and Borg CR100 for scaling perceived exertion," Acta Psychol. **109**, 157–175 (2002).
[6]H. Fastl, in *Sensory Research, Multimodal Perspectives*, edited by R. T. Verrillo (Lawrence Erlbaum Associates, Hillsdale, NJ, 1993), pp. 199–210.
[7]D. Berckmans, K. Janssens, P. Sas, W. Desmet, and H. Van der Auweraer, "A new method for aircraft noise synthesis," in Proceedings of the International Conference on Noise and Vibration Engineering, ISMA06, Leuven, Belgium (2006), pp. 4257–4270.
[8]U. Müller and M. Shütte, "Sound Engineering for Aircraft (SEFA), first results of listening tests," presented at Inter-Noise 2006, Honolulu, HI (2006).
[9]H. Gulliksen, "A least squares solution for paired comparisons with incomplete data," Psychometrika **21**, 125–134 (1956).
[10]C. Dickson, *A Quicker Method of Using Paired Comparisons for the Sound Quality Evaluation* (EuroNoise, Tampere, Finland, 2006).

Amplifying effect of a release mechanism for fast adaptation in the hair bundle (L)

Bora Sul[a] and Kuni H. Iwasa
Section on Biophysics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, 5 Research Court, Room 1B03, Rockville, Maryland 20850-3211

(Received 17 March 2009; revised 4 May 2009; accepted 4 May 2009)

A "release" mechanism, which has been experimentally observed as the fast component in the hair bundle's response to mechanical stimulation, appears similar to common mechanical relaxation with a damping effect. This observation is puzzling because such a response is expected to have an amplifying role in the mechanoelectrical transduction process in hair cells. Here it is shown that a release mechanism can indeed have a role in amplification, if it is associated with negative stiffness due to the gating of the mechonoelectric transducer channel.
 [DOI: 10.1121/1.3143782]

PACS number(s): 43.64.Bt, 43.64.Ld, 43.64.Nf, 43.64.Kc [BLM] Pages: 4–6

I. INTRODUCTION

Fast adaptation of the mechonoelectric transducer (MET) channel in hair bundle has been a focus of recent hair cell physiology because it is considered to be a reverse transduction mechanism with an amplifying effect (LeMasurier and Gillespie, 2005; Fettiplace, 2006; Hudspeth, 2008; Vollrath *et al.*, 2007). However, experimental examinations tend to show that the partial closure of the MET channel is accompanied by reduction in tension at the tip-link (Stauffer *et al.*, 2005; LeMasurier and Gillespie, 2005), which is attributed to elongation of the link between the MET and an unconventional myosin that is responsible for adaptation (Gillespie *et al.*, 1993; Holt *et al.*, 2002; Bozovic and Hudspeth, 2003; Martin *et al.*, 2003). These observations are puzzling in view of its presumed biological role because such a phase delayed elongation is the property of a damper and not of an amplifier.

In this report, we show that such a mechanism can have indeed an amplifying effect if it is associated with negative stiffness of the MET. In the following, we examine a specific example, which is called a "release model" (Stauffer *et al.*, 2005; LeMasurier and Gillespie, 2005) for fast adaptation. However, the applicability of the conclusion is not limited to this particular model as it will become clear by the analysis.

II. RELEASE MECHANISM

Here we give a brief description of a release mechanism. Let x_r be the length of the link, which serves as a release element that connects the MET and the slow adaptation motor. In response to displacement x at tip-link, the MET responds with force F, given by

$$F = -k_g(x - x_g P_o - x_r), \tag{1}$$

where k_g is the stiffness of the gating spring and x_g is the gating distance. The open probability P_o of the channel is determined by gating energy if its gating is much faster than relaxation,

$$P_o = \frac{1}{1 + \exp[-\beta k_g x_g (x - x_g - x_r)]}, \tag{2}$$

where $\beta = 1/k_B T$ with Boltzmann's constant k_B and the temperature T. We assume the distance x_r obeys a relaxation equation with time constant τ,

$$\frac{d}{dt}x_r = (x_{\max}P_r - x_r)/\tau. \tag{3}$$

Here $x_{\max}P_r$ is the equilibrium distance for the intracellular Ca^{2+} concentration that corresponds to open probability P_o. If this release element has one Ca^{2+}-binding site, P_r may be expressed as

$$P_r = \frac{P_o}{\nu + P_o}, \tag{4}$$

where ν is a constant.

III. RESPONSE TO SMALL DISPLACEMENTS

Let us assume that displacement x has a time-dependent component Δx,

$$x = \bar{x} + \Delta x.$$

If this displacement Δx is small, it elicits small responses ΔP_o, ΔP_r, and Δx_r in the open probability, Ca^{2+}-binding, and the release distance, respectively. Equations (2) and (4) respectively lead to

$$\Delta P_o = \beta k_g x_g \bar{P}_o(1 - \bar{P}_o)(\Delta x - \Delta x_r), \tag{5}$$

$$\Delta P_r = \frac{\nu}{(\nu + \bar{P}_o)^2}\Delta P_o, \tag{6}$$

where $\bar{P}_o$ is the steady state open probability. Equation (3) turns into

[a]Author to whom correspondence should be addressed. Electronic mail: sulb@nidcd.nih.gov

$$\frac{d}{dt}\Delta x_r = (x_{\max}\Delta P_r - \Delta x_r)/\tau = (C\Delta x - (1+C)\Delta x_r)/\tau \quad (7)$$

with $C = \beta\nu x_{\max} k_g x_g \bar{P}_o(1-\bar{P}_o)/(\nu+\bar{P}_o)^2$.

If Δx is sinusoidal and we let $\Delta x = \delta x \cos \omega t$ and $\Delta x_r = \delta x_r \cos(\omega t + \phi_r)$, Eq. (7) leads to (see Appendix)

$$\delta x_r \sin \phi_r = -\frac{\omega\tau C \delta x}{(1+C)^2 + (\omega\tau)^2}. \quad (8)$$

IV. WORK DONE DURING ONE CYCLE

For a given hair bundle displacement $x = \bar{x} + \delta x \cos \omega t$, the work W done by force F [Eq. (1)] at the tip-link during one cycle is

$$W = -k_g \int (x - x_g P_o - x_r) \cdot d\Delta x.$$

By using Eqs. (5) and (6), the integrand can be expressed by a sum of terms proportional to either Δx or Δx_r. Of these terms, only the ones proportional to Δx_r contribute, leading to

$$W = k_g(1 - \beta k_g x_g^2 \bar{P}_o(1-\bar{P}_o)) \int \Delta x_r \cdot d\Delta x.$$

Here $k_g(1-\beta k_g x_g^2 \bar{P}_o(1-\bar{P}_o))$ is known as *gating stiffness* (Howard and Hudspeth, 1988; Martin *et al.*, 2000) and will be denoted by $\tilde{k}_g$. This stiffness is reduced by the gating of the MET channel and can take negative values. Because the phase difference between Δx_r and Δx is ϕ_r, the integration over a cycle results in $\pi \sin \phi_r$. With the aid of Eq. (8), we then obtain

$$W = -\pi \tilde{k}_g \frac{\omega\tau C}{(1+C)^2 + (\omega\tau)^2} \delta x^2.$$

This result shows that the work done by the MET is negative as long as gating stiffness $\tilde{k}_g$ remains positive, implying that the MET functions as a damper for periodic stimuli. However, it should also be noted that the work done is positive, if gating stiffness is negative. Under this condition, the MET functions as an amplifier.

How can this be explained? Negative stiffness proves a 180° delay. An additional phase delay introduced by the release mechanism, in effect, gives a phase advance between 0° and 180°, providing amplification. A 90° phase delay due to the relaxation process, the condition for maximal damping, is also optimal for amplification if it is combined with negative stiffness. This observation is applicable to any relaxation process and is not specific to our model.

To take advantage of negative stiffness to do mechanical work, the system must spend energy to maintain itself in a state with negative stiffness. One such energy source is the Ca^{2+} concentration gradient across the plasma membrane and another is adenosine-5′-triphosphate (ATP) for the myosin motor.

V. DISCUSSION

We showed that the release model that we examined provides amplification when it is associated with negative stiffness. However, it is clear that this property is not specific to this particular release model but is generic to any relaxation mechanism. One such example is the model proposed by Tinevez *et al.* (2007), which posits that fast adaptation is an epiphenomenon that arises from an interplay between gating of the MET channel and the myosin motor that is responsible for slow adaptation. It includes viscoelastic relaxation and relaxation involving the movement of the myosin motor.

Here we have treated linearized response for small stimuli to obtain some insight into the issue. For this reason, we have not analyzed the stability of the system, specifically how the operating point of the MET channel, which makes gating stiffness negative can be maintained. It appears to us that the previously reported analysis (Camalet *et al.*, 2000) on the stability of the operating point of the MET would be applicable to our model.

Because negative stiffness is intrinsically unstable, a relatively large stimulus used for experiments would shift the system into a condition with positive gating stiffness. Such a large stimulus is outside the validity of our treatment. Together with the difficulty of achieving extremely high time resolution in stimulation and recording, it may not be surprising to record only relaxation components during experiments (Stauffer *et al.*, 2005; LeMasurier and Gillespie, 2005).

ACKNOWLEDGMENT

This research was supported by the Intramural Research Program of the NIDCD, NIH.

APPENDIX: DERIVATION OF EQUATION (8)

Since $\Delta x = \mathrm{Re}[\delta x e^{(i\omega t)}]$ and $\Delta x_r = \mathrm{Re}[\delta x_r e^{i(\omega t+\phi_r)}]$, Eq. (7) can be expressed as

$$i\omega\tau \delta x_r e^{i(\omega t+\phi_r)} = C\delta x e^{i\omega t} - (1+C)\delta x_r e^{i(\omega t+\phi_r)},$$

which leads to

$$\delta x_r e^{i\phi_r} = \frac{C}{(1+C)+i\omega\tau}\delta x.$$

The imaginary part of this equation is Eq. (8).

Bozovic, D., and Hudspeth, A. J. (**2003**). "Hair-bundle movements elicited by transepithelial electrical stimulation of hair cells in the sacculus of the bullfrog," Proc. Natl. Acad. Sci. U.S.A. **100**, 958–963.

Camalet, S., Duke, T., Jülicher, F., and Prost, J. (**2000**). "Auditory sensitivity provided by self-tuned critical oscillations of hair cells," Proc. Natl. Acad. Sci. U.S.A. **97**, 3183–3188.

Fettiplace, R. (**2006**). "Active hair bundle movements in auditory hair cells," J. Physiol. **576**, 29–36.

Gillespie, P. G., Wagner, M. C., and Hudspeth, A. J. (**1993**). "Identification of a 120 kD hair-bundle myosin located near stereociliary tips," Neuron **11**, 581–594.

Holt, J. R., Gillespie, S. K. H., Provance, D. W., Shah, K., Shokat, K. M., Corey, D. P., Mercer, J. A., and Gillespie, P. G. (**2002**). "A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells," Cell **108**, 371–381.

Howard, J., and Hudspeth, A. J. (**1988**). "Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the

bullfrog's saccular hair cell," Neuron **1**, 189–199.
Hudspeth, A. J. (**2008**). "Making an effort to listen: Mechanical amplification in the ear," Neuron **59**, 530–545.
LeMasurier, M., and Gillespie, P. G. (**2005**). "Hair-cell mechanotransduction and cochlear amplification," Neuron **48**, 403–415.
Martin, P., Bozovic, D., Choe, Y., and Hudspeth, A. J. (**2003**). "Spontaneous oscillation by hair bundles of the bullfrog's sacculus," J. Neurosci. **23**, 4533–4548.
Martin, P., Mehta, A. D., and Hudspeth, A. J. (**2000**). "Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell," Proc. Natl. Acad. Sci. U.S.A. **97**, 12026–12031.
Stauffer, E. A., Scarborough, J. D., Hirono, M., Miller, E. D., Shah, K., Mercer, J. A., Holt, J. R., and Gillespie, P. G. (**2005**). "Fast adaptation in vestibular hair cells requires myosin-1c activity," Neuron **47**, 541–553.
Tinevez, J.-Y., Jülicher, F., and Martin, P. (**2007**). "Unifying the various incarnations of active hair-bundle motility by the vertebrate hair cell," Biophys. J. **93**, 4053–4067.
Vollrath, M. A., Kwan, K. Y., and Corey, D. P. (**2007**). "The micromachinery of mechanotransduction in hair cells," Annu. Rev. Neurosci. **30**, 339–365.

An efficient code for environmental sound classification (L)

Raman Arora
Department of Electrical and Computer Engineering and Auditory Behavioral Research Laboratory, University of Wisconsin, Madison, Wisconsin 53706

Robert A. Lutfi
Department of Communicative Disorders and Auditory Behavioral Research Laboratory, University of Wisconsin, Madison, Wisconsin 53706

(Received 19 January 2009; revised 29 April 2009; accepted 30 April 2009)

An efficient code for classifying environmental sounds is described that exploits a recent significant advance in signal processing known as compressed sensing (CS) [cf.Donoho, D. (2006). IEEE Trans. Inf. Theory **52**, 1289–1306]. CS involves a novel approach to sampling in which the salient information in signals is recovered from the projection onto a small set of random basis functions. The advantage of the random basis over traditional Fourier or wavelet representations is that it allows accurate classification at low target-to-interference ratios based on few samples and little or no prior information about signals.
 [DOI: 10.1121/1.3139982]

PACS number(s): 43.66.Ba [DOS] Pages: 7–10

I. INTRODUCTION

Much work has been devoted in recent years to the goal of developing an automated sound recognition system that can accurately and efficiently classify a wide variety of common environmental sounds according to their generating source. The effort is driven, in part, by the desire to understand through computation modeling our own ability to reconstruct from sound an accurate perception of everyday objects and events present in our natural environment (Bregman, 1990; Ellis, 1996, Lutfi, 2008). It is also motivated by the potential for significant benefit arising from the practical application. Some of the most promising applications are now being pursued in the areas of remote surveillance (Cristani *et al.*, 2004, 2007; Cowling and Sitte, 2003; Wang *et al.*, 2008) and multimedia database search (Wang *et al.*, 2006).

The development of an automated environmental sound recognition system poses a number of significant challenges, but first among these is the problem of how best to encode the salient information in signals. The information rates associated with environmental signals are exceedingly high, and not all information in signals will be diagnostic regarding the identity of their source. Preserving a large amount of information in the initial encoding of signals is costly because it requires increased rates of sampling and subsequent computation. At the outset, then, one must decide what information in signals should be preserved and what information can be discarded without significant loss in classification accuracy. The traditional approach to this problem has been to represent signals partially in some basis that shares a common structure with the signals to be encoded, typically a Fourier or wavelet basis. Signals are sampled densely in this basis and all information other than that provided by a small number (M) of significant coefficients is discarded for further processing. This approach preserves the salient information in signals, but is clearly wasteful in that it requires a good deal of upfront processing of information that ultimately will not be used.

A second and equally important issue pertaining to encoding is the interference produced by ambient noise in the environment. Target signals rarely occur in isolation; rather, they are usually accompanied by other unwanted environmental sounds that compete with the targets. In many applications the target signal-to-noise ratio is also low (e.g., military surveillance applications involving the detection of intentionally concealed targets). In such cases much of the information in targets may be lost in Fourier or wavelet basis since the maximum-valued coefficients are likely to be determined predominantly by the noise. Prior knowledge of the interference can help in these cases to isolate the target from the interference, but such knowledge may be incomplete or entirely lacking in many applications.

The present paper offers a new approach to the classification of environmental sounds that deals effectively with the high information rates associated with potential signals and the interference produced by extraneous sound sources. It does so by exploiting a recent significant advance in the efficient encoding of signals known as compressed sensing (cf. Donoho, 2006). Compressed sensing (CS) is an emergent technology that has found increasing application in the areas of broadband signal monitoring and image reconstruction. It involves a novel approach to sampling in which the salient information in signals is recovered from the projection onto a small set of random basis functions. In the present paper the advantage of CS-based classification over traditional Fourier-based classification is demonstrated by applying both to a representative case of an environmental sound classification task.

II. COMPRESSED SENSING

To appreciate the power of CS and the insight on which it is based, it is instructive to revisit the logic underlying the

traditional approach to signal encoding. This logic says that we should project signals onto a basis that mirrors as closely as possible the underlying structure of the signals. In so doing, we can expect that a small number of basis functions, those having the largest coefficients, will capture most of the salient information in signals. The logic is correct, but it provides no recipe on how to identify the salient functions without first projecting signals onto the complete basis; a rather wasteful exercise. CS avoids this problem by taking an exactly opposite approach to encoding. Instead of matching the basis to the inherent structure of signals, it requires that we project onto a basis that is void of any structure, one for which the basis functions share no feature in common with signals. The one basis having this property for all signals is the random basis, which has noise waveforms as basis functions. The logic may seem counterintuitive; but, consider that, unlike a Fourier or wavelet basis, each function of the random basis is virtually ensured to have some measurable correlation with the signal, positive or negative. This is a direct consequence of the noise basis functions being broadband. The insight of CS is that only a small number of these correlations are required to recover the signal without error—we need not project onto the entire basis. All that is required is that the signals be sparse in some basis; that they be well-represented in some basis by a small number of nonzero coefficients.[1] Sparsity, in fact, is a feature of environmental sounds that underlies the success of popular compression schemes such as mp3. Consider that a vast number of the sounds we encounter in our everyday environment are emitted by objects having relatively simple geometries; bells, bars, plates, and membranes with different supports. Because of their simple geometries these objects have relatively few prominent modes of vibration, and so most of the salient information in their emitted sounds is captured by a few nonzero Fourier coefficients. Other environmental sounds, not sparse in the frequency domain, are nonetheless represented by a small number of nonzero values in time. The sound of footsteps, birds chirping, hands clapping, and many machine sounds are but to name a few. Indeed, with the exception of continuous broadband noise, it is difficult to imagine a natural sound that cannot practically be considered sparse in either frequency or time. In what follows, then, we use the sparse property of environmental sounds to advantage in their classification.

III. METHOD

The 50 environmental sounds used in the simulation (25 targets and 25 interferers) were taken from the high-quality sound effects CDs "Hollywood Leading Edge Sound FX, The General." They are listed in Table I. These sounds were selected because they have been shown to be easily identified by human listeners in a previous psychophysical study (see Gygi *et al.*, 2004); hence, they were deemed, in this sense, to be "common" environmental sounds. The sounds were also selected so that they would span a broad range of different sound categories (e.g., machine, human, weather, and animal sounds). The original sound recordings ranged in duration from 0.5 to 3.6 s. For the simulation they were normalized in duration to 3.6 s by zero padding where necessary. They were also equated in total rms. Finally, because of the significant amount of computation required for the simulation, the sounds were down-sampled from 44.1 to 4 kHz. Each waveform then was effectively low-pass filtered at 2000 Hz and contained a total of $N=3.6\ \text{s}\times 4000/\text{s}=14\,400$ samples.

TABLE I. Waveforms used in the simulation as named in the Hollywood Leading Edge Sound FX CDs.

Targets	Interferers
89_BellChrch.wav	89_CarStart.wav
89_GlassSmash.wav	89_HelicPassby.wav
89_IceDropIntoGlass.wav	89_SprtBowlingStrike.wav
BABYCRY.wav	BASKBALL.wav
BIRD.wav	BUBLES.wav
CATX.wav	CHICOUGH.wav
CLAPSA.wav	CLOCK.wav
COWA.wav	CRASHA.wav
CRICKETS.wav	CYMBALA.wav
DOGX2.wav	DOOROC.wav
DRUMS.wav	FOOTSTP.wav
GUN.wav	HAMMERNG.wav
HARP.wav	HORSERUN.wav
HORSEWNA.wav	LAUGH.wav
MATCH.wav	PEELOUT.wav
PHONE.wav	PINGPA.wav
POURWATR.wav	PlanePropPassby.wav
ROOSTER.wav	SHEEP.wav
SHOVEL.wav	SIREN.wav
SNEEZE.wav	SPLASH.wav
STAPLER.wav	ScissorsHrCut.wav
TENNIS.wav	THUNDER.wav
TOILET.wav	TRAIN.wav
TYPEWRI.wav	WistlePlc.wav
WtrRain.wav	ZIPPERA.wav

The simulation was conducted as follows: (1) Select at random M Gaussian noise waveforms each of length N to construct an $M\times N$ matrix $\mathbf{R}$ as the random basis to be used in all conditions. (2) Compute and store the $M\times 1$ vector $\mathbf{a}_i$ of random basis coefficients for each of the $i=1\cdots 25$ targets,

$$\mathbf{a}_i=\mathbf{R}\mathbf{x}_i, \tag{1}$$

where the $N\times 1$ vector $\mathbf{x}_i$ is the ith discrete target waveform. (3) Compute the $M\times 1$ vector $\mathbf{b}_j$ of basis coefficients for each of the $j=1\cdots 625$ possible combinations of target and interferer,

$$\mathbf{b}_j=\mathbf{R}(\mathbf{x}_i+\alpha\mathbf{y}_k), \tag{2}$$

where the $N\times 1$ vector $\mathbf{y}_k$ is the kth discrete interferer waveform and α determines the target-to-interference ratio. (4) For each target+interferer combination given by j identify the target by the index $\hat{i}$, where $\mathbf{a}_{\hat{i}}$ yields the largest inner-product with $\mathbf{b}_j$,

$$\hat{i}=\arg\max_{i=1\cdots 25}\langle\mathbf{a}_i,\mathbf{b}_j\rangle. \tag{3}$$

(5) Repeat steps 1–4 for values of M ranging from 1 t 256 and at different target-to-interference ratios ranging from −20 to 20 dB. (6) Repeat steps 1–5 replacing the $M\times N$

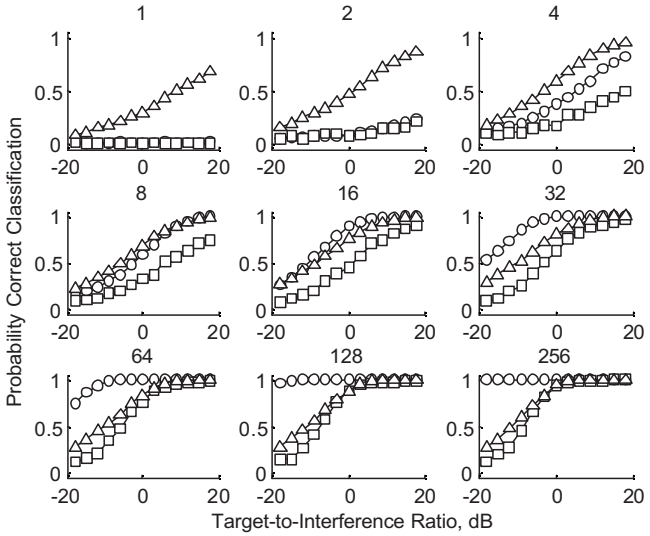


FIG. 1. Percent correct identification performance for CS (circles), MAXF (triangles), and RANDF (squares) is plotted as a function of the target-to-interference ratio for different values of M indicated at the top of each panel.

random basis $\mathbf{R}$ with the $N \times N$ Fourier basis and choosing for $\mathbf{a}_i$ and $\mathbf{b}_j$ the M maximum Fourier coefficients in each case (MAXF classifier). (7) Repeat steps 1–4 replacing the $M \times N$ random basis $\mathbf{R}$ with the incomplete $M \times N$ Fourier basis, using the same M randomly-selected rows of the complete Fourier basis in all conditions (RANDF classifier).

IV. RESULTS

Figure 1 shows the performance of the three classifiers as a function of the target-to-interference ratio with the number of coefficients M as parameter. Note first that the poorest performance overall is achieved by RANDF. Only for $M > 32$ and a target-to-interference ratio greater than 0 does RANDF begin to match the performance of CS. This shows that random sampling alone cannot account for the high performance levels achieved by CS. Compared to MAXF the performance of CS is everywhere poorer for $M < 8$, but somewhere between $M = 8$ and 16 the performance of CS begins to better that of MAXF as M is increased. The reversal is particularly evident at the lower target-to-interference ratios where the performance of CS continues to improve and is near perfect by $M = 128$. The failure of MAXF performance to improve similarly over this range is due to the interference, which at the lower target-to-interference ratios determines the significant Fourier coefficients. This point merely underscores one of the fundamental differences between CS and MAXF described in the Introduction: MAXF can only improve its performance at low target-to-interference ratios by using prior knowledge of the interference to separate target and interference. CS makes no attempt at separation and so requires no such knowledge. The projection onto the random basis ensures that any portion of the target not obscured by the interference will contribute to the classification. Finally, we note that CS is overall considerably more efficient than MAXF. CS performs as well or better than MAXF at all target-to-interference ratios with as few as 16 projections, compared to $N = 3.6\ \text{s} \times 4000/\text{s} = 14\,400$ projections for MAXF.

V. DISCUSSION

We have shown that, at target-to-interference ratios as low as −20 dB, CS achieves near perfect classification of an arbitrary set of environmental sounds with only 128 projections, and equal or better classification performance than the M maximum Fourier coefficients with as few as 16 projections. The results, of course, are for a highly restricted case in which there is only one exemplar of each target class, and in which targets and distracters can reasonably be considered sparse. A much stronger test of the algorithm would require its application to more realistic tasks in which there are multiple exemplars of each target class, including non-sparse targets and distracters. Notwithstanding, the results encourage speculation as to how CS might be recruited in the effort to model human sound source classification. Traditional approaches based on the extraction of structured features and programed schema for separating sound sources require high information rates and much prior knowledge of signals (Ellis, 1996; Martin, 1999). Yet, they still fall pitifully short of the human capacity for classifying sounds in everyday listening. The present results suggest that the problem may be more manageable than these efforts imply. The information rates, perhaps, need not be as high if compression can occur simultaneously with sampling, and the amount of prior signal knowledge, perhaps, need not be as great if classification can occur frequently without the need to separate sound sources. It remains to be seen whether the attractive features of CS could be incorporated into a computational model that would eventually approach the remarkable performance of the human classifier in more realistic listening situations than considered here. The present results are, at least, encouraging.

ACKNOWLEDGMENT

This research was supported by NIDCD Grant No. 5R01DC006875-04.

[1]The reconstruction in this case involves minimization of the L-1 norm, which virtually ensures errorless reconstruction. For a very readable introduction to the theory of compressed sensing the reader is referred to the review article by Baraniuk (2007).

Baraniuk, R. (**2007**). "Compressive sensing," IEEE Signal Process. Mag. **24**, 118–121.
Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
Cowling, M., and Sitte, R. (**2003**). "Time-frequency environmental sound recognition for autonomous robot surveillance," Proceedings of AMiRE, Brisbane, Australia.
Cristani, M., Bicego, M., and Murino, V. (**2004**). "On-line adaptive background modelling for audio surveillance," in Proceedings of the International Conference on Pattern Recognition (ICPR 2004), pp. 399–402.
Cristani, M., Bicego, M., and Murino, V. (**2007**). "Audio-visual event recognition in surveillance video sequences," IEEE Trans. Multimedia **9**, 257–267.
Donoho, D. (**2006**). "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306.
Ellis, D. P. W. (**1996**). "Prediction-driven computational auditory scene analysis," Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
Gygi, B., Kidd, G. R., and Watson, C. S. (**2004**). "Spectral-temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–1265.

Lutfi, R. A. (**2008**). "Human sound source identification," *Auditory Perception of Sound Sources* (Springer, New York), pp. 13–42.
Martin, K. D. (**1999**). "Sound-Source Recognition: A Theory and Computational Model," Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
Wang, J. C., Lee, H. P., Wang, J. F., and Lin, C. B. (**2008**). "Robust environmental sound recognition for home automation," IEEE Trans. Autom. Sci. Eng. **5**, 25–31.
Wang, J. C., Wang, J. F., He, K. W., and Hsu, C. S. (**2006**). "Environmental sound classification using hybrid SVM/KNN and MPEG-7 audio low-level descriptor," International Joint Conference on Neural Networks, Vancouver, BC, Canada.

Pile driving zone of responsiveness extends beyond 20 km for harbor porpoises (*Phocoena phocoena* (L.)) (L)

Jakob Tougaard,[a] Jacob Carstensen, and Jonas Teilmann
National Environmental Research Institute Aarhus University, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark

Henrik Skov
DHI Water and Environment, DK-2800 Hørsholm, Denmark

Per Rasmussen
G.R.A.S. Sound and Vibration A/S, DK-2840 Holte, Denmark

(Received 3 April 2008; revised 27 March 2009; accepted 17 April 2009)

Behavioral reactions of harbor porpoises (*Phocoena phocoena*) to underwater noise from pile driving were studied. Steel monopile foundations (4 m diameter) for offshore wind turbines were driven into hard sand in shallow water at Horns Reef, the North Sea. The impulsive sounds generated had high sound pressures [source level 235 dB re 1 μPa_{pp} at 1 m, transmission loss 18 log(distance)] with a strong low frequency emphasis but with significant energy up to 100 kHz. Reactions of porpoises were studied by passive acoustic loggers (T-PODs). Intervals between echolocation events (encounters) were analyzed, and a significant increase was found from average 5.9 h between encounters in the construction period as a whole to on average 7.5 h between first and second encounters after pile driving. The size of the zone of responsiveness could not be inferred as no grading in response was observed with distance from the pile driving site but must have exceeded 21 km (distance to most distant T-POD station).
 [DOI: 10.1121/1.3132523]

PACS number(s): 43.80.Nd, 43.30.Nb, 43.50.Pn [WA] Pages: 11–14

I. INTRODUCTION

The process of driving large diameter steel monopiles into the seabed in connection to offshore construction activities can generate underwater sound pressures with source levels in excess of 230 dB re 1 μPa (Nedwell *et al.*, 2005). These sounds are likely to disturb marine mammals at considerable distance and even be dangerous at close range (see review by Madsen *et al.*, 2006).

The North Sea is an area where a high incidence of pile driving activity can be expected in the coming years due to the establishment of offshore wind farms in the German Bight and along the east coast of Great Britain. The most common cetacean in the North Sea is the harbor porpoise (*Phocoena phocoena*). As harbor porpoises are known to show aversive reactions to unfamiliar sounds of low to moderate intensity (e.g., Teilmann *et al.*, 2006), there is a potential for significant impact of pile driving on this species. During construction of Horns Rev Offshore Wind Farm in the Danish North Sea, we had the opportunity to study reactions of harbor porpoises to pile drivings by means of stationary acoustic monitoring.

II. MATERIALS AND METHODS

Horns Rev Offshore Wind Farm was built in 2002 and is located about 25 km west of the Danish coast on Horns Reef (55°29”N 7°50”E). The wind farm consists of 80 turbines placed on a hard sandy bottom in 6–12 m of water. Turbine foundations consist of 4 m diameter steel monopiles, which were driven approximately 30 m into the seabed with a hydraulic hammer (IHC Hydrohammer S-600) from a jack-up rig. The hammer delivered on average about one blow per second. Each pile driving operation took between 0.5 and 2.5 h to complete, with total duration determined both by total number of blows delivered and duration of breaks in pile driving due to readjustment of the monopile. Pile driving was performed in the period March 30 to August 1, 2002 on days with calm weather.

Mitigation measures were taken to protect seals and porpoises from exposure to excessive and possibly dangerous levels of noise during pile driving. Acoustic pingers (Aquamark100) were deployed on all anchors of the rig, and a seal scarer (Lofitek) was lowered into the water from the rig whenever the jack-up rig was anchored at a new position. Furthermore a ramp-up of energy delivered to the pile was performed at the beginning of each pile driving. This ramp-up did not follow a standardized scheme but occurred as a natural consequence of the gradual increase in delivered energy over the first series of blows as the pile was gradually advanced into the top meters of the seabed.

Effects on harbor porpoises were studied by static acoustic monitoring: T-PODs (Chelonia Inc., UK, Version 1) placed inside the wind farm area and outside in reference areas. Three stations were operational and provided useful data in the period during which pile driving was undertaken. One station was located inside the 4×4 km^2 wind farm

[a]Author to whom correspondence should be addressed. Electronic mail: jat@dmu.dk

area, a second station 7.5 km to the east of the wind farm and the third station 21 km west of the wind farm. T-PODs were mounted in 500 kg steel cages with the hydrophone about 1.5 m above the seabed in depths between 5 and 10 m.

Briefly described, the T-POD consists of a sensitive hydrophone and a detection circuit designed to detect harbor porpoise signals and logs only the time of occurrence of detected clicks. For a detailed description of the T-POD see Kyhn *et al.* (2008). Settings used were as follows: A (target) filter: 130 kHz, short integration time; B (reference) filter: 90 kHz, long integration time; A/B ratio: 5; A-filter threshold: 0; scan limit: 160; minimum click duration 50 μs. Duty cycle was 75% (5 periods of 9 s monitored every minute). Following download to computer the data were analyzed by the associated software (Tpod.exe Version 5.41), where clusters of associated clicks were grouped into trains and filtered into five categories by the train classification algorithm of the software, according to the likelihood that the clicks originated from porpoises. Only click trains from the two best categories, "cetacean high" and "cetacean low," were used for analysis. Trains, which occurred close to each other in time, were further grouped into clusters, termed encounters. An encounter was defined as a group of click trains with no gaps between individual trains exceeding 10 min (Carstensen *et al.*, 2006). Inter-encounter intervals were determined as the silent periods between encounters and by definition of the encounters; by definition, intervals were always at least 10 min long.

A. Statistical analysis

The effect of pile driving was investigated by comparing inter-encounter intervals associated with pile driving operations to inter-encounter intervals recorded during the entire construction period from March 30 to August 1, 2002. The first three encounters after end of each pile driving operation were identified, and the corresponding inter-encounter intervals prior to these encounters were analyzed. Even in the absence of a reaction to the pile driving the first inter-encounter interval will have an expected value larger than the mean of all inter-encounter intervals when intervals are sampled in this way. This is due to a sampling bias often referred to as the "bus paradox" (Ito *et al.*, 2003) and dictates that if one picks a random minute of the day (end point of pile driving) and evaluates the duration of the interval into which this minute falls, then there is a higher probability of picking a long interval than a short interval simply because there are more minutes to pick from in the long intervals. To circumvent this serious sampling bias inter-encounter intervals associated with pile driving were compared to an equivalent number of inter-encounter intervals randomly sampled from the entire ensemble of intervals during the construction period (minus those associated with pile driving).

The statistical analyses were carried out within the framework of mixed linear models by means of PROC MIXED in the SAS system (SAS Institute Inc., Cary, NC). Inter-encounter intervals were log-transformed after subtraction of 10 min as the intervals by definition have a lower limit of 10 min. Intervals were then modeled as

$$Y_{t(ij)} = \mu + t_i + s_j + ts_{ij} + e_{t(ij)}. \quad (1)$$

Y is the observed interval duration, equal to the sum of μ (is the overall mean log-transformed interval), t (contribution from pile driving; two levels: pile driving or not pile driving), and s (variation among stations; three levels for positions 1, 6, and 7). The interaction term ts describes a differential response to pile driving across stations, thus allowing, for example, for a stronger response close to the pile driving site. The error term e contains residual variation not explained by the model with different variances for intervals associated with pile driving and remaining intervals.

A formal statistical test of differences in the two distributions could not be performed as appropriate tests for this scenario are not available. Instead the sampling and comparison procedure where intervals were sampled at random from the entire construction period and compared to intervals associated with pile driving was repeated 1000 times in a boot strap-like fashion, and medians of the test statistics were used to describe differences in the two distributions. Thus, pile driving activity terms [t and ts of Eq. (1)] and station term s were tested for significance with a partial F-test, i.e., considering the specific contribution of the given factor in addition to all other factors. A level of significance of 5% was used.

Second inter-encounter intervals after pile driving were compared to the intervals immediately following the intervals sampled for the first comparison, and similarly for third inter-encounter intervals. As second and third inter-encounter intervals thus were not sampled independently, but followed directly once the first inter-encounter interval was sampled, they are expected to have distributions similar to the overall distribution of intervals, i.e., not subject to the bus paradox sampling bias.

B. Acoustic measurements

Sound measurements were made on May 2, 2002 during pile driving of the last 10 m of the monopile at turbine foundation No. 17 on a line almost straight east from the foundation. Water depth was approximately 6.5 m at the four measuring stations. Measurements were done with a Reson TC4034 hydrophone and analyzed online with a spectrum analyzer (Stanford research analyzer, SR785). The setup was calibrated on site by means of a pistonphone (G.R.A.S 42AC) but for technical reasons the power spectrum could not be absolutely referenced and total signal energy could not be calculated.

Measurements were made from a small ship with the hydrophone 4 m below the water surface.

III. RESULTS

A. Behavioral reactions

In total 1811 encounters were recorded on the three T-PODs in the study period. Ninety-one inter-encounter intervals were associated with pile driving (29, 15, and 47 pile

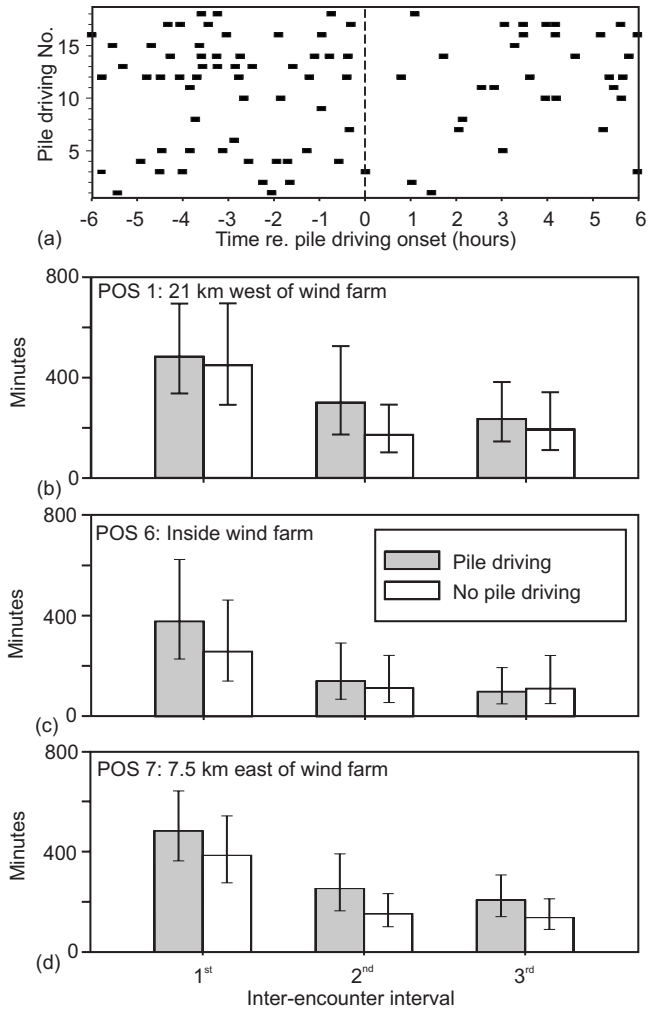


FIG. 1. (a) Example of acoustic detection of harbor porpoises (encounters) in the hours before, during, and after pile driving at Horns Reef. Data from station 6 (inside construction area) during the last month of pile driving activity (July 2002), where 18 pile driving operations were conducted. (b)–(d) Mean inter-encounter intervals prior to first, second, and third encounters after pile driving compared to mean inter-encounter intervals sampled from the entire construction period from March 30 to August 1, 2002. Error bars show 95% confidence intervals of the mean values.

driving events for positions 1, 6, and 7, respectively) and were used in the analysis as first inter-encounter intervals. Thirty-eight of these first inter-encounter intervals (11, 4, and 23 for positions 1, 6, and 7, respectively) started during pile driving whereas the remainder intervals started prior to onset of pile driving. By definition, all first intervals ended at the first encounter following cessation of the associated pile driving.

First inter-encounter intervals after pile driving operations were longer and less variable in duration than the overall sample of inter-encounter intervals, seen in Fig. 1(a) as an increased gap between encounters following onset of pile driving. The statistical analysis showed a significant general increase in the first inter-encounter interval after pile driving operation had ceased, whereas the increase in second and third inter-encounter intervals was not significant [Figs. 1(b)–1(d); Table I; term t in Eq. (1)]. The variation across stations during the whole study period was not significant [term s in Eq. (1); p_s=0.1894 for the first inter-encounter interval, p_s=0.1791 for the second inter-encounter interval, and p_s=0.0742 for the third inter-encounter interval]. The relative increase in inter-encounter intervals [interaction term ts in Eq. (1)] was not significantly different among stations for neither the first (p_{st}=0.5604), second (p_{st}=0.6840), nor third inter-encounter interval (p_{st}=0.4845). In other words, there was no grading in response with distance from pile driving site. The data did not permit a test of whether inter-encounter intervals changed over the entire experimental period, but there were no clear indications in the data of a habituation of the harbor porpoises to the pile driving sounds over the period.

B. Sound measurements

Measurements of impact sound from a single pile driving are shown in Fig. 2. The energy delivered to the pile in the period when measurements were obtained varied between 360 and 450 kJ per blow.

Impulse sounds had durations around 0.2 s. Signals had peak energy at 160 Hz but also significant energy at considerably higher frequencies, up to and possibly beyond 100 kHz, which was the upper frequency limit of the recording equipment (Fig. 2, insert). Sound pressures decreased with distance from the pile driving site with a transmission loss of nearly 6 dB per doubling of distance. Based on this transmission loss, the sound pressure level was back-calculated to a source level of 235 dB re 1 μPa$_{pp}$ at a distance of 1 m. Actual sound pressure levels close to the foundation were likely considerably lower due to near field effects created by the very large transduction area (4 m diameter cylinder in 6 m of water).

IV. DISCUSSION

Harbor porpoises at Horns Reef reacted to pile driving operations at all three measuring stations. The fact that the

TABLE I. Statistics of inter-encounter intervals between acoustic encounters of harbor porpoises in connection to pile driving and for the construction period as a whole. Results of statistical test on first, second, and third intervals following end of pile driving shown in rightmost columns.

	Following pile driving			All other				
Inter-encounter interval	Mean (min)	Median (min)	Variance (log transf.)	Mean (min)	Median (min)	Variance (log transf.)	F	P
First	447	273	1.01	355	175	1.48	6.54	0.0114
Second	221	78	2.27	146	59	2.06	1.71	0.1791
Third	169	74	1.81	145	54	2.16	1.88	0.1727

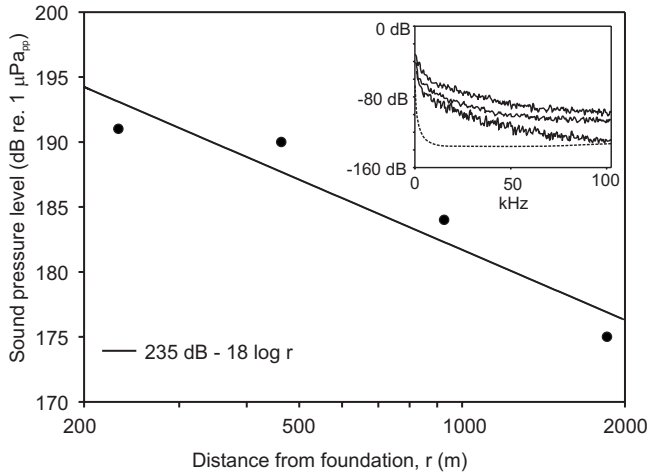


FIG. 2. Sound pressure level measured at four different distances from pile driving of a 4 m diameter steel monopile into hard sand on Horns Reef. Straight line is best fitting simple transmission loss model. Insert: Power spectra of pile driving impact pulses measured at three different distances from the pile driving site at Horns Reef (from top to bottom: 230, 460, and 930 m), where a 4 m diameter steel monopile was driven into hard sand at approximately 6 m water depth. Broken line indicates background noise/system noise.

response was not graded across stations with increasing distance from the construction site is surprising, yet consistent with similar measurements during pile driving at another offshore wind farm (Nysted, Western Baltic; Carstensen *et al.*, 2006). All three measuring stations were thus well within the zone of responsiveness for porpoises despite the large difference in the exposure to sound. The zone of responsiveness may thus have extended well beyond 20 km but in the absence of a grading in responses we are unable to extrapolate beyond the furthest measuring station.

When discussing behavioral effects of any impact, the important point is how this behavioral change affects the long term fitness of the individual animals and the local population as a whole (Bejder *et al.*, 2006). Each pile driving clearly represents a considerable impact on the individual porpoise within the zone of responsiveness, yet this disturbance should be seen in a broader perspective. If porpoises within the zone of responsiveness are prevented from feeding (or nursing their calves) for the entire duration of the pile driving and for the following couple of hours, the impact may not be trivial, in particular, if many pile driving operations are undertaken in the same area within a short time span. However, the porpoises at Horns Reef most likely do not form a stable and stationary population but are part of a much larger and spatially dynamic population in the North Sea. Thus, if porpoises within the zone of responsiveness are forced out of the area by the pile driving, but then are able to resume feeding again as soon as they are out of range of the pile driving sounds, the impact on the individual porpoise may be much less severe. Studies that can address impact on individual animals in terms of energetic consequences are clearly needed in order to assess the long term consequences of the disturbance.

ACKNOWLEDGMENTS

The porpoise monitoring was funded by the Danish Energy Authority as part of the National Demonstration Project on Offshore Wind Farms and conducted under contract with Vattenfall A/S. Underwater sound recordings courtesy of Elsam A/S. Oluf D. Henriksen, NERI, played an important role in developing and implementing the monitoring program and Ole Lindbjerg, Elsam Engineering A/S, was responsible for deployment and recovery of T-PODs. Nick Tregenza is thanked for always being available with help and suggestions on technical problems with the T-PODs. Klaus Lucke is thanked for valuable suggestions to the paper and for allowing references to his work on TTS in harbor porpoises.

Bejder, L., Samuels, A., Whitehead, H., and Gales, N. (**2006**). "Interpreting short-term behavioural responses to disturbance within a longitudinal perspective," Anim. Behav. **72**, 1149–1158.

Carstensen, J., Henriksen, O. D., and Teilmann, J. (**2006**). "Impacts on harbour porpoises from offshore wind farm construction: Acoustic monitoring of echolocation activity using porpoise detectors (T-PODs)," Mar. Ecol.: Prog. Ser. **321**, 295–308.

Ito, J., Nikolaev, A. R., Luman, M., Aukes, M. F., Nakatani, C., and Leeuwen, C. V. (**2003**). "Perceptual switching, eye movements, and the bus paradox," Perception **32**, 681–698.

Kyhn, L. A., Tougaard, J., Teilmann, J., Wahlberg, M., Jørgensen, P. B., and Bech, N. I. (**2008**). "Harbour porpoise (*Phocoena phocoena*) static acoustic monitoring: Laboratory detection thresholds of T-PODs are reflected in field sensitivity," J. Mar. Biol. Assoc. U.K. **88**, 1085–1091.

Madsen, P. T., Wahlberg, M., Tougaard, J., Lucke, K., and Tyack, P. L. (**2006**). "Wind turbine underwater noise and marine mammals: Implications of current knowledge and data needs," Mar. Ecol.: Prog. Ser. **309**, 279–295.

Nedwell, J., Workman, R., and Parvin, S. J. (**2005**). *The Assessment of Likely Levels of Piling Noise at Greater Gabbard and Its Comparison With Background Noise, Including Piling Noise Measurements Made at Kentish Flats*, (Subacoustec, Hampshire, UK).

Teilmann, J., Tougaard, J., Miller, L. A., Kirketerp, T., Hansen, K., and Brando, S. (**2006**). "Reactions of captive harbor porpoises (*Phocoena phocoena*) to pinger-like sounds," Marine Mammal Sci. **22**, 240–260.

Use of forward pressure level to minimize the influence of acoustic standing waves during probe-microphone hearing-aid verification

Ryan W. McCreery,[a] Andrea Pittman,[b] James Lewis,[c]
Stephen T. Neely, and Patricia G. Stelmachowicz
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 8 January 2009; revised 29 April 2009; accepted 1 May 2009)

Probe-microphone measurements are a reliable method of verifying hearing-aid sound pressure level (SPL) in the ear canal for frequencies between 0.25 and 4 kHz. However, standing waves in the ear canal reduce the accuracy of these measurements above 4 kHz. Recent data suggest that speech information at frequencies up to 10 kHz may enhance speech perception, particularly for children. Incident and reflected components of a stimulus in the ear canal can be separated, allowing the use of forward (incident) pressure as a measure of stimulus level. Two experiments were conducted to determine if hearing-aid output in forward pressure provides valid estimates of *in-situ* sound level in the ear canal. In experiment 1, SPL measurements were obtained at the tympanic membrane and the medial end of an earmold in ten adults. While within-subject test-retest reliability was acceptable, measures near the tympanic membrane reduced the influence of standing waves for two of the ten participants. In experiment 2, forward pressure measurements were found to be unaffected by standing waves in the ear canal for frequencies up to 10 kHz. Implications for clinical assessment of amplification are discussed.
 [DOI: 10.1121/1.3143142]

PACS number(s): 43.25.Gf, 43.64.Ha, 43.66.Ts, 43.58.Vb [BLM] Pages: 15–24

I. INTRODUCTION

Measures of sound level in the ear canal are an integral part of clinical hearing-aid verification. Modern hearing-aid verification systems determine sound pressure level (SPL) in the ear canal using speech or speech-like stimuli in an effort to characterize the audibility of speech as a function of frequency for individual listeners. These measures account for the acoustic variability of individual ear canals. Probe-microphone measures of the hearing-aid response are preferable to behavioral verification methods because they provide better test-retest reliability (Hellstrom and Axelsson, 1993), a continuous representation of the frequency response instead of data only at discrete frequencies, and the ability to test infants or young children who are not able to participate reliably in behavioral assessments (Zemplenyi *et al.*, 1985).

Despite these advantages, interactions between the incident SPL in the ear canal and the acoustic reflections from the tympanic membrane (TM) limit the accuracy of these measurements for frequencies above 4 kHz. Specifically, partial cancellation of the acoustic signal occurs for frequencies with wavelengths less than four times the distance between the termination of the probe microphone and the TM. Such variations in SPL are therefore dependent on the position of the probe microphone relative to the TM (Gilman and Dirks, 1986; Chan and Geisler, 1990). Gilman and Dirks (1984) demonstrated that, when a probe microphone is placed at a fixed insertion depth, marked frequency-dependent differences in SPL will occur due to individual variations in ear-canal length. The purpose of the current study was to examine two different approaches to quantifying *in-situ* probe-microphone measurements at frequencies above 4 kHz.

Because many hearing aids have limited usable gain above 4 kHz (Boothroyd and Medwetsky, 1992), clinical recommendations regarding probe-tube insertion depth have not previously considered the potential influence of pressure minima in the ear canal at higher frequencies. Several investigators have recommended procedures to minimize the influence of standing waves at frequencies that are within the bandwidth of most hearing aids. Burkhard and Sachs (1977) recommended a probe-tube insertion depth of 5 mm past the termination of the earmold (EM). The authors argued that this depth would limit errors within the frequency range for most hearing aids and avoid acoustic irregularities, such as evanescent modes, stemming from the increase in diameter of the sound channel from the EM sound bore to the ear canal. Caldwell *et al.* (2006) measured consonant spectra and speech-weighted noise at 1, 5, and 10 mm past the termination of the EM and found that the 10 mm position provided the highest overall sound level, as well as the highest output at 6.3 and 8 kHz. Despite mean results showing that the sound level in the high frequencies is greatest at the deepest insertion depth, the impact of standing waves on individual probe-microphone measurements cannot be determined from the group data presented by Caldwell *et al.* (2006).

[a]Author to whom correspondence should be addressed. Electronic mail: mccreeryr@boystown.org
[b]Present address: Department of Speech and Hearing Science, P.O. Box 870102, Tempe, Arizona 85287.
[c]Present address: Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, Iowa 52242.

Current clinical recommendations have been developed to minimize cancellation from standing waves for frequencies below 4 kHz. However, investigators have demonstrated that hearing-aid bandwidth can be extended to higher frequencies. Amplification at frequencies as high as 16 kHz was demonstrated more than 25 years ago (Killion and Tillman, 1982). The current ANSI standard (ANSI, ANSI S3.22-2003) for determination of a hearing aid's frequency range (bandwidth) is determined by calculating the average full-on gain of the device at 1000, 1600, and 2500 Hz and drawing a line 20 dB below this average parallel to the abscissa. The lowest and highest intersecting frequencies represent the bandwidth of the device. This procedure tends to over-estimate the bandwidth available for a hearing-impaired listener because of insufficient gain to amplify the relatively low amplitude of speech energy at frequencies above 4 kHz. Moore *et al.* (2008) conducted a study to determine the gain required to make speech audible at frequencies up to 12.5 kHz for listeners with mild to moderate hearing losses. Despite limitations in hearing-aid gain and reduced speech energy at high frequencies, Moore *et al.* (2008) found that speech could be made audible at 10 kHz in approximately 40% of ears in their cohort. Although most current prescriptive formulas for hearing-aid gain do not provide targets for frequencies above 6 kHz, these studies suggest that making speech audible at frequencies up to 10 kHz is possible for some individuals with mild to moderate hearing loss.

Investigators have evaluated the effects of extended bandwidths on ratings of listener preference and sound quality. Ricketts and colleagues (2008) found a preference for the sound quality of a signal with 9 kHz bandwidth over a 5.5 kHz bandwidth for adults with normal hearing and those with hearing loss. Other investigators have suggested that rating of sound quality for music is greatest with an upper frequency of 16 kHz or greater, and that sound quality ratings for speech can suffer when the low-pass frequency cutoff is less than 10 kHz (Moore and Tan, 2003). Children, in particular, may experience additional improvements in speech recognition with bandwidths that exceed those of currently available devices. Stelmachowicz *et al.* (2001) systematically varied the cut-off frequency of a low-pass filter to assess the perception of /s/ for children and adults with normal hearing and hearing loss. Results suggested that for all listeners, perception of /s/ for female and child talkers reached maximum performance only when the upper bandwidth extended to 9 kHz. Further investigation (Moeller *et al.*, 2007) found that children with hearing loss were delayed in their acquisition of fricative sounds, even when amplification was provided at an early age. The authors concluded that these delays may be related to the limited bandwidth of hearing aids relative to the high-frequency gain necessary to achieve audibility for fricative sounds.

Recent advances in hearing-aid technology, which have resulted in devices with broader bandwidth, have created the need for reliable and valid probe-microphone measures at frequencies above 4 kHz. Theoretically, placement of the probe microphone at or in close proximity to the TM minimizes acoustic standing waves for frequencies up to 10 kHz because the distance between the probe microphone and TM would be less than the wavelengths for frequencies greater than 10 kHz (Dirks and Kincaid, 1987; Gilman and Dirks, 1986). However, the frequencies of standing waves in a closed ear canal are not determined solely by the distance between the probe microphone and the TM. Because the TM is not perpendicular at the termination of the ear canal, the distance between the probe microphone and the TM cannot be accurately represented by a single value. Consequently, the acoustic impedance at the eardrum has also been evaluated to account for the influence of acoustic reflections in the ear canal on estimates of *in-situ* sound level (Stinson *et al.*, 1982). Previous studies have also demonstrated large variations in SPL measurements taken at or near the TM at high frequencies (Dreisbach and Siegel, 2001; Khanna and Stinson, 1985). The sound level measured in the ear canal is a combination of the incident or forward sound presented to the ear as well as an outgoing reflected component. Therefore, accurate estimation of sound pressure in the ear canal should take into account impedance characteristics in order to estimate both forward and reflected pressure components.

The problem of standing waves in the ear canal is not unique to probe-microphone measures for hearing-aid verification. Acoustic standing waves have also been shown to affect *in-situ* calibration for evoking and measuring otoacoustic emissions. Variations in the length of the ear canal, depth of probe insertion, and reflectance characteristics of the TM can lead to significant differences in the level of the stimulus used to generate an emission. These challenges have led researchers to propose several approaches to improve the validity of sound level measurements in the ear canal. Farmer-Fedor and Rabbitt (2002) proposed a measure of *in-situ* sound level in a manner that differentiates the incident acoustic intensity in the ear canal from the outgoing reflected acoustic intensity. Their results suggested that such a method is more reliable than SPL and less likely to be affected by variations in the sound level related to reflectance. Scheperle *et al.* (2008) used a measure of forward pressure level (FPL) that also allows for an estimation of the incident acoustic pressure in the ear canal without the influence of reflected components that are responsible for acoustic standing waves. They compared *in-situ* calibration for distortion product otoacoustic emissions using SPL, sound intensity level (SIL) (Neely and Gorga, 1998), and FPL. Their results suggested that for *in-situ* calibration, SIL and FPL resulted in a more stable calibration and constant measure of sound level across frequency than SPL.

More recently, Withnell *et al.* (2009) found that behavioral thresholds expressed in incident pressure were not affected by standing waves when compared to *in-situ* measures of SPL. Because the acoustic impedance of the ear varies significantly across individuals, the amount of reflected sound energy will also lead to variability in the frequency and extent of standing waves, as well as in the sound transmitted to the middle ear. Estimates such as SIL and FPL, which account for the variability associated with incident and reflected components of the signal, seem well-suited to address the problem of quantifying sound level in the ear canal. These measures have not previously been applied to the measures typically used for hearing-aid verification.

The goal of the present study was to evaluate two different approaches to minimize the influence of acoustic standing waves on real-ear probe-microphone measurements for hearing-aid verification. In experiment 1, measurements at four probe-microphone placements in the ear canal were compared to determine the influence of standing waves at frequencies up to 10 kHz for adults with personal EMs. In experiment 2, behavioral threshold measures referenced to transducer voltage (dB re 1 μV), ear-canal SPL (ecSPL), and FPL in the ear canal were obtained in children to determine if a measure of incident pressure can provide a more accurate representation of input to the middle ear than ecSPL for frequencies above 4 kHz. It was hypothesized that the variation across frequency of FPL would be more similar to decibel referenced to transducer voltage than to ecSPL values, because FPL is not affected by acoustic standing waves in the ear canal. For the purposes of this discussion, ecSPL and FPL will be used to describe measurements taken at the medial end of the probe tube in the ear canal, while SPL and dB μV describe measurements referenced to coupler and voltage values, respectively.

II. METHOD: EXPERIMENT 1

A. Participants

Ten adults (five females and five males) age 25–54 years (mean=38.7 years) participated in experiment 1. All participants had normal middle-ear pressure ($\pm$50 daPa) and TM mobility as evidenced by 226 and 1000 Hz tympanograms (Tympstar, Version 2, GSI). None of the participants had a history of ear surgery. Otoscopic evaluation confirmed that none of the participants had excessive cerumen in the ear canal. Participants were selected for this study because they had a personal EM. Each EM included in the study was constructed of acrylic material. An equal number of right and left ears were included in the study. Three participants had EMs with parallel vents, while the remaining seven participants had unvented EMs. EM vents were occluded with adhesive putty to ensure that the acoustic effects of venting would be minimized.

B. Procedure

For each subject, ear-canal measurements were obtained in one ear using an Etymotic Research ER-7C probe-microphone system. Two Etymotic Research ER 7-14C probe-microphone tubes were glued together to maintain a constant 2 mm separation between the two probe-microphone locations. Measurements were taken at four positions in the ear canal: 2 mm past the sound bore of the EM (EM+2 mm), 4 mm past the sound bore of the EM (EM +4 mm), 2 mm distal to the TM (TM−2 mm), and at the TM (TM). A 5-s segment of broadband noise (70 dB SPL) was delivered to the ear canal via an Etymotic Research 2A earphone, which was coupled to the EM. The stimulus presentation and microphone recording were processed by a personal computer using a Digital Audio Laboratories CardDeluxe sound card. The sampling rate was 32 kHz, and the probe-microphone response was low-pass filtered at 10 kHz. All measurements were completed in a sound-treated booth.

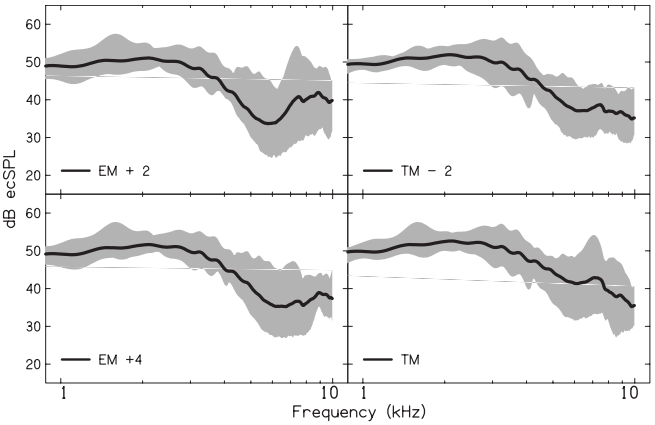


FIG. 1. Each panel displays ecSPL as a function of frequency for each probe-microphone placement. The solid black line is the mean across subjects, and the shaded region represents the range of ecSPL values measured at each frequency.

Results were obtained in a single session approximately 30 min in duration. Subjects were seated in a chair, and the probe-microphone tubes were placed at the eardrum for the first set of measurements. TM placement was verified using a tactile method in which the subject indicated when the probe tube touched the TM. The probe was then withdrawn slightly and secured to the inter-tragal notch using adhesive putty to prevent the probe tube from moving when the EM was placed in the ear. The EM was placed in the ear canal adjacent to the probe tubes for all subjects. In rare cases, acoustic measures revealed that placement of the EM collapsed the probe tube, resulting in a significantly reduced acoustic response. In these instances, the EM was removed and repositioned. Three repetitions of each measurement were made at the TM and at TM−2 without repositioning the probe microphone. The EM was then removed, the probe tubes were placed at EM+2 mm and EM+4 mm positions, and the procedure was repeated. The output files were saved and analyzed using custom software.

III. RESULTS: EXPERIMENT 1

A. Frequency and depth of pressure minima

Figure 1 shows the mean probe-microphone responses across subjects (in spectrum level) as a function of frequency at the four probe placements. The gray shaded area denotes the range of values across subjects. The largest range of responses across subjects occurred at frequencies above 2 kHz. Comparisons across the four probe placements were made for both the frequency of the minimum ecSPL (i.e., notch frequency) and the notch depth, defined as the difference between the maximum ecSPL below the notch and the ecSPL at the notch frequency. These calculations were analyzed to determine if placements at or near the TM resulted in a smaller notch depth or if the notch occurred at higher frequencies than for probe placements near the terminal end of the EM. Means and standard deviations for notch depths and notch frequencies for each placement are reported in Table I.

Analyses of variance (ANOVAs) were used to analyze differences between probe placements with notch frequency and notch depth as within-subjects factors. Notch depth is

TABLE I. Mean and standard deviation notch depth and frequency.

	EM+2	EM+4	TM−2	TM
Notch depth (dB)	23.8 (3.25)	22.04 (3.11)	21.19 (2.56)	19.7 (5.37)
Notch frequency (Hz)	6422 (1431)	7043 (1201)	7832 (1726)	8567 (1470)

plotted as a function of frequency for each probe placement in Fig. 2. Mean notch frequency increased significantly with proximity to the TM [$F_{3,27}=6.06$, $p=0.003$, $\eta^2_p=0.402$]. Post hoc tests using Bonferroni adjustments made for multiple comparisons ($p=0.0125$) revealed a significant difference between the TM and EM+2 conditions only. The depth of the notch did not vary significantly across probe placement [$F_{3,27}=3.01$, $p=0.15$, $\eta^2_p=0.104$]. The smallest individual notch depth at the TM position was greater than 10 dB, suggesting that clinically-significant notches were present in all participants even at the TM placement.

B. Test-retest reliability

Figure 3 displays the mean ecSPL differences across the three trials for each probe-tube placement, as well as a shaded area representing the range of ecSPL differences across trials. Below 4 kHz, the average difference in ecSPL was less than 2 dB for all four positions. Above 4 kHz, the average difference was less than 5 dB for all four positions. Variations greater than 10 dB were observed for each placement, except for the TM position, where the maximum variation across trials was approximately 8 dB. The maximum variation in ecSPL occurred between 4 and 10 kHz, which corresponds with the frequency range where significant pressure minima are most frequently observed.

IV. DISCUSSION: EXPERIMENT 1

Previous clinical recommendations for probe-tube insertion depth are based on minimizing the influence of acoustic pressure minima on *in-situ* probe-microphone measurements by placing the probe microphone in close proximity to the TM. However, the results from experiment 1 suggest that clinically-significant standing waves influence probe-microphone measurements even when the probe is placed in close proximity to the TM. Several participants had pressure minima at frequencies significantly lower than expected based on the distance between the probe microphone and the TM and the frequency corresponding with $\frac{1}{4}$ wavelength. All participants had marked pressure minima below 10 kHz and several participants showed pressure minima at the TM for frequencies within the bandwidth of current hearing aids. A probe-microphone response that decreases at higher frequencies could lead clinicians to make unnecessary adjustments to the response of a hearing aid, which could lead to patient discomfort and/or rejection of amplification.

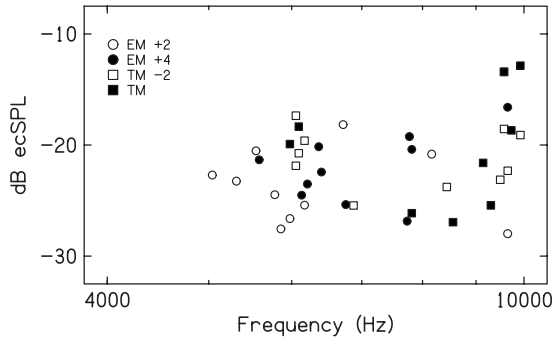


FIG. 2. Notch depth in ecSPL is plotted as a function of frequency for each participant. Probe placements are represented by symbols with circles for EM placements and squares for TM placements. Open symbols correspond with distal probe positions, while solid black symbols correspond with medial positions.

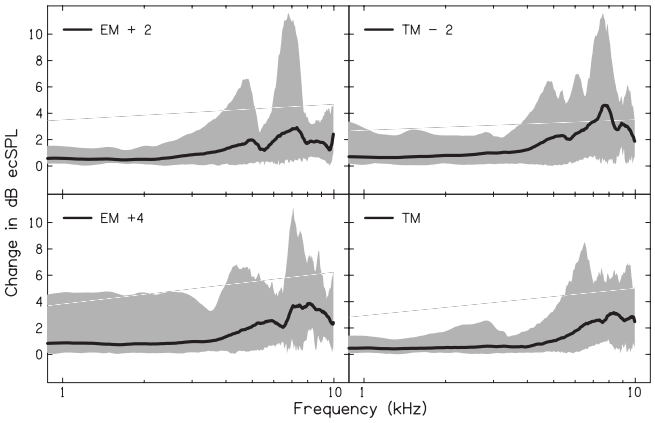


FIG. 3. The solid black line represents the mean differences in ecSPL across three trials as a function of frequency. The shaded area represents the minimum and maximum deviations observed at each frequency across three trials. Each panel represents a different probe placement.

Significant standing waves measured in close proximity to the TM in all participants raises concerns that the tactile method of probe placement may not have resulted in the desired probe insertion depth or that notches influence measurements near the TM. While some previous studies have revealed that the frequency of the pressure minimum in the ear canal is dependent on the distance between the probe and the TM (Gilman and Dirks, 1986), other studies where *in-situ* measurements of sound level were taken at or near the TM revealed significant pressure minima that would not be predicted based solely on this distance (Dreisbach and Siegel, 2001; Khanna and Stinson, 1985). Given significant individual differences in the impedance and reflectance between individuals at frequencies above 4 kHz (Voss and Allen, 1994; Stinson *et al.*, 1982), the fact that pressure minima cannot be easily predicted based on the measurement distance from the TM is not surprising. Even if probe-microphone placement at the TM provided an estimate of *in-situ* sound level that was not affected by standing waves, the potential for patient discomfort and difficulty in establishing and maintaining this position makes the clinical recommendation impractical.

Data from experiment 1 also suggest that the within-subject test-retest reliability for all probe positions was good. Below 4 kHz, the mean variability was less than 3 dB. Greater variability was present at higher frequencies where changes in probe position resulted in larger deviations across trials. As might be expected given the complex interaction between the incident and reflected acoustic energies for fre-

quencies above 4 kHz, the largest deviations across trials occurred in the frequency region of the notch. The amount of variability at frequencies >4 kHz across trials where attempts were made to maintain the position of the probe at the same location is indicative of the difficulty faced by clinicians in making consistent measurements at higher frequencies. While probe-microphone measurements have acceptable within-subject test-retest reliability, the presence of significant pressure minima for frequencies above 4 kHz suggests that ecSPL may not provide a valid measure of ear-canal sound level for hearing aids with wider bandwidths.

Distance between the probe and TM is one factor that influences the frequency of ecSPL pressure minima in the ear canal, but the oblique orientation of the TM at the end of the ear canal means that the probe distance from the TM is never a single value. Alternatively, the acoustic impedance of the sound source and ear canal may help to more accurately characterize *in-situ* sound levels. Scheperle *et al.* (2008) used a calibration method for otoacoustic emissions that measures the Thevenin-equivalent source impedance and pressure to isolate the incident and reflected components of the ear-canal response. Their results indicated that the ear-canal response expressed in FPL was less likely to contain substantial variations in pressure than measurements expressed in ecSPL. Therefore, experiment 2 was conducted to determine if FPL, which accounts for the acoustic impedance of both the transducer and ear canal, would provide a more accurate estimate of the sound input to the middle ear compared to traditional probe-microphone measures expressed in SPL.

V. METHOD: EXPERIMENT 2

A. Participants

16 normal-hearing children (8 males and 8 females), ages 9–15 years (mean=12.1), participated in the second experiment. This age range was selected to reduce the likelihood that behavioral thresholds were elevated at high frequencies due to noise-induced hearing loss. None of the participants had a history of ear surgery. Otoscopic evaluation was completed to confirm that participants did not have excessive cerumen in the ear canal or tympanostomy tubes.

B. Procedure

Data collection took place in a sound-treated room over two sessions approximately 30–45 min in length. At the first visit, hearing thresholds were measured in the test ear of each subject at 4, 6, 8, 9, and 10 kHz using Sennheiser HD 25-1 earphones. These results were compared to normative data collected from 32 children in a previous study using the same transducer (Stelmachowicz *et al.*, 2007). An automated method of limits with adaptive step-size was used until the standard deviation of the threshold response was less than or equal to 2.5 dB at each frequency. If a participant did not meet the inclusion criteria in one ear, evaluation of the other ear was attempted. Individuals with variations in behavioral thresholds were excluded to limit the presence of notches that were the result of threshold variability instead of standing waves. Four participants were excluded on the basis of hearing outside the normative range or variations in threshold >15 dB between adjacent frequencies, reducing the number of participants to 12. Following high-frequency audiometric testing, an impression of the ear canal was taken for the test ear of each participant in order to make a custom EM. Ear impressions were sent to a laboratory for fabrication. All EMs were a full-shell style constructed from clear, vinyl material and were tubed with No. 13 standard EM tubing with a 1.9 mm internal diameter and 3.6 mm external diameter. EMs were either unvented due to limitations in ear-canal size or vents were occluded with putty. The tubing of the EM extended to the termination of the sound bore at the canal and extended 2 cm from the lateral surface of the EM to simulate the approximate tubing length needed for a behind-the-ear (BTE) hearing aid. The result was an average transmission-line length of 34.17 mm (range 29–38 mm). Following receipt of the fabricated EM, participants returned for a second session approximately 2 weeks after the first session.

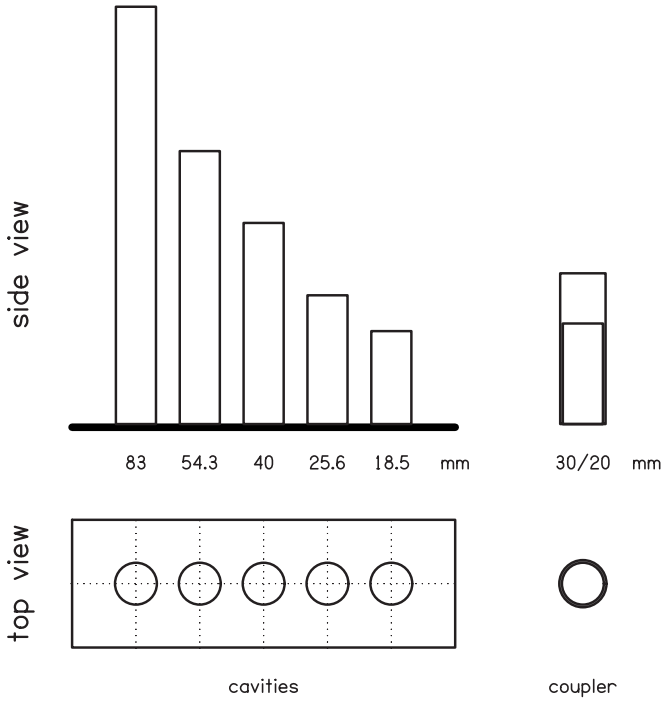


FIG. 4. Schematic of brass cavities and coupler used to obtain source calibrations in experiment 2.

Prior to testing, a calibration of the sound source was conducted with each participant's EM coupled to the transducer to determine the Thevenin-equivalent source impedance and pressure. The source calibration is similar to that used in previous studies to obtain separate estimates of incident and reflected sound levels in the ear canal (Allen, 1986; Keefe *et al.*, 1992; Scheperle *et al.*, 2008). Specifically, five brass tubes with lengths of 83.0, 54.3, 40.0, 25.6, and 18.5 mm were attached to a brass plate at one end. Tube lengths were selected to provide $\frac{1}{2}$ wave resonant peaks at 2, 3, 4, 6, and 8 kHz and allowed for accurate impedance measurements through 10 kHz. A diagram of the brass tubes and 30 mm brass coupler is shown in Fig. 4. The transducer for both sound source calibration and threshold measurements was an

Etymotic Research ER-2A insert earphone. The earphone tube was coupled directly to the EM tubing in the same manner used for a BTE hearing-aid coupling.

To record sound levels in the ear canal, a probe-microphone tube (ER-7C, Etymotic Research) was placed 4 mm past the termination of the EM sound bore either by threading the probe tube through a parallel pressure vent in the EM or by attaching the probe tube to the outside of the canal portion of the EM using 3M Transpore surgical tape. A distance of 4 mm was chosen to approximate clinical probe distances recommended in previous studies. Stimulus presentation and probe-microphone recording were processed by a personal computer using a Digital Audio Laboratories Card-Deluxe sound card. Ear-canal responses were recorded digitally with a sampling rate of 32 kHz and saved for later analysis.

First, the Thevenin-equivalent source characteristics were derived for each participant's EM. Each EM was attached to the brass coupler using adhesive putty. Broadband noise with a flat (voltage) spectrum and 256 ms duration was presented at 61 dB SPL (as measured in a 2 cc coupler) through the EM to each of the five cavities. Using custom software (EMAV, Neely and Liu, 1994), multiple measurements were taken and averaged in each brass tube to reduce the noise level. The obtained source impedance and frequency response for each brass tube were compared to the expected impedance and frequency response based on the length of the tubes, and an error was calculated based on the deviation from the expected response. If significant deviations between the measured and expected responses were obtained, the EM coupling was adjusted and measurements were repeated until the obtained output closely approximated the expected output for all five brass tubes.

Based on the acoustic response as measured at 4 mm past the termination of the sound bore in each of the brass tubes, a Thevenin-equivalent source pressure and impedance was derived for each individual's EM. Next, the same broadband noise was presented through the EM in the participant's ear to determine the pressure and load impedance characteristics of the ear canal at the same probe position. In rare cases, acoustic measures revealed that placement of the EM collapsed the probe tube, resulting in a significantly reduced acoustic response. In these instances, the EM was removed and repositioned. Custom software determined the frequency corresponding to the minimum ecSPL in each participant's ear canal (i.e., the notch frequency). Behavioral thresholds were then measured through the EM at octave frequencies from 0.5 to 8 kHz plus 9 and 10 kHz. To provide additional resolution in the frequency region of the notch, thresholds were also measured at the notch frequency and in $\frac{1}{4}$-octave steps from one octave below the notch to $\frac{1}{2}$ octave above the notch. An automated method of limits with a 5 dB step-size was used until the standard deviation of the threshold response was less than or equal to 2.5 dB at each frequency. Measurements of load impedance were used to derive the three estimates of sound level for each participant's threshold responses. Thresholds were expressed in (a) dB ecSPL as measured at the probe microphone, (b) voltage at the ER-2A transducer (dB re 1 μV), and (c) ear-canal dB FPL measured at the probe microphone, which was derived using load impedance of the ear canal in a manner similar to that used by Scheperle *et al.* (2008). The decibel relative to transducer voltage was used as a reference because auditory thresholds should not be influenced by local pressure minima resulting from interactions between incoming and outgoing sounds in the ear canal.

TABLE II. Notch frequencies and thresholds for each subject in ecSPL, FPL, and transducer voltage (dB re 1 μV).

Subject	Notch (Hz)	ecSPL	FPL	dB μV
1	8623	23.2	27.1	24.1
2	8289	−4.7	−0.5	16.5
3	7436	−9.8	4.4	19.8
4	7429	22.5	38.9	34.7
5	6965	−4.8	7.4	22.1
6	5907	−1.5	9.7	25.1
7	5773	−20	5	24.4
8	5591	−1.5	6.7	19.6
9	5152	7.1	19.4	29.7
10	3958	14.4	18.5	15.7
11	2933	22.5	22	15.6
12	2848	3.4	4.9	11.7
Mean	5908.67	4.23	13.63	21.59
Std	1942.65	13.98	11.61	6.57

VI. RESULTS: EXPERIMENT 2

The frequency of the minimum ecSPL (i.e., the *notch* frequency) and corresponding values for both transducer voltage and FPL are presented in Table II. Comparisons of sound levels at the notch frequency reveal that 10 of 12 subjects had FPL responses greater than ecSPL responses, while 2 subjects had no difference between FPL and ecSPL (S11 and S12). Figure 5 shows the mean threshold data in relative pressure as a function of frequency in ecSPL, FPL, and transducer voltage (dB μV) for all 12 subjects in experiment 2. The profile of threshold values was similar across all three estimates of sound level with two exceptions. First, at frequencies $\leq$2 kHz, the mean ecSPL response is greater than the mean ear-canal FPL response by approximately 6 dB. In the frequency region $>$2 kHz corresponding with the minimum ecSPL, FPL was more similar to dB μV than ecSPL. A repeated-measures ANOVA for frequencies $\leq$2 kHz indicated that the mean difference between ecSPL, FPL, and voltage was significant [$F_{2,80}=9.90$, $p=0.003$, $\eta^2_p=0.198$]. Post hoc tests using Bonferroni adjustments for multiple comparisons ($p<0.0167$) indicated that at low frequencies, SPL was significantly higher than FPL, with a mean difference of 5.5 dB. However, the 1.5 dB mean difference between dB voltage and ecSPL for low frequencies was not significant.

An additional repeated-measures ANOVA compared thresholds at the notch frequency. The mean differences between the three estimations of sound level at the notch (ecSPL, FPL, and dB μV) were significant [$F_{2,22}=19.183$, $p<0.001$, $\eta^2_p=0.636$]. Post hoc tests using Bonferroni adjustments for multiple comparisons ($p<0.0167$) indicated

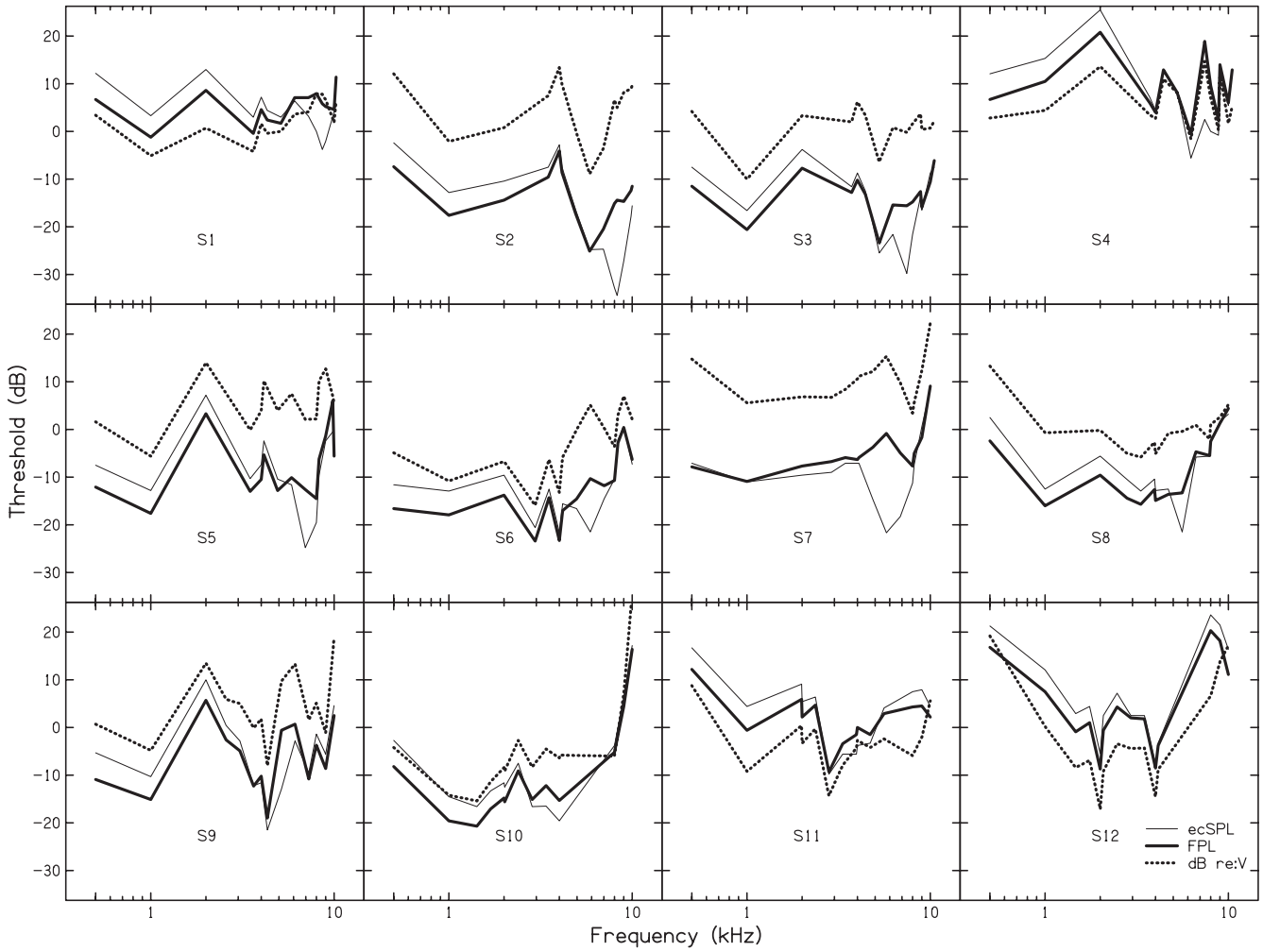


FIG. 5. Behavioral thresholds in relative pressure plotted across frequency for ecSPL (light-shaded line), FPL (dark-shaded line), and dB μV (dotted line). Each panel represents data from an individual subject in descending order by ecSPL notch frequency.

that the differences between all three measures were statistically significant with ecSPL resulting in the lowest estimate of sound level at the notch frequency and FPL resulting in a higher estimate of sound level than ecSPL. The mean difference between ecSPL and dB FPL at the notch frequency was 12.2 dB, indicating that FPL results in higher estimations of ear-canal sound level than ecSPL at the frequency where the influence of standing waves is significant. This finding suggests that thresholds referenced to FPL are less influenced by standing waves at the notch frequency than ecSPL. The variation in FPL near the notch frequency is similar to that of dB μV, which is assumed to be independent of measurement error from acoustic standing waves.

Figure 6 displays the mean threshold data across participants normalized to the notch frequency. To assess the relationship between ear-canal measures of ecSPL, FPL, and dB μV thresholds, a series of linear regression models was calculated and compared using R^2-change tests for nested models. Given that the pattern of the relationships between FPL, ecSPL, and dB μV was different at frequencies below 2000 Hz compared to frequencies above 2000 Hz, the regression models for comparison across the notch frequency were constrained to frequencies greater than 2000 Hz. The full regression model with dB μV as the criterion and both ec-

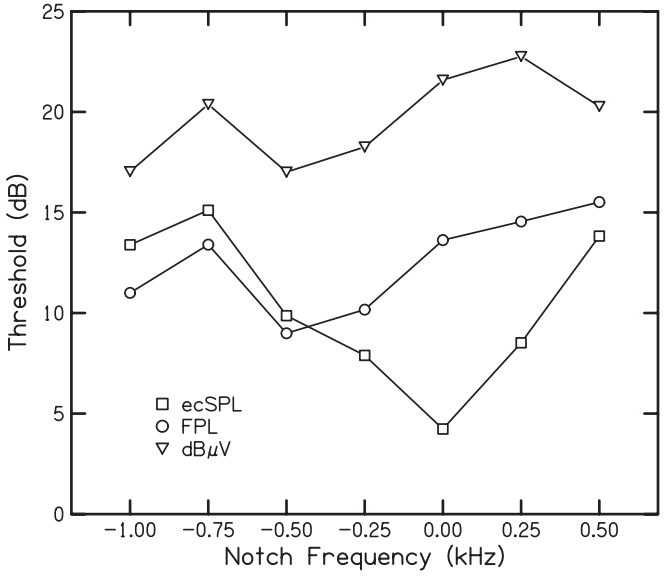


FIG. 6. Mean behavioral thresholds in relative pressure plotted as a function of the range from one octave below the notch frequency to one-half octave above the ecSPL notch frequency. Inverted triangles represent thresholds in dB μV, circles represent thresholds in FPL, and squares represent thresholds in ecSPL.

SPL and FPL as predictors had an $R^2=0.351$, [$F_{2,136}=36.731$, $p<0.001$]. Both ecSPL and FPL had significant regression weights at frequencies above 2000 Hz, but the standardized regression weight (β) for FPL was positive, while β for ecSPL was negative. The pattern of results suggests that as dB μV increases, FPL also increases for frequencies around the notch. Additionally, ecSPL decreases compared to dB μV for the range of frequencies around the notch, consistent with observed patterns of standings waves measured in the ear canal.

Nested comparisons were completed with ecSPL and FPL as individual predictors of dB μV for frequencies above 2000 Hz to determine if either FPL or ecSPL alone was able to predict thresholds in dB μV, as well as the full model that contained both estimates. The linear regression model with ecSPL as the predictor of dB μV threshold revealed an $R^2=0.168$ [$F_{1,137}=27.63$, $p<0.001$]. While ecSPL was found to have a significant relationship with dB μV at frequencies above 2000 Hz, the model with only SPL as the predictor did not predict thresholds in dB μV as well as the full model (R^2-change=−0.183, $F_{1,136}=38.31$ $p<0.001$) with significantly less variance accounted for. The linear regression model with FPL as a predictor of dB μV threshold had an $R^2=0.303$, [$F_{1,137}=59.483$, $p<0.001$]. Comparison of the FPL as the sole predictor of dB μV threshold was significantly better than the ecSPL-only model (R^2-change $=0.135$, $F_{1,137}=48.45$ $p<0.001$) and was not significantly different from the model with both ecSPL and FPL as predictors of dB μV threshold (R^2-change=−0.48, $F_{1,136}=10.23$ $p=0.15$).

VII. DISCUSSION: EXPERIMENT 2

Results from experiment 1 indicated that probe placement at or near the TM did not reduce the influence of pressure minima on probe-microphone measurements for any of the participants. These findings highlight the difficulty of using distance from the TM as a clinical guideline for probe-microphone measurements. Using a method described by Scheperle *et al.* (2008), FPL for behavioral threshold was calculated, in addition to voltage at the transducer and ecSPL. FPL is an estimate of ear-canal sound level that separates incoming and outgoing pressures in the ear canal when the impedance of the ear canal is known. The primary hypothesis of experiment 2, based on theoretical expectations, was that FPL would be less affected by standing waves than ecSPL, and more similar to the voltage at the transducer in the shape of its frequency response near the notch frequency.

Data from experiment 2 confirmed two distinct patterns of results for the relationship between the three estimates of sound level. At frequencies ≤2000 Hz, ear-canal measurements in ecSPL are higher than FPL measurements by approximately 6 dB. This pattern was statistically significant and observed consistently across all 12 subjects. Just as pressure minima at the notch frequency are the result of cancellation between incoming and outgoing sound pressures with opposing phases, the enhancement at frequencies ≤2000 Hz is potentially the result of summation of the two components in phase. The enhancement of ecSPL related to the summation of incoming and outgoing acoustic pressures occurs for frequencies ≤2000 Hz, which have wavelengths >6.75 in. Given the fact that these wavelengths exceed the length of the ear canal, this increase in ecSPL should be independent of the location at which measurements are taken in the ear canal.

At frequencies >2000 Hz, ear-canal measurements of threshold in ecSPL exhibit minima that are not evident in either FPL or dB μV. Substantial pressure minima in ecSPL, but not in FPL or dB μV, were observed in 10 of the 12 subjects. The two exceptions (S11 and S12) did not have significant differences between ecSPL and FPL, but did have an enhancement of ecSPL at frequencies ≤2000 Hz. An enhancement of ecSPL at lower frequencies should also be accompanied by a corresponding pressure minimum in each case. One explanation could be that both subjects had pressure minima at frequencies above 10 kHz, which were not measured in the current study. Both subjects also had minimum ecSPL values at frequencies close to 2 kHz, which was much lower than the other participants. Given that the relationship between each estimate of sound level (ecSPL and FPL) with dB μV is different for frequencies above and below 2 kHz, the close proximity of the ecSPL notch to 2 kHz may have resulted in no difference between the two estimates. Despite the lack of an apparent pressure minimum in each case, FPL and ecSPL were otherwise similar for both participants above 2000 Hz. This observation is an indication that FPL would not likely lead to additional errors in estimation of *in-situ* hearing-aid gain in cases where a SPL pressure minimum is not measured.

VIII. CONCLUSIONS

The purpose of the current experiments was to identify methods of minimizing the influence of acoustic standing waves in the ear canal on *in-situ* probe-microphone measurements, which are used clinically to estimate hearing-aid gain and output. Current methods are based on placing the probe-microphone tube close enough to the TM to limit the frequency of standing waves to frequencies above the bandwidth of commercially-available hearing aids (e.g., Burkhard and Sachs, 1977). Recent research suggests that there may be advantages to designing hearing aids with usable frequencies as high as 10 kHz to improve speech perception (Stelmachowicz *et al.*, 2001) and perception of sound quality (Moore and Tan, 2003). As manufacturers begin to extend the upper frequency limits of hearing aids, current probe-microphone measurement schemes are unlikely to provide valid measures of ear-canal sound levels. The goal of experiment 1 was to determine if probe placements close to the TM could reduce the influence of standing waves to frequencies approaching 10 kHz. Four probe-microphone placements at different positions relative to the EM and TM were used to measure the ear-canal response in ten subjects. Pressure minima ranging from 12 to 26 dB were present at all four probe placements, and even when the probe tube was placed near the TM, clinically-significant pressure minima were observed below 10 kHz. While the use of a tactile method of probe placement at the TM without visual confirmation might not have re-

sulted in insertion depths as close to the TM as desired, our finding of significant variations in ecSPL for measurements taken near the TM is consistent with previous studies (Dreisbach and Siegel, 2001; Khanna and Stinson, 1985). Although additional efforts to place the probe tube closer to the TM might reduce the influence of standing waves, the potential for patient discomfort and difficulty establishing and maintaining such a TM placement would still limit its clinical utility for hearing-aid verification.

The test-retest reliability of probe-microphone measurements at all positions for each subject was acceptable with differences less than 5 dB over most of the frequency range. More substantial errors were observed with the probe placements closest to the TM, where small changes in probe position are likely to have larger effects on the measured response. Larger deviations between repeated-measurements for the same probe position at frequencies above 4 kHz highlight the difficulty in maintaining consistency across measurements, even when care is taken to maintain a constant probe position. In addition to being potentially uncomfortable for patients and not feasible with children, the current data suggest that placement of the probe microphone at or near the TM does not appear to adequately minimize the impact of standing waves on estimates of *in-situ* sound level. While placement close to the TM resulted in a pressure minimum with an average frequency greater than 6 kHz, 18 of the 40 placements revealed significant pressure minima in the ecSPL at frequencies that would be within the bandwidth of most current hearing aids. This finding suggests that acoustic standing waves in the ear canal compromise the validity of probe-microphone measurements even for hearing aids with an upper bandwidth of 6 kHz. Errors in the estimation of ear-canal sound level of this magnitude may lead to inaccurate assignment of gain and output and could pose potential risk to residual hearing sensitivity.

In experiment 2, FPL was used to obtain an estimate of sound level in the ear canal that is independent of outgoing acoustic reflections from the TM. The hypothesis was that FPL would be more similar to decibel based on voltage at the transducer than ecSPL, since the decibel based on the transducer voltage should be theoretically independent of acoustic reflections. For frequencies >2000 Hz, FPL was found to be a better predictor of dB voltage than ecSPL. At frequencies ≤ 2000 Hz, however, ecSPL was found to be approximately 6 dB higher than FPL. The enhancement of ecSPL at frequencies ≤ 2000 Hz is most likely the result of summation between the incoming and reflected sound energy in the ear canal, which is a counterpart to the cancellation at frequencies >2000 Hz, which results in pressure minima in the estimation of the incoming signal. Because summation occurs at lower frequencies with $\frac{1}{4}$ wavelengths greater than the length of the ear canal, the 6 dB enhancement in ecSPL compared to FPL will be constant along the length of the ear canal and at the TM. Therefore, the 6 dB difference between ecSPL and FPL at low frequencies would only need to be taken into account in hearing-aid verification if probe-microphone measurements are taken in FPL and thresholds are referenced to ecSPL. Alternatively, cancellation has the ability to impact clinical decisions about the assignment of amplification. For higher frequencies where cancellation can result in significant pressure minima, the response of the hearing aid will appear to be inadequate. The frequency, gain, and output characteristics of the hearing aid may be adjusted unnecessarily by the clinician to compensate for what appears to be insufficient gain. The use of FPL for hearing-aid measurements could potentially result in smaller errors in estimating *in-situ* hearing-aid gain by taking into account the influence that reflected sound pressure has on *in-situ* measurements of hearing-aid gain and output.

ACKNOWLEDGMENTS

This study was supported by grants from the National Institute of Deafness and Other Communication Disorders (Grant Nos. R01 DC04300, R01 DC-008318, P30 DC-4662, and T35 DC-008757).

Allen, J. B. (**1986**). "Measurement of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.
ANSI (**2003**). ANSI S3.22-2003 Specification of Hearing Aid Characteristics (American National Standards Institute, New York).
Boothroyd, A., and Medwetsky, L. (**1992**). "Spectral distribution of /s/ and the frequency response of hearing aids," Ear Hear. **13**, 150–157.
Burkhard, M. D., and Sachs, R. M. (**1977**). "Sound pressure in insert earphone couplers and real ears," J. Speech Hear. Res. **20**, 799–807.
Caldwell, M., Souza, P. E., and Tremblay, K. L. (**2006**). "Effect of probe tube insertion depth on spectral measures of speech," Trends Amplif. **10**, 145–154.
Chan, J. C., and Geisler, C. D. (**1990**). "Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal," J. Acoust. Soc. Am. **87**, 1237–1247.
Dirks, D. D., and Kincaid, G. E. (**1987**). "Basic acoustic considerations of ear canal probe measurements," Ear Hear. **8**, 60S–67S.
Dreisbach, L. E., and Siegel, J. H. (**2001**). "Distortion-product otoacoustic emissions measured at high frequencies in humans," J. Acoust. Soc. Am. **110**, 2456–2469.
Farmer-Fedor, B. L., and Rabbitt, R. D. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.
Gilman, S., and Dirks, D. D. (**1984**). "A probe EM system for measuring eardrum SPL under hearing-aid conditions," Scand. Audiol. **13**, 15–22.
Gilman, S., and Dirks, D. D. (**1986**). "Acoustics of ear canal measurement of eardrum SPL in simulators," J. Acoust. Soc. Am. **80**, 783–793.
Hellstrom, P. A., and Axelsson, A. (**1993**). "Miniature microphone probe tube measurements in the external auditory canal," J. Acoust. Soc. Am. **93**, 907–919.
Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). "Method to measure acoustic impedance and reflection coefficient," J. Acoust. Soc. Am. **91**, 470–485.
Khanna, S. M., and Stinson, M. R. (**1985**). "Specification of the acoustical input to the ear at high frequencies," J. Acoust. Soc. Am. **77**, 577–589.
Killion, M. C., and Tillman, T. W. (**1982**). "Evaluation of high-fidelity hearing aids," J. Speech Hear. Res. **25**, 15–25.
Moeller, M. P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., *et al.* (**2007**). "Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I—Phonetic development," Ear Hear. **28**, 605–627.
Moore, B. C., Stone, M. A., Fullgrabe, C., Glasberg, B. R., and Puria, S. (**2008**). "Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss," Ear Hear. **29**, 907–922.
Moore, B. C., and Tan, C. T. (**2003**). "Perceived naturalness of spectrally distorted speech and music," J. Acoust. Soc. Am. **114**, 408–419.
Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.
Neely, S. T., and Liu, Z. (**1994**). "EMAV: Otoacoustic emission averager,"

Technical Memo No. 17 (Boys Town National Research Hospital, Omaha, NE).
Ricketts, T. A., Dittberner, A. B., and Johnson, E. E. (**2008**). "High-frequency amplification and sound quality in listeners with normal through moderate hearing loss," J. Speech Lang. Hear. Res. **51**, 160–172.
Scheperle, R. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2008**). "Influence of in situ, sound-level calibration on distortion-product otoacoustic emission variability," J. Acoust. Soc. Am. **124**, 288–300.
Stelmachowicz, P. G., Lewis, D. E., Choi, S., and Hoover, B. (**2007**). "Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children," Ear Hear. **28**, 483–494.
Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (**2001**). "Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults," J. Acoust. Soc. Am. **110**, 2183–2190.
Stinson, M. R., Shaw, E. A., and Lawton, B. W. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.
Voss, S. E., and Allen, J. B. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.
Withnell, R. H., Jeng, P. S., Waldvogel, K., Morgenstein, K., and Allen, J. B. (**2009**). "An in situ calibration for hearing thresholds," J. Acoust. Soc. Am. **125**, 1605–1611.
Zemplenyi, J., Dirks, D., and Gilman, S. (**1985**). "Probe-determined hearing-aid gain compared to functional and coupler gains," J. Speech Hear. Res. **28**, 394–404.

Impact of meteorological conditions on noise propagation from freeway corridors

N. C. Ovenden[a)]
Department of Mathematics, University College London, Gower Street, London WC1E 6BT, United Kingdom

S. R. Shaffer and H. J. S. Fernando
Department of Mechanical and Aerospace Engineering, Center for Environmental Fluid Dynamics, Arizona State University, Tempe AZ 85287-9809

(Received 20 November 2008; revised 1 April 2009; accepted 10 April 2009)

This paper examines the impact of meteorological conditions on the propagation of vehicular noise from urban freeways. A parabolic equation model coupled to an analytical Green's function solution close to the source field is used to compute the refracted sound field up to half a mile from the freeway to predict the noise exposure of residential areas nearby. The model was used in conjunction with meteorological and sound-level measurements taken at two freeway sites over the course of four days in Phoenix, AZ. From the data collected, three test cases of varying levels of atmospheric stratification and wind shear are presented and discussed. The model demonstrates that atmospheric effects are able to raise sound levels by 10–20 dB at significant distances away from the highway, causing violations of acceptable limits imposed by the Federal Highway Administration in residential areas that are normally in compliance.
 [DOI: 10.1121/1.3129125]

PACS number(s): 43.28.Gq [JWP] Pages: 25–35

I. INTRODUCTION

Noise pollution is a serious and worsening environmental concern in urban areas. Not only does it diminish the quality of human life[1–3] but it also alters wildlife habitats and breeding sites.[4] Highway traffic, airports, heavy industry, railways, and even leisure activities located close to built-up areas all contribute to the noise menace, and thus urban planners and managers pay close attention to mitigate it. This paper concerns a main contributor to noise pollution in urban areas—the freeway noise—which varies considerably in time and space in the proximity of roadways. The noise level therein depends on a myriad of factors, to name a few, the traffic speed and volume, vehicle type, ground conditions, terrain, sound barriers, atmospheric absorption and meteorological variables (e.g., temperature, wind velocity, and turbulence), and their spatial and temporal profiles.[5,6] While a majority of these factors are accounted for in operational sound prediction models, currently available models do not take all the salient factors into account.[7–9] For example, the latest version of the Federal Highway Administration's (FHWA) traffic noise model (TNM) Version 2.5 released in 2004 does not include the effects of temperature and wind variability; i.e., uniform, isothermal atmospheric conditions are assumed in the calculations. The latter is a reasonable assumption for shorter (less than 200 m) distances from the sound source, but errors can be substantial when predicting intermediate and far-field noise, since refraction of sound due to temperature and wind causes anomalous intensity variations at a significant distance from the source. For example, noise measurements conducted in Scottsdale, AZ, following complaints by residents living more than about 400 m ($\sim 1/4$ mile) from the East Loop 101 freeway, suggest that ground-level inversions (surface stable temperature stratifications) can increase the sound level by as much as 10–15 dB.[10] While the noise level therein under neutral atmospheric conditions is well within the FHWA noise abatement criterion (NAC), an inversion can cause the dBA level to exceed NAC. FHWA-NAC recommends implementing abatement procedures such as noise walls or modified pavement types (quiet pavements) when the energy averaged or equivalent sound-level (L_{eq}) approaches an A-weighted value of 67 dBA. The challenge, however, is accounting for inversions and wind shear.

The influence of atmospheric factors becomes particularly critical when noise mitigation is realized via a combination of techniques, for example, noise walls and quiet pavements. The Arizona Department of Transportation (ADOT) has received approval from the FHWA for the Quiet Pavement Pilot Program to investigate the usefulness of pavement surface type as a noise mitigation strategy, subject to the condition that Arizona would be a pilot program with specific research objectives and requirements.[11] ADOT would overlay Portland Cement Concrete Pavement in the Phoenix valley with a 1 in. thick asphalt rubber friction course (ARFC) surface. Where the ARFC is placed and noise walls are required, the walls may be reduced in height in view of the extra mitigation offered by ARFC surfacing.[11] Beginning in 2003, ADOT has been monitoring six sites across the Phoenix Metropolitan Area for traffic-generated noise to evaluate the effectiveness of ARFC. While measure-

[a)]Author to whom correspondence should be addressed. Electronic mail: nicko@math.ucl.ac.uk

ments show that ARFC has reduced freeway noise appreciably (8–10 dB) at close-in community locations, sound refraction due to environmental conditions can defeat the noise abatement approaches (e.g., the use of walls) at some distances away. Noise walls are expensive and typically cost ~$1M/mile, and hence merits of their installation should be carefully evaluated *a priori*.

The effectiveness of AFRC pavements, sound walls, and environmental factors become dominant only at certain intrinsic frequency ranges. The relationships between these variables and *A*-weighted noise levels in the field thus are intricate and can only be delineated via models that properly quantify fundamental relationships between sound levels and environmental factors. Such models will help in design, in interpretation of measurements taken at different positions/times, and in placing results on a unified scientific basis. For full forecasting, both the sound and environmental variables need to be predicted and their interaction should be quantified. A straightforward (yet onerous, because of computational expense) method is the nesting of an acoustic model with an environmental forecasting model. A simpler methodology is to use available representative atmospheric data from the area to feed the acoustic model, assuming local smaller scale variations are unimportant. The research reported herein is of this latter ilk and includes a meteorological measurement component. The aim is to examine how different meteorological conditions, especially ground based inversions and shear, can affect freeway noise, by taking Phoenix as a case of interest.[12,13]

A suite of computational approaches are presently being used for atmospheric sound propagation studies,[5] which include (i) Gaussian-beam methods, (ii) fast-field program models, and (iii) parabolic equation (PE) models. Ray theories, although robust for indoor acoustics, rapidly become highly cumbersome to compute in downward refracting media where many rays are needed and caustics are problematic. Additional complications, such as diffraction by obstacles, turbulence, and prediction of acoustic shadow regions, further urge the use of alternative methods. The key to PE models is the use of an effective sound speed based on temperature and wind speed of the actual mean flow field that both modify the isotropic adiabatic sound speed.[14,15] When assuming a line (or an axisymmetric) source, the two dimensional wave operator is factored into left- and right-traveling components transverse to the source. The unsteady pressure field due to a source can then be resolved by marching the computation across the domain away from the source by discounting any waves that propagate toward the source. Major disadvantages of this method are that it becomes inaccurate at high elevation angles and cannot directly account for back scatter unless the more difficult task of handling propagation in both directions is addressed. It has many advantages, however, including the ease of incorporating atmospheric absorption, varying boundary conditions, and geometries (e.g., complex terrain) along with actual spatially varying meteorological profiles. For these reasons, methodologies based on the PE prove highly popular,[16–22] although it is common practice to use hybrid models combining several methods to exploit features of the problem at hand in an attempt to circumvent potential caveats of any individual methods.[23–27]

In order to understand and quantify the effects of atmospheric temperature and velocity profiles on sound propagation, we have combined a field measurement campaign (Sec. II) with modeling efforts (Sec. III). The field measurements are to provide realistic vertical profiles of temperature and crosswind velocities to the model and were performed over 4 days at two freeway sites in Scottsdale, AZ and Mesa, AZ, where meteorological and sound data were taken and recorded over roughly a 6 h period between 6 a.m. and 12 p.m. (Sec. IV). For the modeling, the sound data are inputted into a Green's function model to evaluate the near source field generated from the freeway traffic (Sec. V). This source field along with the meteorological data is then inputted into a PE model to compute the refracted sound field out to a distance of 600 m. The results are compared to neutral atmospheric conditions, and the effect of stratification and wind shear is separated and quantified in three 20 min time-averaged cases selected from the field data (Sec. VI). The conclusions of the study are given in Sec. VII.

II. EXPERIMENTS

The experiments discussed herein were conducted by the Center for Environmental Fluid Dynamics at Arizona State University (EFD-ASU) in collaboration with ADOT and Illingworth & Rodkin, Inc. The EFD-ASU team made detailed measurements of atmospheric meteorological conditions, Illingworth & Rodkin Inc. provided sound measurements, and ADOT videotaped the traffic and recorded its speed.

The field experiments were conducted on October 10 and 11, 2006, in a location just on the west side of Phoenix loop 101 (ADOT location 3E, 33°30′05.95″ N 111°53′17.09″ W) and on November 7 and 8, 2006, just on the north side of Phoenix loop 202 (ADOT location 3D: 33°28′56.65″ N 111°45′48.16″ W). The details of motivation for site selection are outlined in Ref. 10. Although both sites are situated in urban locations, the freeways are relatively new so that housing and other buildings are located some distance (at least 0.5 km) away. Hence, the terrain neighboring in both freeways is relatively flat and homogeneous with hard sandy soil and sparse bushes. A cross section of the terrain for the route 202 site is shown in Fig. 1.

Measurements were taken from 7 a.m. to 11 a.m. in order to better understand how noise levels change during a period of a temperature inversion, typical daytime convective conditions, and during the morning transition period. It is interesting to note that the temperature conditions near the surface were found to be unstable even in the early morning hours, and this is believed to be due to retention of heat by the roadway surfaces even after the sunset, because of the high thermal capacity of road surfaces.

A number of instruments were employed, which included 3D sonic anemometers and a sound detection and ranging (SODAR) with radio acoustic sounding system (RASS). SODAR measures vertical wind profiles of all three components whereas RASS measures the vertical tempera-

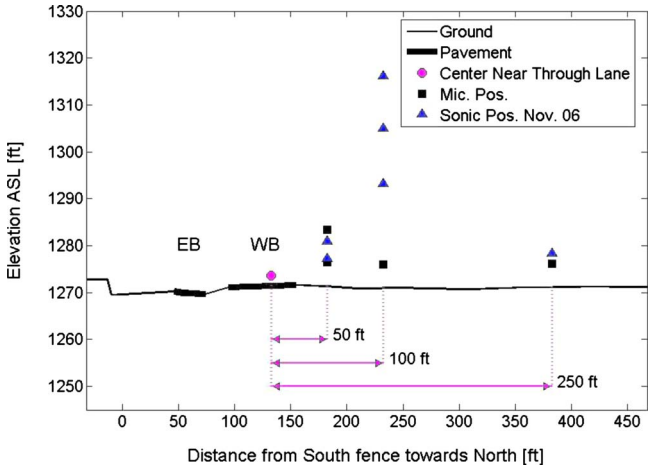


FIG. 1. (Color online) Cross section of Loop 202 site as elevation above sea level. Horizontal distance is measured in feet from the fence on the south side. Positions of instruments are shown as squares for microphones and triangles for sonic anemometers in the November 2006 field campaign. Arrows indicate distances from the center of the nearest travel lane (filled circle) on the West Bound side (WB).

ture profile. Sound measurement instruments at the 3E site were also located at the following positions, where range is the horizontal distance from the center of the nearest travel lane and height is measured above ground level (agl).

Location	Range	Height
1	15.24 m (50 ft)	1.52 m (5 ft)
2	15.24 m (50 ft)	3.66 m (12 ft)
3	30.48 m (100 ft)	1.52 m (5 ft)

Note at the 3D site there was an additional sound meter at 76.22 m (250 ft) from the center of the nearest travel lane at a height of 1.52 m (5 ft) agl. The sonic anemometers were located on towers at the same distance from the highway as the sound measurement instruments, while the SODAR/RASS systems were located further away to avoid possible contamination of the sound-level measurements (Fig. 1).

The sonic anemometers were operated at a frequency 10 Hz, providing all three velocity components and temperature. The data collected enable us to obtain detailed information on mean flow and temperature close to the surface, as well as turbulent statistics. During October 10 and 11 two sonic anemometers were located on a tripod 15.24 m (50 ft) from the center of the nearest travel lane and three on a tower 30.48 m (100 ft) from the center of the nearest travel lane. The heights of instruments on the tripod were 1.8 and 2.9 m agl and the height of those on the tower were 2, 4, and 6 m agl. During measurements on November 7 and 8, a tripod was located at 15.24 m (50 ft) where the heights of the sonics were 1.8 and 2.9 m agl, while sonics at the tower were placed at levels 6.8, 10.4, and 13.8 m agl. On November 8, one more sonic was placed on a tripod at a location 76.22 m (250 ft) from the center of the nearest travel lane at 2.2 m agl to measure atmospheric conditions close to the furthest sound measurement point.

The SODAR/RASS system was utilized to measure wind speed and temperature profiles, respectively, between roughly 20 to 600 m agl during the October and November deployments. This system provided more details on the structure of the atmospheric boundary layer during the periods of measurements at high altitudes, but for the present study the most important is the data taken up to 200 m or so in height. Both the SODAR/RASS and the sonic anemometers were set up to enable the wind velocity component across the highway to be separated from the wind velocity component parallel to the highway. Only the component perpendicular to the highway is inputted into the sound propagation model.

III. MODELING

Based on the initial sound data from the field experiments, we construct a two-dimensional model of acoustic propagation from a single monofrequency coherent line source in a vertically layered atmosphere. A rectangular xy coordinate system is used with y measuring the vertical height and x the horizontal range from the center of the nearest travel lane. All lengths are non-dimensionalized on a typical source height L_0, velocities are non-dimensionalized on the sound speed measured at ground level C_0, density is non-dimensionalized on the density of air at 1 atm ($\rho_0 = 1.2$ kg m^{-3}), and pressure p is non-dimensionalized on $\rho_0 C_0^2$. For a given frequency f (Hz), we define the Helmholtz number as $\omega = 2\pi f L_0 / C_0$ and by writing the acoustic pressure perturbation as $p(x,y,t) = p_c(x,y)e^{-i\omega t}$ the Helmholtz equation for a line source at $\mathbf{x} = \mathbf{x}_0$ of strength S in a vertically layered atmosphere is obtained:

$$\frac{\partial^2 p_c}{\partial x^2} + \frac{\omega^2}{\tilde{c}^2(y)} \frac{\partial}{\partial y}\left(\frac{\tilde{c}^2(y)}{\omega^2} \frac{\partial p_c}{\partial y} \right) + \frac{\omega^2}{\tilde{c}^2(y)} p_c = S\delta(\mathbf{x} - \mathbf{x}_0). \quad (1)$$

Here, $\tilde{c}$ is the non-dimensional *effective* sound speed which includes the effects of both temperature and crosswind. Given the measured vertical temperature $T(y)$ and crosswind $U_0(y)$ profiles, the effective sound speed is defined in a standard manner to be

$$\tilde{c}(y) = \frac{\sqrt{\gamma R T(y)} + U_0(y)}{C_0},$$

where γ is the ratio of specific heats and R is the ideal gas constant. The boundary conditions imposed are a far-field Sommerfeld radiation condition as $r = \sqrt{x^2 + y^2}$ becomes large, which takes the form

$$\frac{\partial p_c}{\partial r} - i\left(\frac{\omega}{\tilde{c}} \right) p_c = o(r^{-1/2}) \text{ and } p_c = \mathrm{O}(r^{-1/2}) \quad \text{as } r \to \infty, \quad (2)$$

and an impedance boundary condition at the ground surface

$$-\frac{1}{i\omega} \frac{\partial p_c}{\partial y} = \frac{1}{Z} p_c \quad \text{at } y = 0. \quad (3)$$

Throughout this paper, the empirical impedance model of Delany and Bazley[28] is used where, for a ground surface with flow resistivity σ (Pa s m^{-2}), the impedance Z is given by

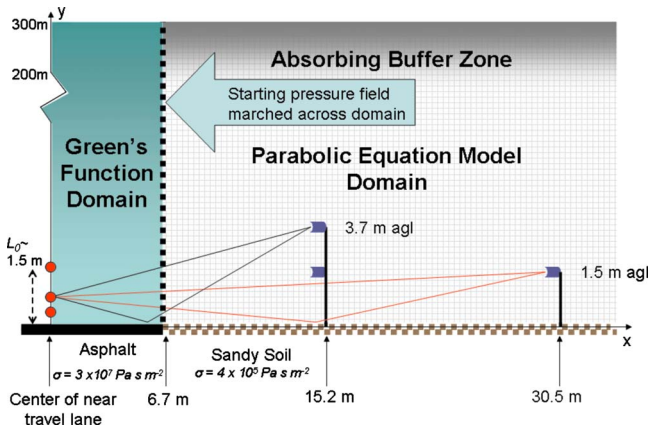


FIG. 2. (Color online) A schematic of the coupled models used to resolve the far-field propagation of traffic noise from a freeway corridor. The filled circles represent monofrequency coherent effective line sources above the centerline of the nearest lane of traffic. A Green's function method is utilized both to determine virtual source heights and strengths from the sound meter data and also to initialize the sound field along the vertical dashed line at the edge of the pavement. A PE model then marches this input pressure field across the domain, handling each frequency component separately.

$$Z = 1 + 0.0511\left(\frac{\sigma}{f}\right)^{0.75} + i\ 0.0768\left(\frac{\sigma}{f}\right)^{0.73}. \tag{4}$$

Two models are used in tandem to compute the far-field sound propagation: (i) a near-field analytic Green's function solution assuming a homogeneous atmosphere and (ii) a PE approximation. Figure 2 shows the regions of the x-y domain where each model is used.

The near-field Green's function solution[29] is used to obtain the acoustic field in the vicinity of the line source where the refractive effects of atmospheric factors are assumed to be negligible. In other words, the Green's function solution assumes a constant effective sound speed $\tilde{c}=1$ and solves Eqs. (1)–(3) with this assumption up to the edge of the highway, at 6.7 m (22 ft), obtaining the sound field

$$\frac{p_c(x,y;y_0)}{S} = -\frac{i}{4}H_0^{(1)}\left(\omega\sqrt{x^2+(y-y_0)^2}\right) - \frac{i}{4}H_0^{(1)}\left(\omega\sqrt{x^2+(y+y_0)^2}\right) + P_Z(x,y;y_0), \tag{5}$$

where $H_0^{(1)}$ is the zeroth order Hankel function of the first kind, and the term $P_Z(x,y;y_0)$ represents the correction to the hard-wall solution for Z finite. This correction is derived by Chandler-Wilde and Hothersall[29] and is given in terms of

$$\lambda = \omega\sqrt{x^2+(y+y_0)^2},$$

$$\gamma = (y+y_0)/\sqrt{x^2+(y+y_0)^2},$$

$$a_+ = 1 + \frac{\gamma}{Z} - (1-Z^{-2})^{1/2}(1-\gamma^2)^{1/2},$$

with the result

$$P_Z(x,y;y_0) = \frac{e^{i\lambda}}{\pi Z\sqrt{\lambda}}\int_0^\infty s^{-1/2}e^{-s}g(s/\lambda)ds + \frac{e^{i\lambda(1-a_+)}}{2\sqrt{Z^2-1}}\,\mathrm{erfc}\left(e^{-i\pi/4}\sqrt{\lambda a_+}\right),$$

where

$$g(t) = -\frac{[Z^{-1}+\gamma(1+it)]}{(t-2i)^{1/2}[t^2-2i(1+\gamma/Z)t-(Z^{-1}+\gamma)^2]} - \frac{e^{-i\pi/4}\sqrt{a_+}}{2(1-Z^{-2})^{1/2}(t-ia_+)}.$$

The first integral expression is calculated using Gauss–Laguerre quadrature and the second *surface wave* term (due to its strong exponential decay away from the ground) is evaluated using the formula given in Ref. 30. We assume over the near-field calculation that the ground impedance is typically of porous asphalt with $\sigma=3\times10^7$ Pa s m^{-2}, which is given in Table 4.9 of Attenborough *et al.*[6]

The near-field Green's function solution provides an acoustic field at the edge of the freeway $p_{\mathrm{ini}}(y)=p_c(x_{\mathrm{edge}},y)$, which is subsequently used as an initial condition for a two-dimensional Cartesian variant of the standard axisymmetric PE model, first derived by Gilbert and White.[14]

The PE model used is the parabolic wide-angle approximation of Eq. (1) assuming a two-dimensional line source. The pressure field is rewritten as $p_c(x,y)=\psi(x,y)e^{i\omega x}$ and $\psi(x,y)$ is obtained by solving the equation

$$\frac{\partial\psi}{\partial x} + \frac{1}{4\omega^2}\left[\frac{1}{\tilde{c}^2}\frac{\partial}{\partial y}\left(\tilde{c}^2\frac{\partial^2\psi}{\partial y\partial x}\right) + \omega^2\left(\frac{1}{\tilde{c}^2}-1\right)\frac{\partial\psi}{\partial x}\right] = \frac{i}{2\omega}\left[\frac{1}{\tilde{c}^2}\frac{\partial}{\partial y}\left(\tilde{c}^2\frac{\partial\psi}{\partial y}\right) + \omega^2\left(\frac{1}{\tilde{c}^2}-1\right)\psi\right]. \tag{6}$$

Equation (6) and the impedance boundary condition (3) are finite-differenced and the solution is obtained by marching forward in the x direction. Sandy soil is taken to be the ground surface type beyond the freeway with $\sigma=4\times10^5$ Pa s m^{-2} and we assume the ground to be completely flat to concentrate strictly on atmospheric effects in this study.

The radiation condition (2) is dealt with numerically by a buffer zone[15,31,32] occupying approximately the upper one-third (100 m) of the grid domain, $y_{\mathrm{att}}<y<y_{\max}$, where the effective sound speed $\tilde{c}$ in Eq. (6) is replaced by

$$\bar{c}(y) = \tilde{c}(y)\left[1 + iA\left(\frac{y-y_{\mathrm{att}}}{y_{\max}-y_{\mathrm{att}}}\right)^3\right]^{-1}.$$

Here, A is a real parameter that can be optimized for each frequency component. To ensure the effectiveness of the buffer zone, the initial pressure profile obtained from the near-field Green's function method $p_{\mathrm{ini}}(y)$ must also be smoothly reduced to zero within the buffer zone to prevent spurious reflections from the truncated top of the grid domain. Thus,

$$\psi(x_{\text{edge}},y) = p_{\text{ini}} \exp\left(-\frac{B\omega^2}{2}\left(\frac{y-y_{\text{att}}}{y_{\max}-y_{\text{att}}}\right)^2 - i\omega x_{\text{edge}}\right),$$

where $1 \leqslant B \leqslant 4$ is another optimized parameter dependent on frequency.

Effects of atmospheric absorption are additionally incorporated following the method outlined in Ref. 5 (Sec. B.5) by applying an attenuation rate dependent on the local humidity, temperature, and atmospheric pressure in dB m^{-1} to each frequency band at 1 m agl before summing to form the L_{eq} versus range plots (Figs. 8, 10, and 12). This method follows the International Standard ISO 9613-1:1993(E). The attenuation rate used here is based on a relative humidity of 20%, which is typical for the city of Phoenix, and the temperature profiles obtained from the measurements taken. The pressure in the absorption calculation is assumed to be 101.325 kPa.

IV. CHOSEN TEST CASES AND MODELING PARAMETERS

Based on the large amount of meteorological and sound data collected, three test cases are presented here. To exclusively illustrate the strong dependence of acoustic properties on environmental conditions, the main focus will be on one site (Rt. 202). Temperature and crosswind profiles above 40 m are obtained from the SODAR/RASS measurements in 10 m increments, whereas data at lower altitudes are gleaned from the sonic anemometers. The meteorological profiles are time-averaged over a period of 20 min. To obtain the surface-layer velocity profile for an unstable convective boundary layer (below 60 m), theoretical curves of the Monin–Obukhov (MO) stability theory are fitted to the sonic data. The MO theory suggests that near the ground both vertical temperature and velocity gradients have the form

$$\frac{\partial \zeta}{\partial y} \sim A_\zeta (1 - B_\zeta y)^{2/3} y^{-4/3} \quad \text{for } \zeta = U_0(y),\ T(y), \tag{7}$$

where A_ζ and B_ζ are parameters fitted to the data.[33] Since $\partial\zeta/\partial y$ diverges like $y^{-4/3}$ as $y \to 0$, the chosen temperature profile is made linear near the ground so that $T(y) \sim Ay + B$ and the velocity takes instead a standard logarithmic form, $U_0(y) \sim A \log(y/y^*)$, where y^* is the aerodynamic roughness length. Above 60 m the theoretical curve smoothly transitions into the SODAR-RASS data. If the useful range of data from the SODAR-RASS is less than 300 m, the theoretical curve is held constant at the last entry from the SODAR-RASS. Measurements and theoretical profiles for the three chosen cases are shown in Fig. 3. The following representative cases were selected for study:

A November 7, 2006 (Rt 202) 11 a.m. Wind shear at very high altitudes but little temperature stratification. Note that in this case, the SODAR-RASS data were usable up to 250 m compared with 200 m in the other cases.

B November 7, 2006 (Rt 202) 8 a.m. Significant stratification and shear flow.

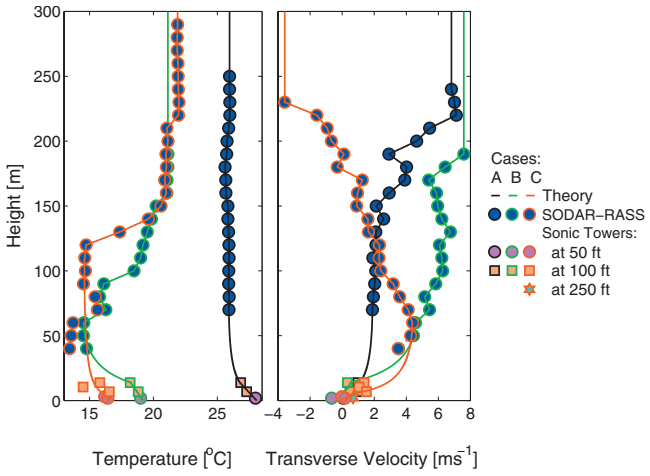


FIG. 3. Temperature and crosswind (to the freeway) data with fitted theoretical profiles for the three cases. All data points above 20 m are given by the SODAR-RASS with lower height information obtained from the sonic anemometers as shown in the legend.

C November 8, 2006 (Rt 202) 8 a.m. Strongly stratified with a sharp change in temperature at approximately 120 m above the ground and a crosswind jet at approximately 50 m above the ground.

V. ANALYSIS OF TRAFFIC SPECTRA TAKEN BY NOISE METERS

The overall acoustic source field we are attempting to replicate consists of a six-lane highway (three lanes in each direction) with multiple moving sound sources that vary according to their speed, traffic density, and vehicular type. Without knowledge of the exact acoustic signature of every car and truck, a number of severe but unavoidable assumptions need to be made about the nature of the sound sources. We emphasize here that the focus of this paper is on the meteorological aspect of noise transmission from freeways as opposed to understanding the composition of sound sources emitted.

Our sound data consist of 5 min time-averaged 1/3 octave data from three sound meters placed close to the highway. We have no information about the sound generated from separate lanes of traffic or the frequency output of different vehicle types traveling at different speeds. Hence, the principal aim of our model noise source must be to generate a *representative* sound field that matches the three sound meter measurements taken at the site. Figure 4 shows the difference between the 5 min averaged dBA level taken from the sound meter at 1.5 m (5 ft) above the ground and located 15.2 m (50 ft) away from the center of the nearest travel lane and the sound meter at 1.5 m (5 ft) above the ground and located 30.5 m (100 ft) away from the center of the nearest travel lane. This clearly shows a geometric attenuation of 3 dB as the distance from the source doubles, providing justification to the assumption that the freeway can be treated as series of line sources; we assume that this holds true for the entire study domain. Note that because the SODAR-RASS meteorological data are only available in 20 min time-averaged periods, the 5 min time-averaged acoustic data $[L_A^{5\ \text{min}}(f_n)$ in A-weighted decibels (dBA) for each frequency

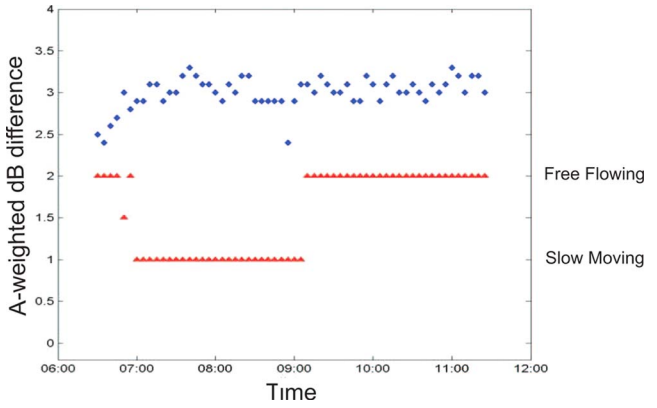


FIG. 4. (Color online) The difference in overall A-weighted sound level on October 11, 2006, measured between the sound meter located 15.2 m (50 ft) from the center of the nearest lane of traffic at a height of 1.5 m (5 ft) and the sound meter located 30.5 m (100 ft) from the center of the nearest lane of traffic at a height of 1.5 m (5 ft). The triangles merely display an indication of the traffic conditions at the time (either free flowing, slow moving or intermediate). A decrease of 3 dB with a doubling of distance corresponds to what is expected for a line source as $P_{\text{line}} \sim r^{-1}$ in a neutral atmosphere.

f_n] are also combined into 20 min average values for consistency in the long-range noise propagation model. This is achieved by extracting four consecutive 5 min time-averaged sound measurements, $[L_A^{5\ \text{min}}(f_n)]_k$ for $k=1$, 2, 3, and 4, and then averaging using the following formula:

$$L_A(f_n) = 10\log_{10}\left(\frac{1}{4}\sum_{k=1}^{4} 10^{[L_A^{5\ \text{min}}(f_n)]_k/10}\right).$$

For the purpose of our model, the traffic noise is approximated as a series of monofrequency coherent line sources positioned vertically above the center of the nearest travel lane (see Fig. 2). The strength and effective height of these *virtual* sources are unknowns that are determined by replicating as closely as possible the 1/3 octave data obtained from the three closest sound meters. As the sound meters are positioned relatively close to the source, the influence of meteorological conditions is regarded as negligible over the range up to the furthest sound meter, and a neutral atmosphere is therefore assumed in the near field. This enables the unknown line source parameters to be determined by using the Green's function model for acoustic propagation from a line source above an impedance plane as detailed in Sec. III. As mentioned before, the flow resistivities chosen are $\sigma=3\times10^7$ Pa s m^{-2} for the asphalt and $\sigma=4\times10^5$ Pa s m^{-2} for the sandy soil where, for the loop 202 experimental site (Fig. 1), the surface is assumed to be asphalt out to a range of 6.7 m (22 ft) from the virtual line sources with sandy soil beyond, as shown in Fig. 2. Repeating the calculation for other flow resistivities suggests that neither representating the asphalt as a hard wall ($Z=\infty$) nor varying the sandy soil flow resistivity between 2×10^2 and 6×10^5 Pa s m^{-2} changes the results significantly.

For a given 1/3 octave interval, the height of a representative line source can be calculated by accurately trying to replicate the differences between the dBA values recorded by the three sound meters. This is done by varying the source height to minimize the sum of the absolute errors between

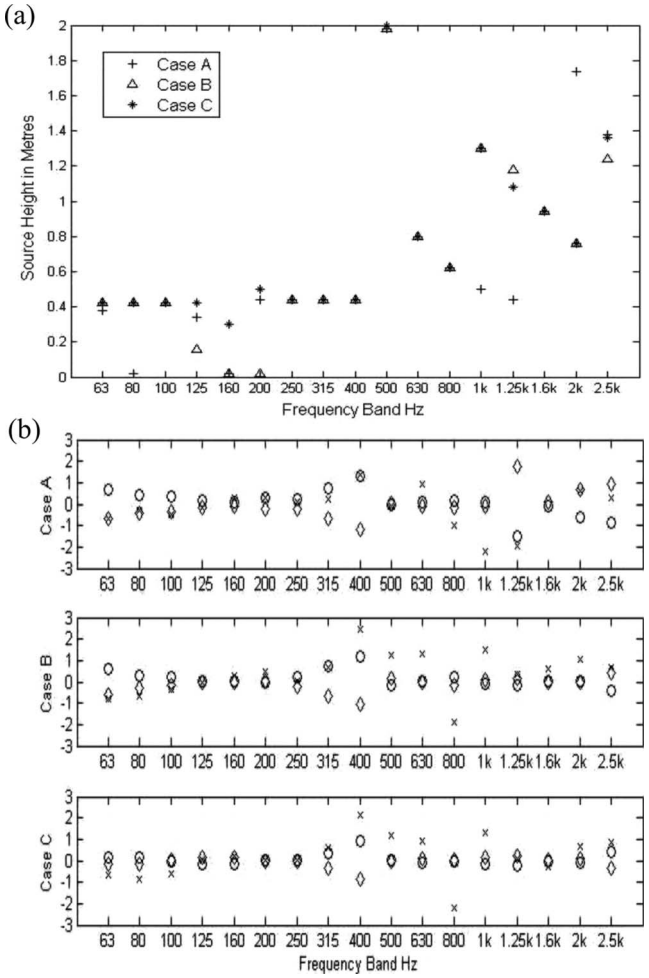


FIG. 5. (a) Virtual source heights for the three cases obtained by minimizing an error norm based on decibel differences between sound meters. (b) Measured dBA minus the dBA obtained from the Green's function solution for each virtual line source at the three sound meter locations. Circles show measured minus computed dBA for the meter at location 1, diamonds for the meter at location 2, and crosses for the meter at location 3.

the differences obtained by the Green's function model in the meter locations and the actual measured differences. However, this can lead to unrealistic virtual source heights and so an additional height penalty is also imposed. A norm based on this premise can be obtained by defining ΔM_{ij}^f to be the dBA difference actually recorded at frequency band f between meter locations i and j (as numbered in Sec. II) and $\Delta G_{ij}^f(H)$ to be the dBA difference obtained between meter locations i and j from a virtual monofrequency coherent line source at frequency f and at height H above the surface. Our virtual source height H is then determined by minimizing

$$\sum_{i<j} |\Delta M_{ij}^f - \Delta G_{ij}^f(H)| + 3H,$$

where the term $3H$ represents the additional height penalty mentioned above. Once the height is determined, the source strength can be obtained by averaging the source strengths required to reproduce the three meter readings. The source heights calculated for each frequency band in Cases A–C are plotted in Fig. 5(a) and Fig. 5(b) shows for all three cases the difference between the dBA measured at each sound meter and that determined from the virtual source obtained through

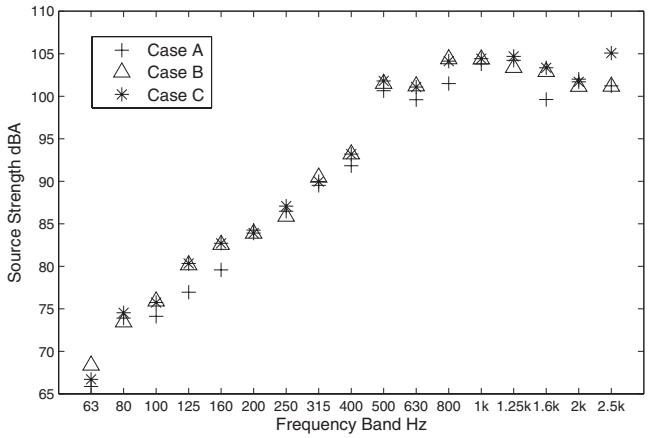


FIG. 6. Virtual source strengths for the three cases obtained by minimising an error norm based on dB differences between sound meters.

the optimization process described above. Observe that the virtual line sources replicate the measured sound field accurately to well within 1 dBA error for most frequencies and meter readings. Note additionally that there is generally very good agreement on the source heights obtained in each case for all frequencies (using data taken on different days at different times) although three obvious exceptions are the significant virtual source height differences for the 1, 1.25, and 2 kHz components between Case A and the other cases. We point out, however, that local norm minima are obtained for Case A at approximately the same heights as the virtual heights obtained for the two other cases but these are not optimal with the chosen norm. Other small discrepancies in Fig. 5(b) for the lower frequencies can be explained as the dBA difference errors do not vary that much with height due to the large wavelengths so that variations of source height do not significantly alter spatially the sound field generated. Perhaps, in fact, the most problematic difficulty in selecting source height here occurs around 400–500 Hz range where the norm error at small heights takes unacceptably high values (possibly 7 dB) but only approaches zero again at source heights of well over 2 m or so. This can be observed in the increase in measured minus computed errors at around 400 Hz and providing some justification for imposing a height penalty.

Following the determination of source heights, it is relatively straightforward to use the Green's function near-field model to obtain the *A*-weighted source strengths and these are shown for Cases A–C in Fig. 6. Note the good agreement in the source strength profile across the frequency ranges 63–2.5 kHz for the three cases. The sound signature is almost identical for Cases B and C, both taken at the same time during rush hour on consecutive days, whereas Case A has lower sound levels particularly in the 100–200 Hz and 800 Hz–2 kHz band, possibly due to the lower traffic levels occurring in the late morning.

VI. CONSTRUCTION OF L_{eq} PLOTS

In each chosen case, the model is run for each frequency component, based on the central frequency of the 1/3 octave band, with and without the influence of meteorological effects for comparison. For efficiency, the frequency range of the computation is reduced from spanning the entire range of 25 Hz–20 kHz to only include those bands between 63 Hz and 2.5 kHz (17 components in all). Such a restriction, due to *A*-weighting, produces an error of less than 0.2% in terms of the final overall sound pressure level (SPL) when compared to the actual values measured by the sound meters.

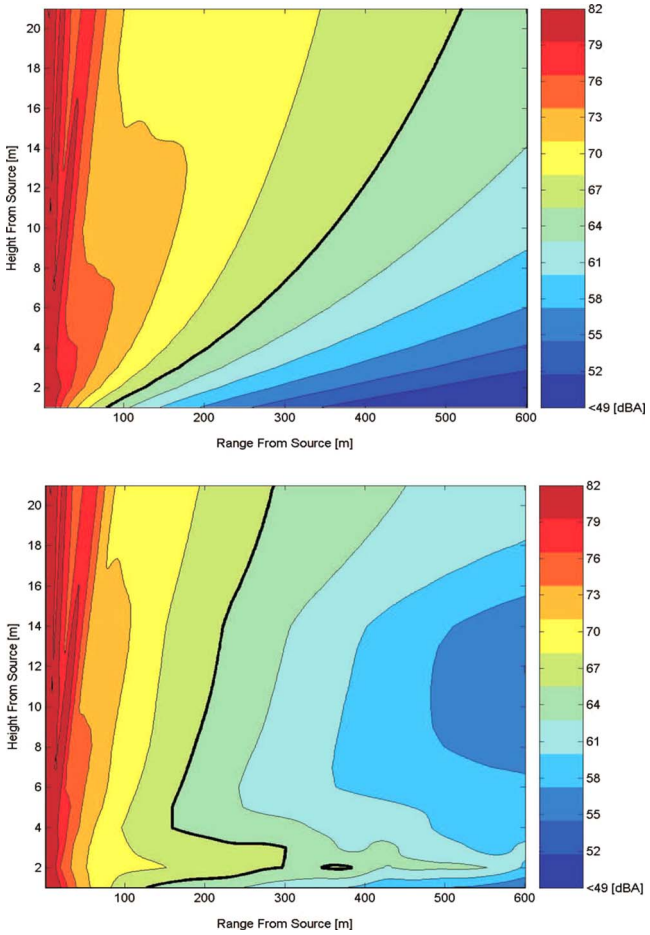


FIG. 7. (Color online) Case A: *A*-weighted SPL contours without meteorological effects (top) and with meteorological effects (bottom). The effect of atmospheric absorption is not included here. Each contour line represents a change of 3 dBA. The bold contour represents the 67 dBA level.

The spatial *A*-weighted SPL distribution for each frequency component is resolved by the PE model on a grid of size and spacing dependent on the wavelength (based on a usual ten grid points per wavelength). These results are subsequently interpolated onto a grid of 1 m spacing with a range of 0–600 m horizontally and 0–300 m vertically. Then at each grid point the *A*-weighted frequency contributions $L_A(f_n)(x,y)$ are combined to produce the overall L_{eq} SPL level by the formula

$$L_{eq} = 10 \log_{10} \sum_{n=1}^{17} 10^{L_A(f_n)/10},$$

with f_n={63, 80, 100, 125, 160, 200, 250, 315, 400, 500, 630, 800, 1000, 1250, 1600, 2000, 2500}.

Results of the spatial SPLs are presented in Figs. 7 and 8 for Case A, Figs. 9 and 10 for Case B, and Figs. 11 and 12 for Case C. For each case, Figs. 7, 9, and 11 show a contour

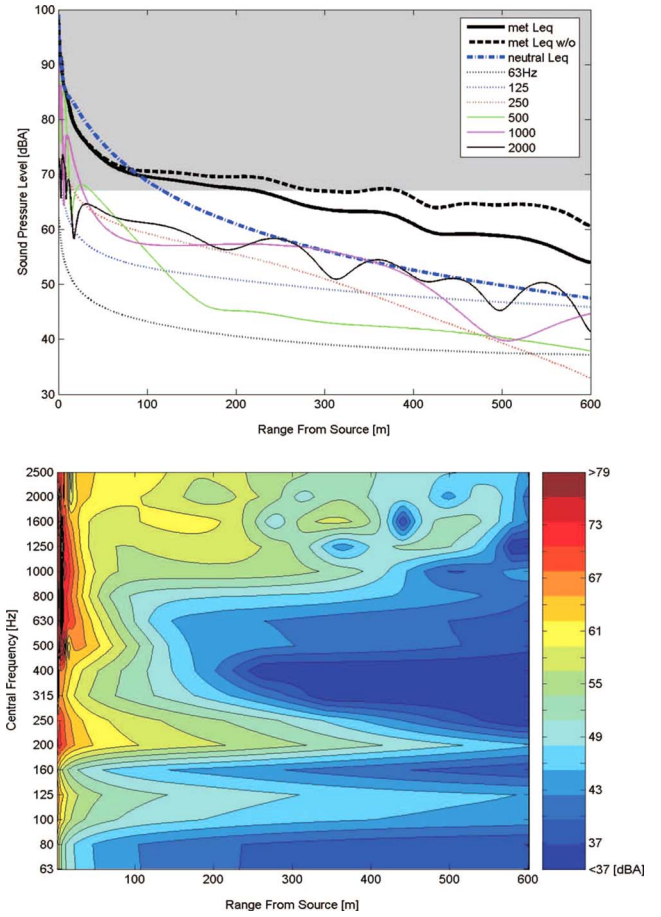


FIG. 8. (Color online) Case A: Overall *A*-weighted SPL and the SPL of each frequency component at a height of 1 m above the ground. The top figure shows the SPL for neutral conditions (bold blue dash-dot line), with meteorological effects but without atmospheric absorption (bold black dashed line) and with both meteorological effects and with atmospheric absorption (bold black solid line and frequency bands). The shaded area in the top figure represents the region where the SPL range exceeds the 67 dBA threshold. The bottom figure shows contours of *A*-weighted SPL with meteorological effects for each frequency component at an altitude of 1 m with atmospheric absorption. Each contour line represents a change of 3 dBA.

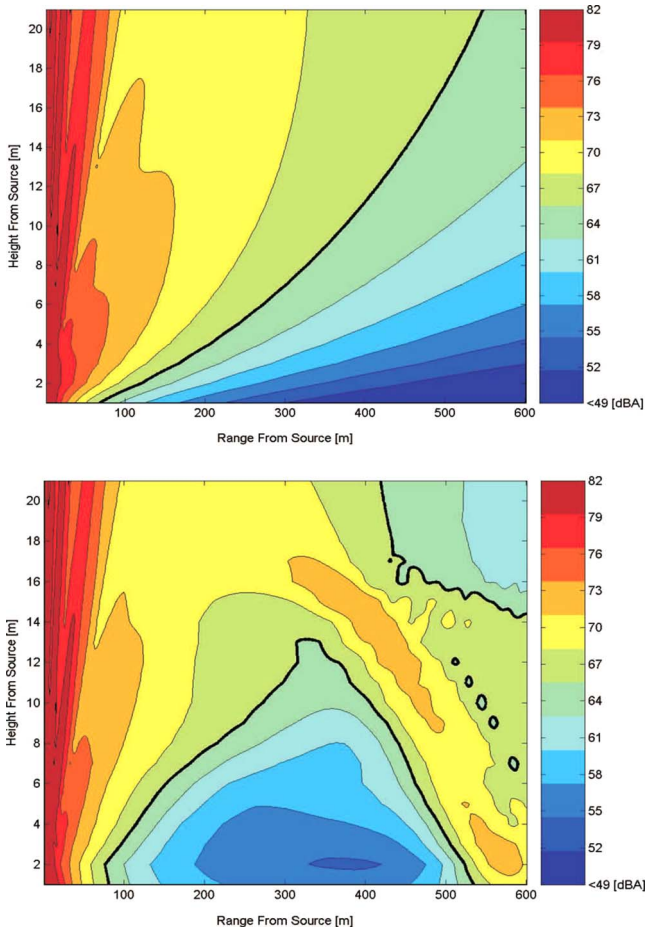


FIG. 9. (Color online) Same caption as for Fig. 7, but for Case B.

plot of the equivalent spatial sound field obtained in a neutral atmosphere directly above the contour plot of the resolved spatial sound field when the temperature and crosswind velocity effects are included. Note that the downwind side of the freeway is always shown and the vertical range displayed is only up to 20 m in height agl. It is clear from these figures that the overall impact of the meteorological effects is significant in all three cases examined. Indeed, significantly higher noise levels are predicted near the ground downwind for all cases. For guidance, FHWA's NAC threshold of 67 dBA is shown as a thick contour line on the spatial contour plots of L_{eq} (Figs. 7, 9, and 11). Furthermore, on the range plots (the top plot in Figs. 8, 10, and 12), the shaded area also represents SPLs exceeding the 67 dBA threshold. Below each case is examined in more detail.

The meteorological effects are weakest for Case A with little temperature stratification and a crosswind of the order of 2 ms^{-1} persisting from about 30 m to around 150 m in altitude. However, Fig. 7 clearly shows how the crosswind shear flow present up to 30 m above the surface focuses sound into a thin layer of 3–5 m in height, where the sound intensity is raised by roughly 10–15 dB. As a result, the sound level close to the ground does not fall below 67 dBA until a horizontal distance of approximately 220 m from the freeway is reached, as opposed to approximately 110 m predicted for a neutral atmosphere (see Fig. 8). A close examination of the impact of the meteorological effects on individual frequency components (Fig. 8, bottom) reveals that the frequency bands 200–250 Hz and 1–2.5 kHz remain the most intense out to the far-field.

Case B occurred during rush-hour traffic on Loop 202 with a stronger wind shear from the ground attaining a crosswind speed of 6 ms^{-1} at 60 m in height. More severe temperature gradients can also be observed, with the temperature falling 5° with increasing altitude before rising back to its ground-level value at an altitude of 100 m. The competition between the near-ground negative temperature gradient and positive wind shear means that overall near-ground sound levels drop more rapidly than they would in neutral atmospheric conditions over the first 150 m or so from the freeway. However, the refractive effects due to wind shear and to the presence of a temperature inversion at higher altitudes lead to sound rays being refracted back toward the ground from above and sound focusing at around 550 m from the freeway. Indeed, Figs. 9 and 10 indicate that the *A*-weighted SPL starts to exceed the 67 dBA threshold close to the ground at a range of 500 m before continuing to exceed

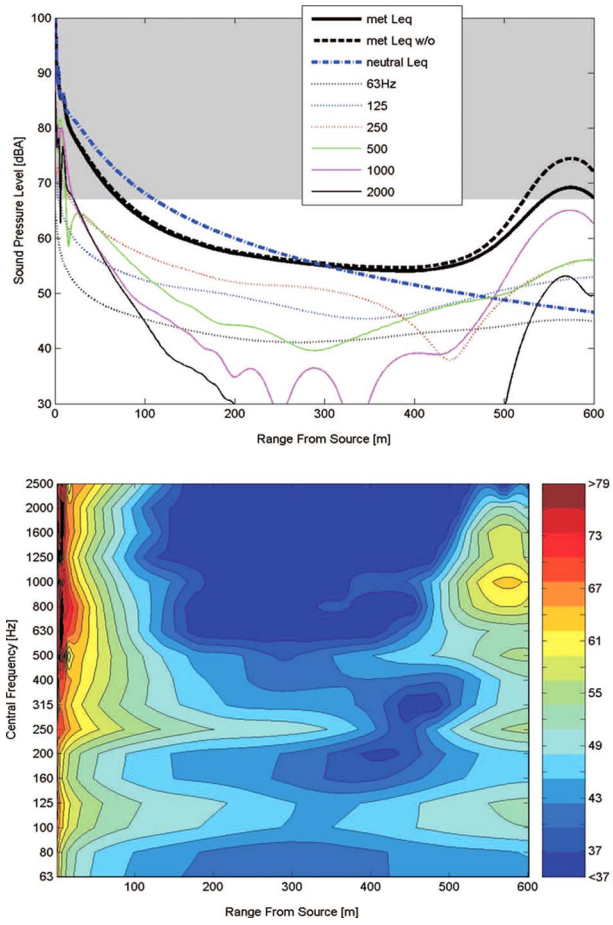


FIG. 10. (Color online) Same caption as for Fig. 8, but for Case B.

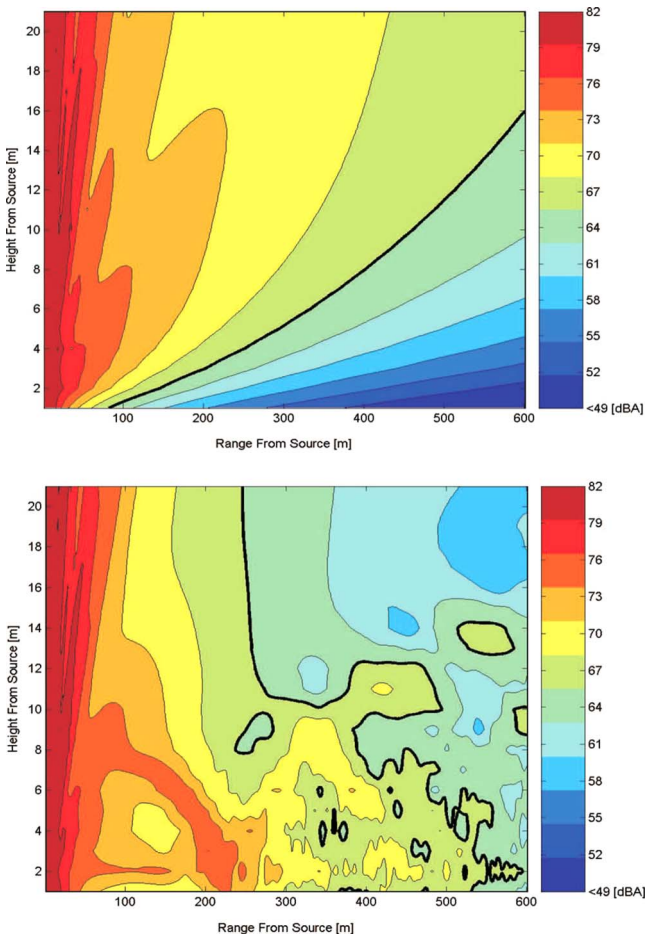


FIG. 11. (Color online) Same caption as for Figs. 7 and 9, but for Case C.

67 dBA almost up to the edge of the calculation domain. It would appear that the frequency range 800 Hz–1.6 kHz is particularly influenced and focused most intensely by the combination of wind shear and temperature gradients (see bottom of Fig. 10) although all frequency ranges appear to be subject to some degree of near-ground focusing at the 500 m range. It is especially interesting to note that the most intensely focused frequency range components appear, in fact, to fall to perceptively low sound levels (below 40 dBA) in the range 200–400 m from the freeway. The spatial contours in Fig. 9 thus strongly suggest that this case could be a typical example of excessive sound levels occurring far from the freeway which are unlikely to be abated by the use of a sound barrier.

Case C is also taken during rush-hour traffic and has the most severely altering meteorological profiles, being strongly stratified and having a crosswind jet peaking at 4 ms^{-1} at a height of 50 m above the ground. Figures 11 and 12 show a concentration of sound rays and pockets of constructive and destructive interference between the rays in a roughly 5–6 m wide layer close to the ground, particularly beyond the 300 m range. As a result, the effect of wind shear with only a mild negative temperature gradient close to the ground leads to the near-ground SPL persisting in excess of the 67 dBA threshold up to almost 350 m from the freeway. Once again the dominant frequencies responsible appear to be 1 and 1.25 kHz with other neighboring frequencies also being strongly influenced by the meteorological conditions.

VII. CONCLUSIONS/FURTHER COMMENTS

This work represents a combined experimental and theoretical study on the impact of meteorological conditions on the propagation of traffic noise from a freeway corridor. Clear indications from the results obtained are that TNMs used to judge the environmental noise impact on nearby communities must incorporate the usual or expected meteorological conditions that occur in that geographical location. This is of particular importance for the case of ARFC deployment that motivated this study, wherein the reduction in the effectiveness of the pavement with use is deduced via measurements made over certain time periods in different years. Without corrections for the effects of meteorology, the validity of such assessments is highly questionable unless only the near field data are utilized. It should also be added that some of the atmospheric effects observed in this paper offer the possibility of rendering traditional mitigation techniques, such as noise walls, ineffective. However, this would not occur with strategies based on controlling the traffic noise at source, by developing quiet pavement materials such as ARFC for instance.

The combined Green's function and PE model has shown its capabilities in taking meteorological data and near-field sound measurements to generate a spatial map of predicted noise levels. The model also enables analysis of individual frequency components (e.g., as in Fig. 8 for Case A),

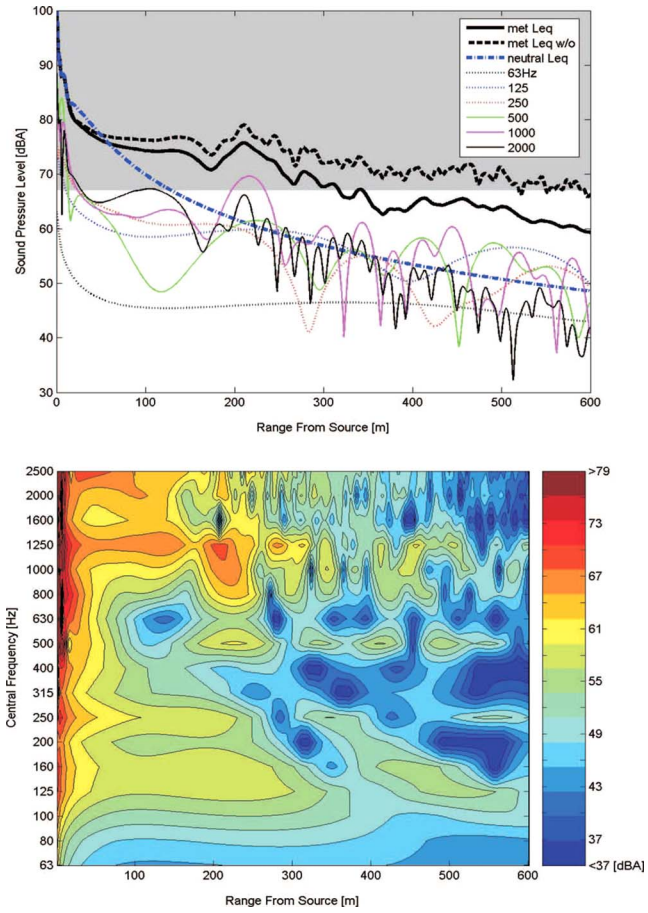


FIG. 12. (Color online) Same caption as for Figs. 8 and 10, but for Case C.

and there is some indication in this study that the frequency range 1–1.6 kHz is the most significantly influenced by meteorological conditions and thus provides the principal contribution to far-field traffic noise levels; however, this prediction requires experimental confirmation. If such further evidence arises, mitigation strategies targeting this frequency band would be the most effective in preventing excessive noise levels at large distances from the freeway corridor.

ACKNOWLEDGMENTS

We are extremely grateful to Arizona Department of Transportation (ADOT), Arizona State University (ASU), and University College London (UCL) for their support of this ongoing collaborative research. In particular, we thank Christ Dimitroplos and Fred Garcia at ADOT for their interest and encouragement. The consultants Illingworth-Rodkin assistance in the project in measuring and processing the sound data is also gratefully acknowledged. We also thank Dragan Zajic, Leonard Montenegro, and Adam Christman for their help with the field experiments and subsequent data analysis and the referees for their very constructive and encouraging comments.

[1]L. Goines and L. Hagler, "Noise pollution: A modern plague," South Med. J. **100**, 287–294 (2007).
[2]J. A. T. Granados, "Reducing automobile traffic: An urgent policy for health promotion," Rev. Panam Salud Publica **3**, 227–241 (1998).
[3]L. M. Ward and P. Suedfeld, "Human responses to highway noise," Environ. Res. **6**, 306–326 (1973).
[4]R. T. T. Forman, "Estimate of the area affected ecologically by the road system in the united states," Conserv. Biol. **14**, 31–35 (2000).
[5]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, Dordrecht, 2001).
[6]K. Attenborough, K. Li, and K. Horoshenkov, *Predicting Outdoor Sound* (Taylor and Francis, London, 2007).
[7]C. Steele, "A critical review of some traffic noise prediction models," Appl. Acoust. **62**, 271–287 (2001/3).
[8]B. Lihoreau, B. Gauvreau, M. Berengier, P. Blanc-Benon, and I. Calmet, "Outdoor sound propagation modeling in realistic environments: Application of coupled parabolic and atmospheric models," J. Acoust. Soc. Am. **120**, 110–119 (2006).
[9]R. L. Wayson, M. Martin, A. M. Edwards, and R. Wasko, "The AAMA traffic noise model: A better approach," SAE Trans. **104**, 2402–2410 (1995).
[10]J. Chambers, H. Saurenman, R. Bronsdon, L. Sutherland, K. Gilbert, R. Waxler, and C. Talmadge, "Effects of temperature induced inversion conditions on suburban highway noise levels," Acta. Acust. Acust. **92**, 1060–1070 (2006).
[11]L. Scofield, P. Donavan, "Early results of the Arizona Quiet Pavement Program," in Proceedings of the 80th Meeting of the Association of Asphalt Paving Technologists, Long Beach, CA (2005).
[12]J. Wang and J. Angell, *Air Stagnation Climatology for the United States (1948–1998)* (NOAA/Air Resources Laboratory, Atlas No. 1, Silver Spring, MD, 1999).
[13]H. Fernando, S. Lee, J. Anderson, M. Princevac, E. Pardyjak, and S. Grossman-Clarke, "Urban fluid mechanics: Air circulation and contaminant dispersion in cities," Environ. Fluid Mech. **1**, 107–164 (2001).
[14]K. E. Gilbert and M. J. White, "Application of the parabolic equation to sound propagation in a refracting atmosphere," J. Acoust. Soc. Am. **85**, 630–637 (1989).
[15]M. West, K. Gilbert, and R. Sack, "A tutorial on the parabolic equation (pe) model used for long range sound propagation in the atmosphere," Appl. Acoust. **37**, 31–49 (1992).
[16]K. E. Gilbert and X. Di, "A fast greens function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).
[17]X. Di and K. E. Gilbert, "The effect of turbulence and irregular terrain on outdoor sound propagation," in Proceedings of the Sixth Symposium on Longrange Sound Propagation (1994), pp. 315–333.
[18]R. Sack and M. West, "A parabolic equation for sound propagation in two dimensions over any smooth terrain profile: The generalised terrain parabolic equation (gt-pe)," Appl. Acoust. **45**, 113–129 (1995).
[19]M. West and Y. Lam, "Prediction of sound fields in the presence of terrain features which produce a range dependent meteorology using the generalised terrain parabolic equation (gt-pe) model," in Proceedings of Inter-Noise 2000 (2000), Vol. **2**, p. 943.
[20]D. K. Wilson, "A turbulence spectral model for sound propagation in the atmosphere that incorporates shear and buoyancy forcings," J. Acoust. Soc. Am. **108**, 2021–2038 (2000).
[21]D. K. Wilson, J. G. Brasseur, and K. E. Gilbert, "Acoustic scattering and the spectrum of atmospheric turbulence," J. Acoust. Soc. Am. **105**, 30–34 (1999).
[22]P. Chevret, P. Blanc-Benon, and D. Juve, "A numerical model for sound propagation through a turbulent atmosphere near the ground," J. Acoust. Soc. Am. **100**, 3587–3599 (1996).
[23]V. Ostashev, *Acoustics in Moving Inhomogeneous Media* (Spon, London, 1997).
[24]V. E. Ostashev, V. Mellert, R. Wandelt, and F. Gerdes, "Propagation of sound in a turbulent medium. I. Plane waves," J. Acoust. Soc. Am. **102**, 2561–2570 (1997).
[25]V. E. Ostashev, F. Gerdes, V. Mellert, and R. Wandelt, "Propagation of sound in a turbulent medium. II. spherical waves," J. Acoust. Soc. Am. **102**, 2571–2578 (1997).
[26]V. E. Ostashev, D. K. Wilson, L. Liu, D. F. Aldridge, N. P. Symons, and D. Marlin, "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation," J. Acoust. Soc. Am. **117**, 503–517 (2005).
[27]F. de Roo and I. Noordhoek, "Harmonoise WP 2 reference sound propagation model," Fortschritte der Akustik **29**, 354–355 (2003).

[28]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[29]S. Chandler-Wilde and D. Hothersall, "Efficient calculation of the Green's function for acoustic propagation above a homogeneous impedance plane," J. Sound Vib. **180**, 705–724 (1995).
[30]K. Attenborough, "Sound propagation close to the ground," Annu. Rev. Fluid Mech. **34**, 51–82 (2002).
[31]E. M. Salomons, "Improved green's function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).
[32]J. S. Robertson, W. L. Seigmann, and M. J. Jacobson, "Low-frequency sound propagation modeling over a locally reacting boundary with the parabolic approximation," J. Acoust. Soc. Am. **98**, 1130–1137 (1995).
[33]R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer Academic, Dordrecht, 1988).

Asynchronous control of vortex-induced acoustic cavity resonance using imbedded piezo-electric actuators

M. M. Zhang,[a] L. Cheng,[b] and Y. Zhou
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Home, Kowloon, Hong Kong, SAR of China

(Received 13 June 2008; revised 5 May 2009; accepted 6 May 2009)

This paper presents an experimental investigation of the control of a vortex-induced acoustic cavity resonance from flow over a bluff body using embedded piezo-ceramic actuators in order to alter the resonant flow-acoustic interactions. The action of the actuators was asynchronous. Experiments were mainly conducted at the flow velocity of acoustic resonance, where the vortex shedding frequency from the upstream bluff body approached the frequency of the first acoustic mode of two downstream cavities. The fluctuating acoustic pressure was measured using a microphone. The perturbed flow field around the bluff body was monitored using two single hot wire anemometers and one *X*-wire. It was found that the induced transverse vibrations were effective to reduce the acoustic resonance. The cavity sound pressure level at resonance was reduced by 8.2 dB in presence of actuation. The physics behind the control mechanism is discussed.
 [DOI: 10.1121/1.3143784]

PACS number(s): 43.28.Ra, 43.50.Ki [AH] Pages: 36–45

I. INTRODUCTION

Flow-induced acoustic resonance occurs as a result of strong interactions between unsteady separated flows and the acoustic modes of a cavity, when the dominant frequency of the flow separation approaches that of one acoustic mode of the cavity.[1–6] This phenomenon is commonly found in many engineering applications, and may be classified into self-induced and vortex-induced resonances. The former is excited by vortices shed from the orifice leading edge over the cavity, and often occurs in landing-gears, weapon bays in aircraft, and open cavities in moving vehicles. The latter is caused by vortex shedding from a bluff body in crossflows and is frequently seen in tube or plate bundles of heat exchangers and boilers,[7] cascades of compressor blades,[8] and guide/turning vanes in ducts and radial diffusers.[9] Flow-induced acoustic resonance may induce an acoustic pressure amplitude sufficiently high to cause very serious noise or vibration problems.[10]

Control of self-induced resonance has been extensively explored using passive and active control techniques.[11] Typical passive methods include modifications of the geometries of the acoustic source regions,[12,13] use of fences or screens within the cavities,[14,15] and use of small leading edge spoilers to control the source.[9] In contrast, active control techniques involve energy input via actuators to manipulate the shear flow, effectively reducing the flow-acoustic interaction, using either independent external disturbance, i.e., open-loop control, or feedback-signal-controlled system, i.e., closed-loop control. Active control techniques can be synchronous or asynchronous, depending on whether the actuation frequency is equal to the system fundamental frequency.[16] To mention a few typical examples, Sarno and Franke[17] successfully suppressed shear layer oscillations in a cavity through 45° steady or pulsating mass flow injection at the cavity leading edge, resulting in a reduction of 10 dB in the cavity sound pressure level (SPL) at the frequency of the acoustic resonance. Using a closed-loop method, Huang and Weaver[18] and Ziada *et al.*[19] used the fluctuating acoustic pressure inside the cavity as feedback signals, which was then amplified and phase-shifted by a controller to drive control loudspeakers at the entrance of the tunnel and at the upstream corner of the cavity, respectively. In this way, the shear layer oscillation across the cavity was attenuated. Cattafesta *et al.*[16,20,21] and Kook *et al.*[22] used an oscillating piezo-electric flap and electrodynamically driven leading edge spoilers, respectively, which were hinged near the cavity leading edge, to disturb the shear layer separation. The action of the flap was controlled by a closed-loop controller with the feedback signals from the fluctuating acoustic pressure within the cavity. As a result, attenuation in cavity pressure was obtained at the frequency of acoustic resonance.

Control of vortex-induced acoustic resonance has been scarcely reported in the literature, apart from a few papers reporting numerical simulations of the phenomenon. This phenomenon may be important in cascade structures, such as turbines, helicopters, and fans, in which a blade interacts with the wake of an upstream blade; the other is the occurrence of acoustic resonance, as evidenced by Mohany and Ziada[10] who measured the acoustic resonant interaction between vortices shed from an upstream cylinder and the first acoustic mode of the tunnel where the cylinder was located and obtained a SPL of 155 dB in a wind tunnel and the work of Roozen *et al.*[23] in which strong vortex-induced noise was measured near the bass-reflex port of a loudspeaker. Their work was mainly focused on investigations of acoustic-structure coupling rather than control.

[a]Present address: Department of Mechanical Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218.
[b]Author to whom correspondence should be addressed. Electronic mail: mmlcheng@polyu.edu.hk

Vortex-induced acoustic resonance induced by vortex shedding from bluff bodies may possibly be controlled by effectively impairing the source, i.e., vortex shedding. Many investigations of passive and active control of vortex shedding from bluff bodies have been reported in the literature. Passive control methods typically include changing the cross section of structures and adding fixed mechanical vortex disturbers such as longitudinal grooves or riblets to alter vortex shedding.[24–26] As examples of active open-loop control, Hsiao and Shyu[27] used acoustic waves emitted from a slot on the surface of a cylinder to actively disturb the fluid field and demonstrated that a local disturbance near the shear layer instability frequency and around the flow separation point caused an increase in lift and a reduction in drag and the vortex strength (Re=420–34 000). Williams *et al.*[28] introduced symmetric and anti-symmetric forcing of water flow (Re=470) at a frequency of about twice the vortex shedding frequency (f_s) through two rows of holes located at $\pm 45°$, respectively, away from the forward stagnation line of the cylinder. They observed a change in shedding frequency and flow structure. Most existing closed-loop vortex shedding control methods rely on feedback signals provided by hot wires in the turbulent wake. Warui and Fujisawa[29] reduced the vortex strength at Re=6700 using electromagnetic actuators installed at both ends of a circular cylinder to create a lateral oscillation. Tokumaru and Dimotakis[30] and Filler *et al.*[31] created cylinder rotary oscillations to produce regulated injection of circulation into the wake. Both led to an attenuation in vortex strength and drag. Ffowcs Williams and Zhao[32] used a loudspeaker mounted on a wind tunnel wall to impair vortex shedding from a cylinder at Re=400. Using the same technique, Roussopoulos[33] observed an increase by 20% in the onset Reynolds number for vortex shedding. Huang[34] used sound within a cylinder to generate a pulsating flow through a thin slit near the separation point on the cylinder surface, and suppress vortex shedding from both sides of the cylinder in the Re range between 4×10^3 and 1.3×10^4.

Cheng *et al.*[35] developed a novel perturbation technique to control both vortex shedding and structural vibration. The essence of the technique is to generate a controllable transverse motion of a structural surface using embedded piezo-ceramic actuators to alter fluid-structure interactions. The effectiveness of this technique has been demonstrated for the case of active control of vortex shedding and associated structural vibrations for different cases of fluid-structure interactions, including resonant flow-structure coupling on a flexible-supported rigid cylinder,[36,37] resonant flow-structure coupling on a fix-supported flexible cylinder,[38] and non-resonant coupling on a fix-supported flexible cylinder.[39] This technique has recently been successfully applied to the control of noise caused by blade-vortex interaction[40] (BVI) and the control of airfoil aerodynamics.[41] Piezo-ceramic actuators are lighter and smaller than other actuation devices such as loudspeakers and electromagnetic actuators. Owing to its special design, the actuator presently used requires a relatively low energy input to generate appreciably large displacements. Typically, without any loading, it can vibrate at a maximum displacement of about 2 mm and a frequency up to 2 kHz.

This paper presents results from an experimental study to extend the aforementioned technique to the control of the vortex-induced acoustic resonance by means of reducing vortex strength by a bluff body. Asynchronous control was carried out in this study to establish the feasibility of the technique. The test configuration comprises an upstream bluff body acting as the vortex generator, and two downstream acoustic cavities, within which acoustic resonance phenomenon occurs. Experiments were focused on the occurrence of acoustic resonance when the vortex shedding frequency coincided with the frequency of the first acoustic resonance of the cavity. Uncontrolled flow-acoustic interactions were first investigated. A simple asynchronous control system was developed. Control performance was assessed in terms of the cavity SPL at resonance. To understand the control mechanism, the perturbed flow field and the resonant flow-acoustic interaction were analyzed in detail.

II. EXPERIMENTAL PROCEDURES

Experiments were carried out in a closed circuit wind tunnel, as shown in Fig. 1(a). This facility, designed for aeroacoustic experiments, was previously used for flow-silencer testing.[42] It has a 1.82-m-long square test section of 0.1×0.1 m^2. A parabolic contraction at the inlet improved the uniformity of the flow velocity profile and reduced boundary layer thickness. A flat-walled diffuser, with a half angle of 14°, was used downstream of the working section to increase pressure recovery. The maximum flow velocity was 50 m/s with a turbulence intensity of less than 0.1% in the upstream section. The background noise of the tunnel was low since the motor and fan noise was mostly absorbed by acoustic lining. This has been experimentally verified.

A rigid thick rectangular plate, with a height of $h=11$ mm, a width of $w=47$ mm, and a length of 98 mm, was used as bluff body. The plate was rigidly fixed on both side walls of the test section and located about 0.37 m downstream of the exit plane of the tunnel contraction, as shown in Figs. 1(a) and 1(b). The plate angle of attack was zero. Note that w/h was about 4.3, falling into the range between 3.2 and 7.6, where only one vortex separated from the leading edge of the plate may develop along the plate at any instant. This is a typical pattern of flow around a rectangular plate.[43] To improve the two-dimensionality of vortex shedding from the plate, the cross section of the plate leading edge was made semi-circular[44] [Fig. 1(b)]. Two identical cavities with square cross sections were located downstream of the plate, symmetrical with respects to x-z plan, on the top and bottom walls of the tunnel, respectively, and at the same streamwise location. The two cavities act as side-branch acoustic resonators. The origin of the coordinate system, shown in Fig. 1(a), was defined at the center of the model trailing edge, with x, y, and z corresponding to the streamwise, transverse, and spanwise directions, respectively. The depth (L) and width (B) of each cavity were 440 and 70 mm, respectively. The first acoustic resonance frequency of the

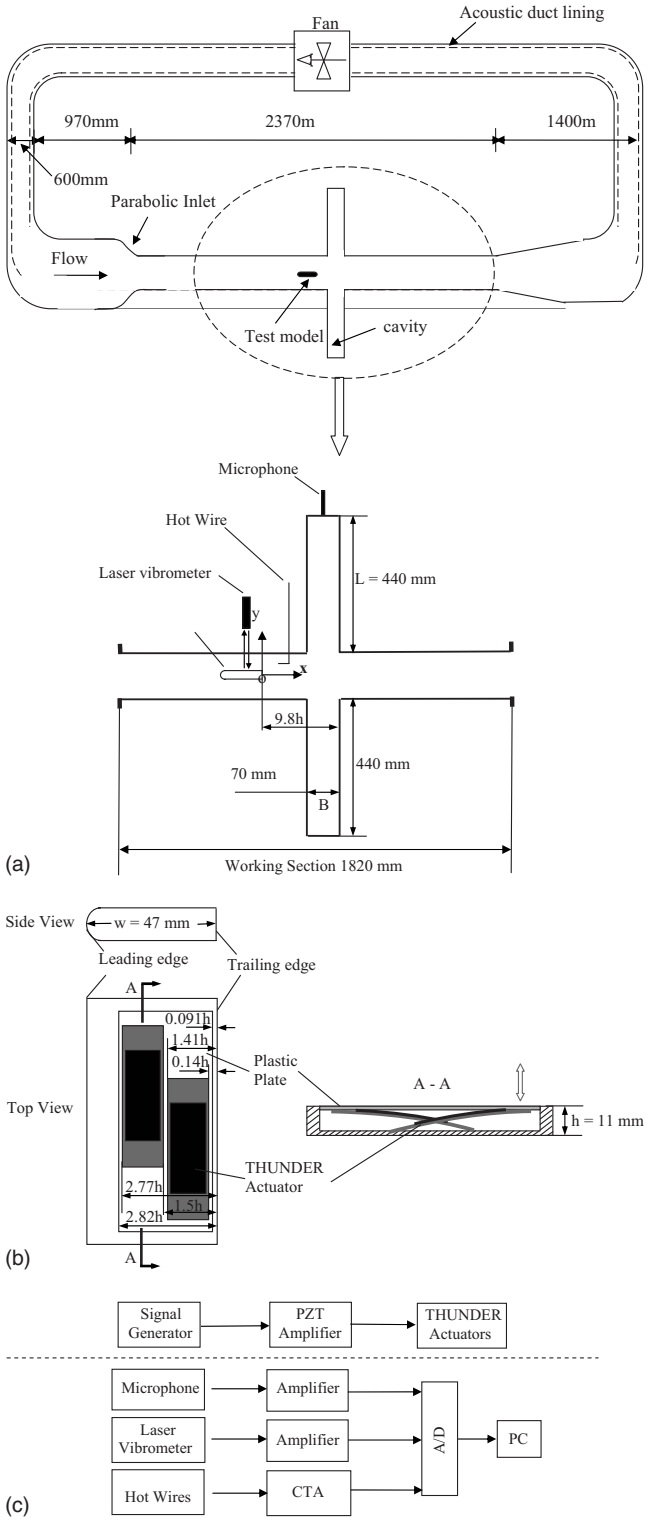


FIG. 1. Experimental setup: (a) Wind-tunnel and sensing configuration; (b) test model in detail; (c) asynchronous control system and measurement system.

cavity (f_a') was approximately $f_a'=c/4L\approx 193.2$ Hz,[45] where c is the speed of sound. The corresponding critical flow velocity U_{cr} ($=f_s h/\mathrm{St}$) at resonance, when shedding frequency $f_s=f_a'$, was estimated to be about 8.5 m/s, using a Strouhal number St of 0.25, as suggested by Welsh *et al.*[44] for similar w/h ratios. The distance between the trailing edge of the plate and the downstream wall of the cavities was about $9.8h$. This distance ensured an effective resonant fluid-acoustic interaction in the near wake of the plate, as demonstrated later.

Two curved piezo-ceramic actuators [thin layer composite unimorph driver and sensor (THUNDER)],[46] with a length of 63 mm and a width of 14 mm, were embedded in a slot of 80 mm long, 30 mm wide, and 7 mm deep on the top side of bluff body and 1.5 mm from the model trailing edge [Fig. 1(b)]. The two actuators were installed at $x=-1.41$ to $0.14h$ and $x=-2.77$ to $-1.50h$ [Fig. 1(b)], respectively. The characteristics of the actuators have been thoroughly discussed previously.[46] Typically, without any loading, the present actuator (THUNDER-8R) can vibrate with a peak displacement of about 2 mm and a frequency up to 2 kHz. The actuators were installed in a cantilever manner to create the maximum perturbation displacement in the transverse y-direction, and thus better control performance for the same excitation condition. The actuators and the walls of the slot around the actuators were lubricated to minimize contact friction. A thin plastic plate with a thickness of 1.2 mm, mounted flush with the upper surface of the plate, was connected with the cantilevered end of the actuators using double-sided glue and was placed at $x=-2.82$ to $-0.091h$ [see Fig. 1(b)]. Driven by the actuators, the plate oscillated to create a uniform transverse vibration of the plate surface, as confirmed by the measurement of velocity at several points over the place by a laser vibrometer. The motion of the plate may have created a small step on the test model surface. But the gap between the plate and the model was small such that leakage effects were deemed negligibly small. Furthermore, the plate had a maximum displacement of less than 0.9 mm, and was rather stiff. Therefore, spanwise variation in transverse displacement (80 mm) due to the actuators could be neglected. The actuators were simultaneously activated by a sinusoidal signal with controllable frequency. The input voltage was generated by a simple asynchronous control system, including a signal generator (Model DS345) and a dual-channel piezo-driver amplifier (Trek PZD 700), as indicated in Fig. 1(c).

A 12.7 mm diameter condenser microphone (B&K 4189), mounted flush with the center of the top wall of the upper cavity [Fig. 1(a)], was used to measure the fluctuating acoustic pressure inside the cavity. To analyze the control effect on the flow field, one or two tungsten single hotwires of 5 μm in diameter, with wire aligned with z direction, were deployed to measure the fluctuating flow velocity and streamwise mean velocity. Figure 1(a) shows a typical arrangement of one single wire case, placed at $x/h=2$ and $y/h=1.5$. Furthermore, the fluctuating flow velocities along x- and y-directions in the wake of the test model were measured using a 5 μm tungsten X-wire. The measured flow velocities were corrected using the method of Durgun and Kafali,[47] in view of the present blockage ratio ($\approx 11\%$). In addition, the perturbation displacement (Y_p) was measured using a Polytec Series 3000 dual beam laser vibrometer [Fig. 1(a)]. After amplification, all measurement signals were recorded using a personal computer through a 12-bit AD board at a sampling frequency of 6 kHz per channel. The duration of each record was about 20 s.

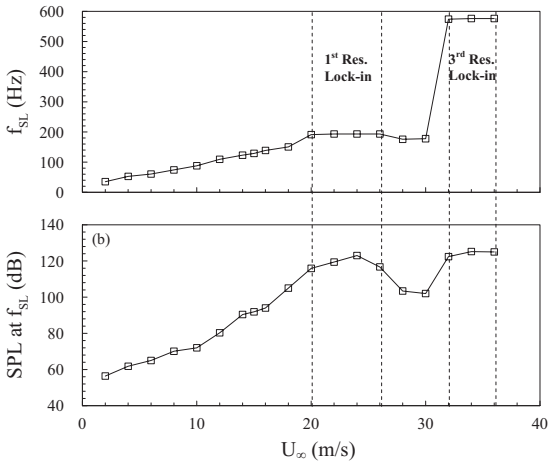


FIG. 2. Dependence of the frequency of the shear layer (f_{SL}) separated from the two cavities (a) and SPL at f_{SL} (b) on the free-stream flow velocity (U_∞) in the absence of the test model.

III. UNCONTROLLED FLOW-ACOUSTIC INTERACTION

Uncontrolled flow-acoustic interactions were first examined to provide a baseline for comparison. As previously mentioned, in the present setting, one would expect the sound inside the cavity to be induced by both self-induced resonance and vortex-induced resonance. It is therefore important to ensure that the two cases have very different critical flow velocities so that they can be separated. In view of this, a test was first conducted before the installation of the bluff body to investigate the variation of the frequency of the shear layer (f_{SL}) separated from the two cavities and SPL at f_{SL} with the free-stream flow velocity U_∞, as shown in Fig. 2. It can be seen from Fig. 2(a) that f_{SL} generally increased with U_∞. Two plateaus appear within the velocity ranges of U_∞=20–26 m/s and U_∞=32–36 m/s [Fig. 2(a)], where f_{SL} is locked-in to the first and third acoustic mode frequencies of the cavity, i.e., f_a'(=193.4 Hz) and f_a'''(=576.2 Hz), respectively. The first range matches the theoretical prediction of the critical flow velocity corresponding to the first mode resonance [U_{cr}=21.3 m/s, calculated from $U_{cr}=f_a'B\pi/(4St)$,[48] with St=0.5 (Ref. 45)]. The SPL at f_{SL} is rather high, up to 126 dB, in the two lock-in ranges, suggesting the occurrence of strong resonance. Note that the second acoustic resonance mode was not excited, which is consistent with the analyses on acoustic resonance in a cavity with one closed end by Morse.[49]

Similar experiments were repeated after the test model was installed. Figure 3 illustrates the dependence of the frequency of the vortices shed from the test model (f_s), SPL at f_s and the magnitude of the power spectrum of the fluctuating flow velocity E at f_s on the free-stream flow velocity (U_∞). The hotwire was located at x/h=1 and y/h=0.75, where the flow velocity fluctuation was observed to be very strong. Although the lock-in phenomenon under the first-mode acoustic resonance still exists, the flow velocity range corresponding to f_a'(=193.4 Hz), i.e., 8.3 m/s$<U_\infty<$9.2 m/s, obviously becomes narrower, consistent with observation of Welsh *et al.*.[44] Recall that the calculated f_a'(=193.2 Hz) is very close to the measured value and the calculated U_{cr}(=8.5 m/s) lies in the lock-in range from 8.3 to 9.2 m/s,

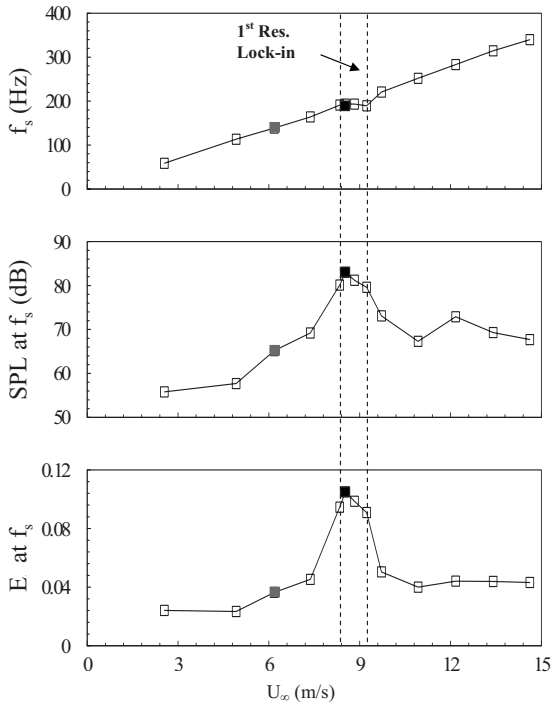


FIG. 3. Dependence of the frequency of the vortices shed from the test model (f_s), SPL at f_s and the magnitude of the power spectrum of the fluctuating flow velocity E at f_s on the free-stream flow velocity (U_∞) in the presence of the test model. The hot wire was located at x/h=1 and y/h=0.75.

in reasonable agreement with the measurement. Furthermore, for 8.3 m/s$<U_\infty<$9.2 m/s, the averaged SPL at f_s is about 81 dB when the test model is installed (Fig. 3). Clearly, the oscillations caused by self-induced resonance and the vortex-induced resonance have very different critical flow velocities so that their effects can be easily separated. The strongest acoustic resonance was observed at U_∞=8.53 m/s, corresponding to a Reynolds number Re ($\equiv U_\infty h/\nu$, where ν is the kinematic viscosity) of 6200, where both the SPL at f_s and the E at f_s reach their maximum value of 83 dB and 0.11, respectively. This working condition was chosen as the control target in the following control tests.

IV. CONTROL PARAMETERS AND PERFORMANCES

Two important parameters, i.e., the perturbation frequency f_p and perturbation voltage V_p of the controller need to be determined. To this end, a series of tests was conducted to document the effect of these two parameters on the SPL at f_a', i.e., the strength of the vortex-induced acoustic resonance. Figure 4(a) shows the variation of the SPL at f_a' and the root mean square (rms) value of perturbation displacement Y_{prms} with f_p. A maximum permissible voltage, with a rms value V_{prms}=141 V, was applied to the actuators as f_p varied within 0–90 Hz. In this low frequency range, the noise generated by the actuator itself, if any, was undetectable by the microphone. Compared with the unperturbed case (f_p=0 Hz and V_{prms}=0 V), the SPL at f_a' was reduced when the actuation was employed. The maximum attenuation occurred at f_p=30 Hz, resulting in a reduction of 8.2 dB in the SPL. At a fixed f_p(=30 Hz), Fig. 4(b) shows the variation of the

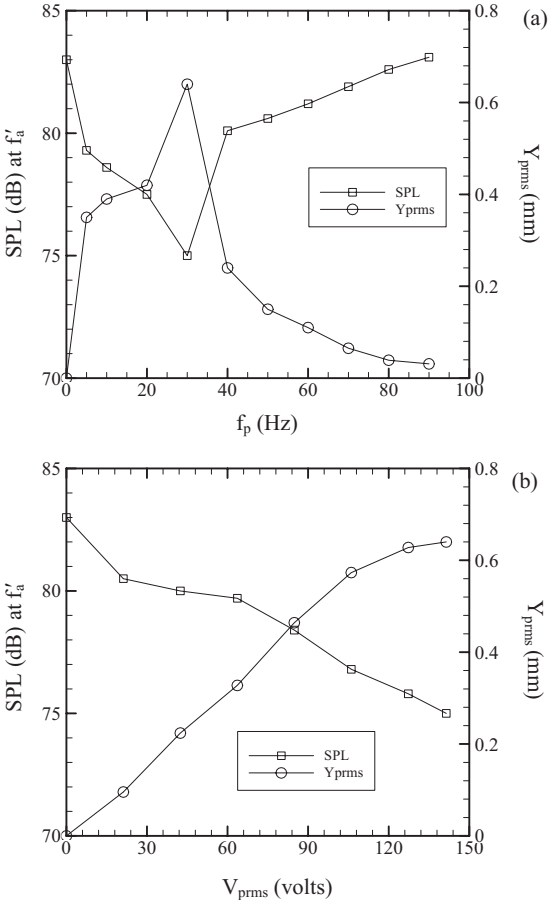


FIG. 4. Effect of control parameters on the SPL at f_a' and the rms value of the perturbation displacement Y_{prms}: (a) different perturbation frequency f_p at V_{prms}=141 V; (b) different V_{prms} at f_p=30 Hz. Re=6200.

SPL at f_a' and Y_{prms} with V_{prms}. It can be seen that the SPL at f_a' monotonously decreased as V_{prms} was increased from 0 to 141 V. Obviously, a greater actuation voltage is desirable. Further iterations between f_p and V_{prms} failed to improve the control performance significantly, implying that the combination of f_p=30 Hz with V_{prms}=141 V were the optimum parameters for the asynchronous controller. In fact, the natural frequency of the actuator presently used was designed to be around 30 Hz, at which the rms value of the perturbation displacement Y_{prms} was largest, up to about 0.64 mm [Fig. 4(a)]. In addition, Y_{prms} increased monotonously with V_{prms} [Fig. 4(b)], though not linearly.[50] The large Y_{prms} may partially account for the better control authority at f_p=30 Hz and V_{prms}=141 V over other combinations of f_p and V_{prms}. Unless otherwise stated, this optimum set of parameters was used in the experiments hereinafter. The power spectral density of the surface oscillating velocity was measured using the laser vibrometer, as shown in Fig. 5. It has been observed that the actuation-produced surface oscillation is predominately controlled by its first harmonic component at 30 Hz. Other harmonics are all below 4% of the dominant frequency. Therefore, the surface motion could be approximately considered to be simple harmonic.

The control performance of the above tuned asynchronous controller was further evaluated in terms of the power spectrum of SPL at f_a'. Figure 6 shows the results, with and

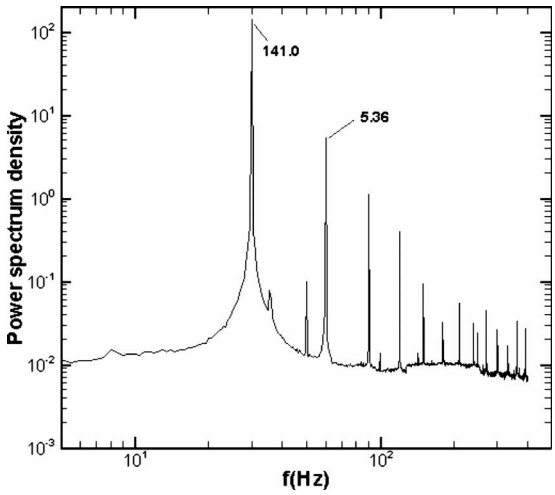


FIG. 5. Power spectrum of perturbation when the perturbation frequency f_p was 30 Hz and rms value of perturbation voltage V_{prms} is 141 V.

without control, at Re=6200. Four evident peaks in the SPL-spectrum appear in the absence of control, corresponding to the natural frequency of shear layer separation, i.e., $f_{SL}'^*$, the frequencies of the first, third, and fifth cavity acoustic modes, i.e., $f_a'^*$ (=vortex shedding frequency $f_s'^*$), $f_a'''^*$, and $f_a^{(5)*}$, respectively. The asterisk used in this paper denotes the normalization of frequency f by h and U_∞, i.e., $f^*=fh/U_\infty$. As mentioned before, only odd modes of the cavity appear. Figure 6 shows that, under control, the SPL at $f_a'^*=f_s'^*$ decreases from 83 to 74.8 dB, although a small peak at $f_p^*=0.039$ emerges, apparently due to the excitation. In addition, reductions in the SPL at $f_a'''^*$ and $f_a^{(5)*}$ are also noticeable. Generally, the control was deemed to be effective.

V. DISCUSSIONS

To understand the underlying physics, the perturbation effect on the flow field in the wake of the test model was examined. Figure 7 displays the power spectrum, E, of fluctuating flow velocity with and without perturbation. The

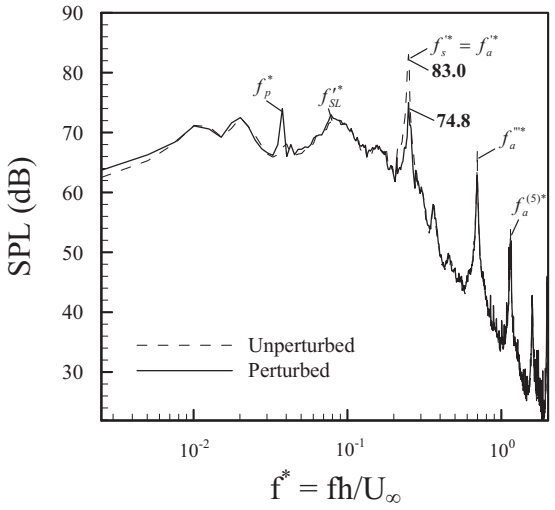


FIG. 6. Spectrum of SPL with and without perturbation. The perturbation frequency and the rms value of perturbation voltage were 30 Hz and 141 V, respectively. Re=6200.

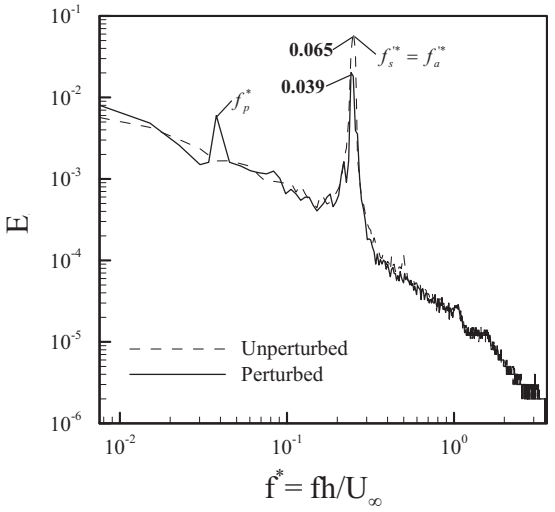


FIG. 7. Power spectrum, E, of the fluctuating flow velocity with and without perturbation (Re=6200). The hotwire was located at $x/h=1.5$ and $y/h=1$.

spectra were normalized in this paper so that $\int_0^\infty E(f)df=1$. The hotwire was placed at $x/h=1.5$ and $y/h=1$. Without perturbation, a pronounced peak, due to vortex shedding from the test model, occurred at $f_s'^*=f_a'^*=0.25$ in E. The peak magnitude at $f_s'^*$ was 0.065. Upon the deployment of control, the peak magnitude at $f_s'^*$ was reduced to 0.039 or around 5 dB, i.e., 60% of its unperturbed counterpart, suggesting an effective impairment in the energy of the vortices. In order to make sure the control does not affect the flow only locally, Fig. 8 compares the cross-flow distributions of the mean velocity $\bar{U}^*$ and the Reynolds stresses $\overline{u^2}^*$, $\overline{v^2}^*$, and $\overline{uv}^*$ measured by an X-wire at $x/h=2$ with and without perturbation. Compared with the unperturbed case, the minimum $\bar{U}^*$ and maximum $\overline{u^2}^*$, $\overline{v^2}^*$ and $\overline{uv}^*$ exhibit a considerable decrease by 9.3%, 21.6%, 36.7%, and 38.9%, respectively. The similar control effect is still discernible at $x/h=5$, as shown in Fig. 9. Clearly, the vortex strength in the wake of the model is impaired. The increased mean velocity deficit in $\bar{U}^*$ is related to the decreased entrainment of high speed fluid from the free-stream, which is induced by the weakened vortex strength.[29] Furthermore, the reduced maximum $\overline{u^2}^*$, $\overline{v^2}^*$ and $\overline{uv}^*$ may be ascribed to the impaired vortex strength.

It is pertinent to mention that the ability of the present technique in suppressing vortex formation from bluff bodies and reducing vorticity has been demonstrated using extensive particle image velocimetry (PIV) measurement in the past.[35–39] Although the present test model differs from the ones used in the previous studies, the basic phenomenon should remain the same. According to Howe's vortex sound theory,[3] the reduction in vortex circulation strength is responsible for the noise source reduction.

Insight may be better gained into the impairment in the vortex strength by examining how the flow around the test model evolves and responds to the control. Figure 10 shows the control effect on the fluctuating flow velocity spectrum E at different streamwise locations with constant $y/h=0.75$. Before perturbation, a very small peak in E (0.0014 in magnitude) appears at $f_s'^*$, as the hotwire is located above the leading edge of the test model, i.e., $x/h=-4$ and $y/h=0.75$ [Fig. 10(a)]. Toward the trailing edge, the magnitude of E at $f_s'^*$ continually increases, though still rather small until the hotwire reaches the model trailing edge at $x/h=0$ and $y/h=0.75$ [Figs. 10(b)–10(e)]. The peak value of E at $f_s'^*$ undergoes drastic changes behind the trailing edge of the test model at $x/h=1$, jumping from 0.011 to 0.11 [Fig. 10(f)], where the peaks at the second and third harmonics of $f_s'^*$, i.e., $f_s''^*$ and $f_s'''^*$, are also evident. The measured E at $f_s'^*$ progressively decreases with the increasing x/h but remains much larger than those at $x/h\leqslant 0$ [Figs. 10(g)–10(j)]. The modification in unperturbed E at $f_s'^*$ can be attributed to the evolution of vortex shedding from the test model. These observations are, in fact, consistent with numerical findings reported by Hourigan *et al.*[51] and Mills *et al.*,[52] who observed that flow over an elongated bluff body separated from the plate leading edge and formed vortices. The vortices then reattached on the plate surface and continually moved along it until reaching the plate trailing edge, where they were shed into the wake. On the other hand, as a leading edge vortex approached the trailing edge, the redeveloped thin boundary

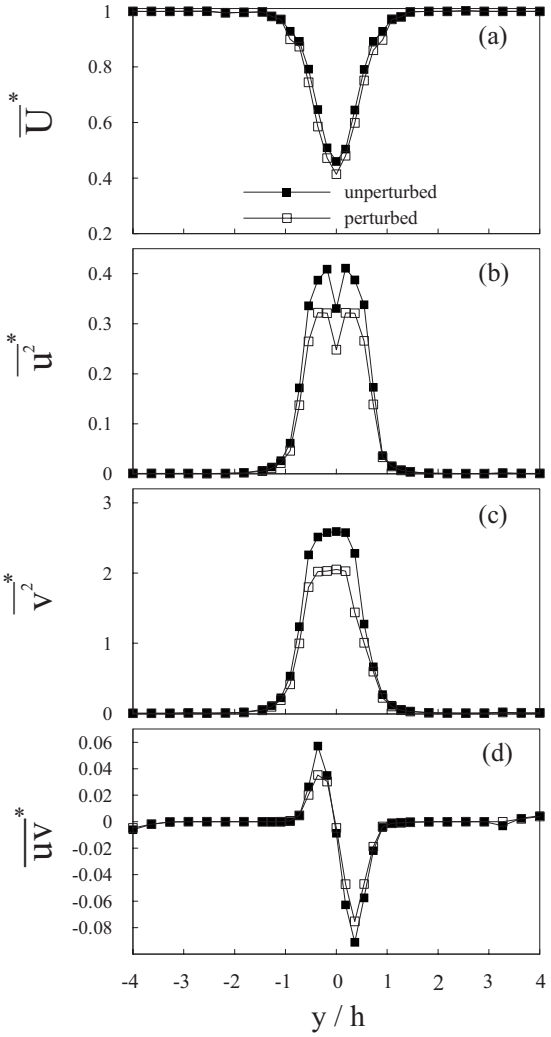


FIG. 8. Cross-flow distribution of mean streamwise flow velocity and Reynolds stresses at $x/h=2$: (a) $\bar{U}^*$, (b) $\overline{u^2}^*$, (c) $\overline{v^2}^*$, (d) $\overline{uv}^*$. Re=6200.

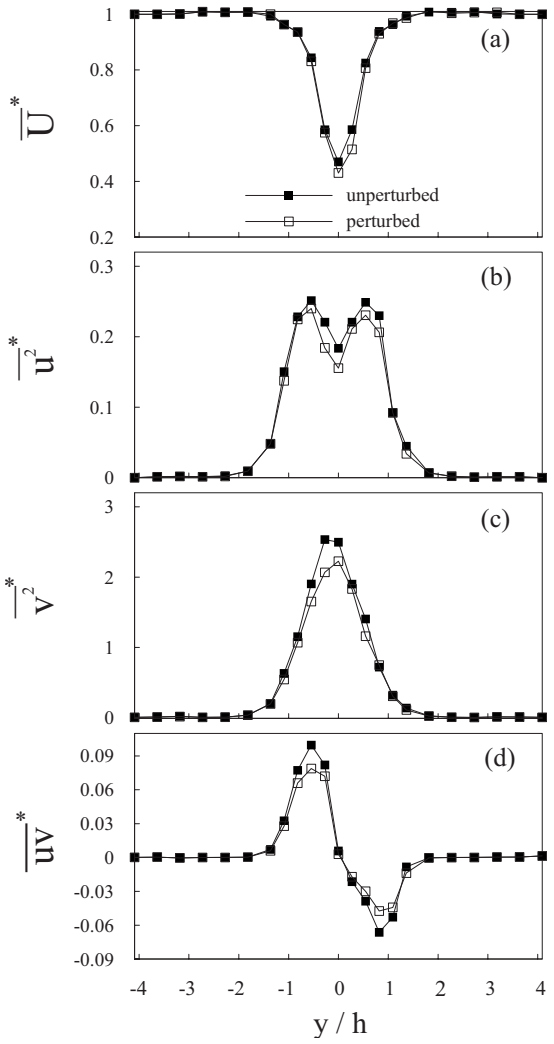


FIG. 9. Cross-flow distribution of mean streamwise flow velocity and Reynolds stresses at $x/h=5$: (a) $\overline{U}^*$, (b) $\overline{u^2}^*$, (c) $\overline{v^2}^*$, (d) $\overline{uv}^*$. Re=6200.

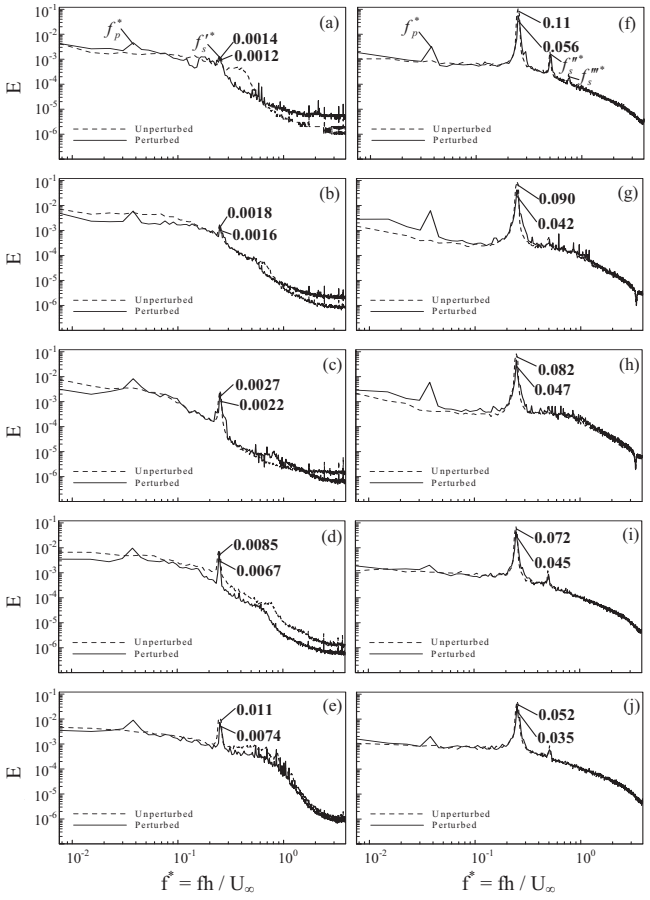


FIG. 10. The fluctuating velocity spectrum E with and without perturbation at different streamwise locations ($y/h=0.75$): (a) $x/h=-4$; (b) $x/h=-3$; (c) $x/h=-2$; (d) $x/h=-1$; (e) $x/h=0$; (f) $x/h=1$; (g) $x/h=2$; (h) $x/h=3$; (i) $x/h=5$; (j) $x/h=7$. Re=6200.

layer in front of it would also separate at the plate trailing edge and rolled up to form a trailing-edge vortex of like sign. The vortices from the leading and trailing edges interact and their interaction occurs alternately on each side of the plate at the upper and lower trailing-edge corners, resulting in a pair of regular vortex streets in the wake. In the present case, the frequency of vortex shedding from the leading/trailing edge is locked-in with f_a'.[53] In addition, the strength of the trailing-edge vortex was much higher than that of the leading edge vortex.[53] This could explain the small peaks in E at $f_s'^*$ over the plate surface and the jump in E at $f_s'^*$ in the wake of the model. Once the perturbation was applied, the perturbation disturbed the redeveloped boundary layer on the plate from which the trailing-edge vortices formed, resulting in a significant reduction in the energy of vortices, especially in the wake of the model (Fig. 10). This statement can be further substantiated by using a more accurate method to quantify the variation in $E_{\Delta f}$ associated with $f_s'^*$ by integrating E over a −3 dB bandwidth about $f_s'^*$ in Fig. 10. The percentage (δ) of energy reduction under control, compared with the uncontrolled case, at different locations are calculated and shown in Fig. 11. It can be seen that the reduction in $E_{\Delta f}$ is most pronounced at the trailing edge near the model; for example, δ are about 67%, 57%, and 51%, corresponding to $x/h=1$, 2, and 3, respectively. The impaired vortices near the model may be possibly responsible for the considerable decrease in the SPL at f_a', consistent with the results in Fig. 5.

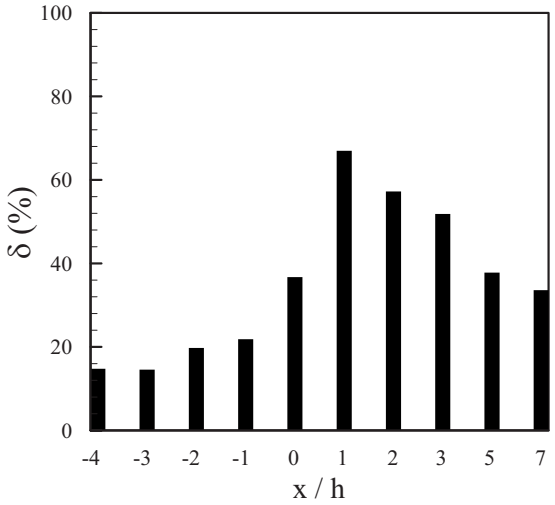


FIG. 11. Comparison in the energy reduction percentage (δ) of the fluctuating flow velocity at different streamwise locations ($y/h=0.75$). Re=6200.

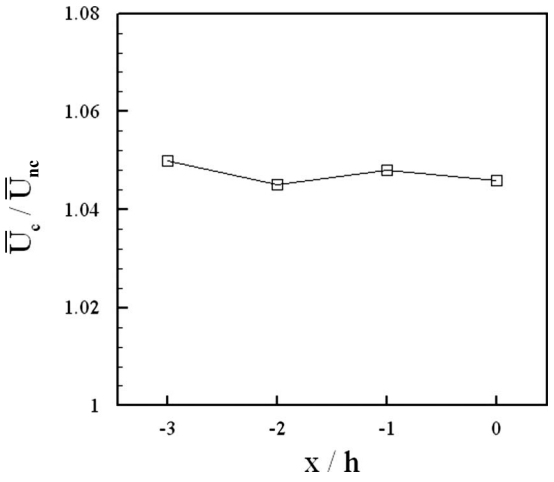


FIG. 12. The streamwise variation of the mean flow velocity under control (y/h=0.75). Re=6200.

From the results in Figs. 10 and 11, one may surmise that the weakened vortex street results from the impaired strength of the dominant trailing-edge vortex. To confirm this, the streamwise mean flow velocity above the surface of the model was measured under control (y/h=0.75), as indicated in Fig. 12. The measured flow velocity $\bar{U}_c$ exceeds the unperturbed velocity $\bar{U}_{nc}$ by about 5% at x/h=−3 to 0. The typical time history of the streamwise total flow velocity signal U, measured by the hotwire placed at x/h=−3 and y/h=0.75, highlights an increase in $\bar{U}_c$ once the perturbation is introduced (Fig. 13). The perturbation displacement (Y_{prms}=0.64 mm) will pump energy into the boundary layer over the plate and lead to transition to turbulence and subsequently early reattachment of the leading edge vortices. This transition may entrain high-speed free-stream fluid to the flow near the plate and increase $\bar{U}_c$. Furthermore, the enhanced $\bar{U}_c$ may accelerate the advection of the leading edge vortex along the model surface, as confirmed by the perturbation effect on the phase shift $\phi_{u_1u_2}$ (Fig. 14) between two streamwise flow velocities, u_1 and u_2, simultaneously measured using two single hotwires above the model leading edge (x/h=−4, y/h=0.75) and trailing edge (x/h=0, y/h=0.75), respectively. Here, $\phi_{u_1u_2}$ is defined by $\phi_{u_1u_2}[\equiv \tan^{-1}(Q_{u_1u_2}/Co_{u_1u_2})]$, where $Co_{u_1u_2}$ and $Q_{u_1u_2}$ are the cospectrum and quadrature spectrum of u_1 and u_2, respectively. The cross-spectrum was computed from the Fourier transform of the correlation $\overline{u_1(t+\tau)u_2(t)}$.[36] Obviously, $\phi_{u_1u_2}$ at $f_s'^*$ decreases from 2.6 to 1.7 after perturbation. The comparison in $\phi_{u_1u_2}$ was conducted within 1 cycle of leading edge vortex shedding since w/h was designed such that only one vortex separated from the leading edge of the plate is developed along the plate at any instant.[43] The decreased $\phi_{u_1u_2}$ implies less time for a vortex to travel from the leading edge to the trailing edge under control. In a test carried out in a water tunnel, Mills *et al.*[52] introduced a transverse disturbance through the oscillation of the top and bottom walls of the working section to disturb the flow field around a rectangular plate with w/h=6–10. Before control, they found that the undisturbed leading edge vortices always passed the plate trailing edge at the same time in one cycle as the leading edge vortex shedding regardless of the w/h value. They further observed that, for steady flow, as the leading edge vortices on one side of the plate passed the corresponding trailing-edge corner, the reattached boundary layer adjacent to the other trailing-edge corner began to roll up, forming trailing-edge vortices, which alternately occurred on both sides of the plate. Based on the observations, they successfully impaired trailing-edge vortices by controlling the disturbance phase within a full disturbance cycle as the leading edge vortices reached the trailing-edge corner. By doing so, the disturbed leading edge vortices at one trailing-edge corner could move downward or upward to interfere with the initial rollup motion of the redeveloped boundary layer from another trailing-edge corner, thus greatly reducing the vortex strength of trailing-edge vortices. Similarly, the present accelerated leading edge vortices on the upper side of the plate may interact with the redeveloped boundary layer separated from the lower trailing-edge corner. Without perturbation, the interaction/disturbance will not occur because the upper

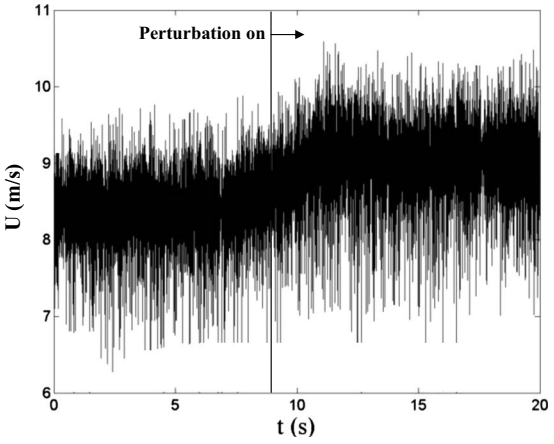


FIG. 13. Typical transition of the total streamwise flow velocity signal U when the perturbation was switched off and on. The hotwire was located at x/h=−3 and y/h=0.75. Re=6200.

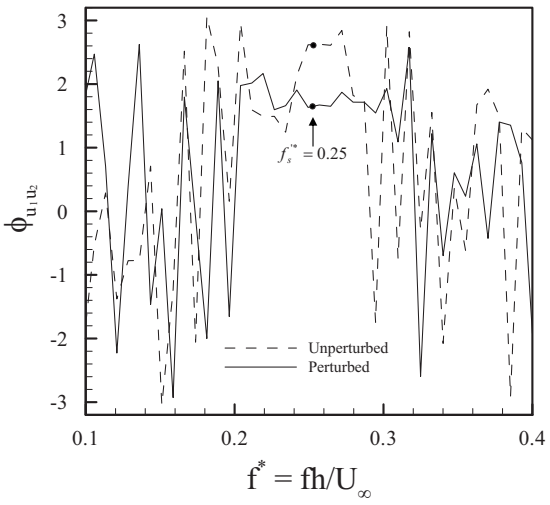


FIG. 14. Spectral phase $\phi_{u_1u_2}$ between the fluctuating flow velocity (u_1) at x/h=0 and y/h=0.75 and the fluctuating flow velocity (u_2) at x/h=−4 and y/h=0.75. Re=6200.

leading edge vortices and lower trailing-edge vortices cannot simultaneously appear near their corresponding trailing-edges. This disturbance of upper leading edge vortices effectively perturbs the rollup of lower trailing-edge vortices by collision, leading to the impaired strength of lower trailing-edge vortices.

As seen in Figs. 8 and 9, the perturbation on the upper side of the model has the equal effect on either side of the wake centerline. This has been observed in other cases[54] and is attributed to the steady formation of vortex streets. In fact, vortex shedding is a result of initial wake instability.[55] In order to form a stable vortex street, it is essential for the two oppositely signed vortices separating from the cylinder to have approximately the same strength through interactions.[56] Therefore, vortex shedding from both sides of the plate appears equally affected.

From a different perspective, the weakened wake behind the model was shown to result in a modification in flow-acoustic correlation. As a matter of fact, the spectral coherence between the fluctuating flow velocity (u) and the fluctuating acoustic pressure (p) showed that the strong correlation between vortex shedding from the model and the first acoustic mode of the cavities receded once the perturbation is introduced.

Based on the above analyses, an interpretation for the impaired vortex-induced acoustic resonance mechanism is now proposed. The vortex-induced acoustic resonance originates from the strong interaction between the coupled vortices shed from both leading and trailing-edges of the bluff body and the first acoustic mode of the cavities. The trailing-edge vortex overwhelms the leading edge vortex in terms of the vortex strength. The asynchronous controlled surface perturbation accelerates the advection of the reattached leading edge vortices along the model surface and thus disturbs the formation of the trailing-edge vortices, leading to a significant reduction in the vortex strength in the wake of the model. The whole process results in an effective impairment in the flow-acoustic interaction and subsequently the vortex-induced acoustic resonance.

VI. CONCLUSIONS

Control of vortex-induced acoustic resonance has been experimentally investigated using a perturbation technique. The investigation leads to following conclusions.

(1) Vortex-induced acoustic resonance may be effectively controlled using an imbedded piezo-electric actuator along with an asynchronous control system; the SPL at the first acoustic resonance was reduced by 8.2 dB.
(2) Analyses of the measured data suggest that the effective control lies in the modification of the vortex strength in the wake of the bluff body. The vortex-induced acoustic resonance results from strong interactions between vortices originated from both leading and trailing-edges of the plate and the acoustic mode of the downstream axisymmetric cavities. The convection of the vortex originated from the leading edge along the plate surface was accelerated by the controllable surface perturbation. This subsequently changes the initial conditions in the formation of the trailing-edge vortex and further enables a vigorous interaction between the two types of vortices, thus weakening remarkably the vortex strength in the wake of the plate and hence the flow-acoustic interaction.

ACKNOWLEDGMENTS

The authors wish to acknowledge support given to them by the Research Grants Council of HKSAR through Grant No. PolyU 5132/07E. The author wishes to acknowledge the constructive comments from the two anonymous reviewers during the review process.

[1]R. D. Blevins, "The effect of sound and vortex shedding from cylinders," J. Fluid Mech. **161**, 217–237 (1985).
[2]D. Rockwell and E. Naudascher, "Review—Self-sustaining oscillations of flow past cavities," ASME J. Fluids Eng. **100**, 152–165 (1978).
[3]M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, Cambridge, 2003).
[4]A. Oengören and S. Ziada, "Vorticity shedding and acoustic resonance in an in-line tube bundle, Part II: Acoustic resonance," J. Fluids Struct. **6**, 293–309 (1992).
[5]P. M. Radavich, A. Selamet, and J. M. Novak, "A computational approach for flow-acoustic coupling in closed side branches," J. Acoust. Soc. Am. **109**, 1343–1353 (2001).
[6]K. S. Peat, J. G. Ih, and S. H. Lee, "The acoustic impedance of a circular orifice in grazing flow: Comparison with theory," J. Acoust. Soc. Am. **114**, 3076–3086 (2003).
[7]R. D. Blevins and M. M. Bressler, "Experiments on acoustic resonance in heat exchanger tube bundles," J. Sound Vib. **164**, 503–533 (1993).
[8]R. Parker and D. C. Pryce, "Wake excited resonances in an annular cascade: An experimental investigation," J. Sound Vib. **37**, 247–261 (1974).
[9]S. Ziada, A. Oengören, and A. Vogel, "Acoustic resonance in the inlet scroll of a turbo-compressor," J. Fluids Struct. **16**, 361–373 (2002).
[10]A. Mohany and S. Ziada, "Flow-excited acoustic resonance of two tandem cylinders in cross-flow," J. Fluids Struct. **21**, 103–119 (2005).
[11]L. Cattafesta, D. Williams, C. Rowley, and F. Alvi, "Review of active control of flow-induced cavity resonance," AIAA Paper No. 2003-3567 (2003).
[12]S. A. T. Stoneman, K. Hourigan, A. N. Stokes, and M. C. Welsh, "Resonant sound caused by flow past two plates in tandem in a duct," J. Fluid Mech. **192**, 455–484 (1988).
[13]B. A. W. Simth and B. V. Luloff, "The effect of seat geometry on gate valve noise," ASME J. Pressure Vessel Technol. **122**, 401–407 (2000).
[14]H. H. Heller and D. B. Bliss, "The physical mechanism flow-induced pressure fluctuations in cavities and concepts for the suppression," AIAA Paper No. 75-491 (1975).
[15]S. F. McGrath and D. J. Olinger, "Control of pressure oscillations in deep cavities excited by grazing flow," J. Aircr. **33**, 29–36 (1996).
[16]L. N. Cattafesta III, S. Garg, M. Choudhari, and F. Li, "Active control of flow-induced cavity resonance," AIAA Paper No. 97-1804 (1997).
[17]R. L. Sarno and M. E. Franke, "Suppression of flow-induced pressure oscillations in cavities," J. Aircr. **31**, 90–96 (1994).
[18]X. Y. Huang and D. S. Weaver, "On the active control of shear layer oscillations across a cavity in the presence of pipeline acoustic resonance," J. Fluids Struct. **5**, 207–219 (1991).
[19]S. Ziada, H. Ng, and C. E. Blake, "Flow excited resonance of a confined shallow cavity in low Mach number flow and its control," J. Fluids Struct. **18**, 79–92 (2003).
[20]L. N. Cattafesta III, S. Garg, and D. Shukla, "The development of piezoelectric actuators for active flow control," AIAA J. **39**, 1562–1568 (2001).
[21]L. N. Cattafesta III, J. Mathew, and A. Kurdila, "Modeling and design of piezoelectric actuators for fluid flow control," Trans. Jpn. Soc. Aeronaut. Space Sci. **109**, 1088–1095 (2001).
[22]H. Kook, L. Mongeau, and M. A. Franchek, "Active control of pressure fluctuations due to flow over Helmholtz resonators," J. Sound Vib. **255**, 61–76 (2002).
[23]N. B. Roozen, M. Bockholts, V. E. Pascal, and A. Hirschberg, "Vortex sound in bass-reflex ports of loudspeakers. Part I. Observation of response to harmonic excitation and remedial measures," J. Acoust. Soc. Am. **104**, 1914–1918 (1998).
[24]M. M. Zdravkovich, "Review and classification of various aerodynamic

and hydrodynamic means for suppressing vortex shedding," J. Wind. Eng. Ind. Aerodyn. **7**, 145–189 (1981).
[25]J. Every, R. King, and D. S. Weaver, "Vortex-excited vibrations of cylinders and cables and their suppression," Ocean Eng. **9**, 135–157 (1982).
[26]F. Wilson and J. C. Tinsley, "Vortex load reduction: experiments in optimal helical strake geometry for rigid cylinders," ASME J. Energy Resour. Technol. **111**, 72–76 (1989).
[27]F. B. Hsiao and J. Y. Shyu, "Influence of internal acoustic excitation upon flow passing a circular cylinder," J. Fluids Struct. **5**, 427–442 (1991).
[28]D. R. Williams, H. Mansy, and C. Amato, "The response and symmetry properties of a cylinder wake subjected to localized surface excitation," J. Fluid Mech. **234**, 71–96 (1992).
[29]H. M. Warui and N. Fujisawa, "Feedback control of vortex shedding from a circular cylinder by cross-flow cylinder oscillations," Exp. Fluids **21**, 49–56 (1996).
[30]P. T. Tokumaru and P. E. Dimotakis, "Rotary oscillation control of a cylinder wake," J. Fluid Mech. **224**, 77–90 (1991).
[31]J. R. Filler, P. L. Marston, and W. C. Mih, "Response of the shear layers separating from the circular cylinder to small amplitude rotational oscillations," J. Fluid Mech. **231**, 481–499 (1991).
[32]J. E. Ffowcs Williams and B. C. Zhao, "The active control of vortex shedding," J. Fluids Struct. **3**, 115–122 (1989).
[33]K. Roussopoulos, "Feedback control of vortex shedding at low Reynolds numbers," J. Fluid Mech. **248**, 267–296 (1993).
[34]X. Y. Huang, "Feedback control of vortex shedding from a circular cylinder," Exp. Fluids **20**, 218–224 (1996).
[35]L. Cheng, Y. Zhou, and M. M. Zhang, "Perturbed interaction between vortex shedding and induced vibration," J. Fluid Mech. **17**, 887–901 (2003).
[36]M. M. Zhang, L. Cheng, and Y. Zhou, "Closed-loop-controlled vortex shedding from a flexibly supported square cylinder under different schemes," Phys. Fluids **16**, 1439–1448 (2004).
[37]M. M. Zhang, L. Cheng, and Y. Zhou, "Closed-loop-controlled vortex shedding from a flexibly supported square cylinder under different schemes," Eur. J. Mech. B/Fluids **23**, 189–197 (2004).
[38]M. M. Zhang, L. Cheng, and Y. Zhou, "Control of vortex-induced nonresonance vibration using piezo-ceramic actuators embedded in a structure," Smart Mater. Struct. **14**, 1217–1226 (2005).
[39]L. Cheng, Y. Zhou, and M. M. Zhang, "Controlled vortex-induced vibration on a fix-supported flexible cylinder in crossflow," J. Sound Vib. **292**, 279–299 (2006).
[40]M. M. Zhang, L. Cheng, and Y. Zhou, "Closed-loop controlled vortex-airfoil interactions," Phys. Fluids **18**, 046102 (2006).
[41]M. M. Zhang, L. Cheng, and Y. Zhou, "Control of post-stall airfoil aerodynamics based on surface perturbation," AIAA J. **46**, 2510–2519 (2008).
[42]Y. S. Choy and L. Huang, "Effect of flow on the drumlike silencer," J. Acoust. Soc. Am. **118**, 3307–3085 (2005).
[43]R. Parker and M. C. Welsh, "Effects of sound on flow separation from blunt flat plates," Int. J. Heat Fluid Flow **4**, 113–127 (1983).
[44]M. C. Welsh, A. N. Stokes, and R. Parker, "Flow-resonant sound interaction in a duct containing a plate, Part I: Semi-circular leading edge," J. Sound Vib. **95**, 305–323 (1984).
[45]S. Ziada and S. Shine, "Strouhal numbers of flow-excited acoustic resonance of closed side branches," J. Fluids Struct. **13**, 127–142 (1999).
[46]J. P. Marouzé and L. Cheng, "A feasibility study of active vibration isolation using THUNDER actuators," Smart Mater. Struct. **11**, 854–862 (2002).
[47]O. Durgun and K. Kafali, "Blockage correction," Ocean Eng. **18**, 269–282 (1991).
[48]J. C. Bruggeman, A. Hirschberg, M. E. H. Van Dongen, A. P. J. Wijnands, and J. Gorter, "Self-sustained aero-acoustic pulsations in gas transport systems: Experimental study of the influence of closed side branches," J. Sound Vib. **150**, 371–393 (1991).
[49]P. M. Morse, *Vibration and Sound* (American Institute of Physics for the Acoustical Society of America, New York, 1981).
[50]S. A. Wise, "Displacement properties of RAINBOW and THUNDER piezoelectric actuators," Adv. Weld. Sci. Technol., Proc. Int. Conf. Trends Weld. Res. **69**, 33–38 (1998).
[51]H. Hourigan, M. C. Thompson, and B. T. Tan, "Self-sustained oscillations in flows around long blunt plates," J. Fluids Struct. **15**, 387–398 (2001).
[52]R. Mills, J. Sheridan, and K. Hourigan, "Particle image velocimetry and visualization of natural and forced flow around rectangular cylinders," J. Fluid Mech. **478**, 299–323 (2003).
[53]A. N. Stokes and M. C. Welsh, "Flow-resonant sound interaction in a duct containing a plate, II: Square leading edge," J. Sound Vib. **104**, 55–73 (1986).
[54]C. Welsh, K. Hourigan, L. W. Welch, R. J. Downie, M. C. Thompson, and A. N. Stokes, "Acoustics and experimental methods: The influence of sound on flow and heat transfer," Exp. Therm. Fluid Sci. **3**, 138–152 (1990).
[55]M. Provansal, C. Mathis, and L. Boyer, "Benard–von Kámán instability: Transient and forced regimes," J. Fluid Mech. **182**, 1–22 (1987).
[56]H. Sakamoto, K. Tan, and H. Haniu, "An optimum suppression of fluid forces by controlling a shear layer separated from a square prism," J. Fluids Eng. **113**, 183–189 (1991).

Green's function approximation from cross-correlation of active sources in the ocean

Laura A. Brooks[a] and Peter Gerstoft
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093

(Received 15 January 2009; revised 1 May 2009; accepted 2 May 2009)

Green's function approximation via ocean noise cross-correlation, referred to here as ocean acoustic interferometry, has been demonstrated experimentally for passive noise sources. Active sources offer the advantages of higher frequencies, controllability, and continuous monitoring. Experimental ocean acoustic interferometry is described here for two active source configurations: a source lowered vertically and one towed horizontally. Results are compared and contrasted with cross-correlations of passive noise. The results, in particular, differences between the empirical Green's function estimates and simulated Green's functions, are explained with reference to theory and simulations. Approximation of direct paths is shown to be consistently good for each source configuration. Secondary (surface reflection) paths are shown to be more accurate for hydrophones with a greater horizontal separation.
 [DOI: 10.1121/1.3143143]

PACS number(s): 43.30.Pc [RCG] Pages: 46–55

I. INTRODUCTION

Using passive noise, good estimates of the acoustic Green's function between two points can be determined from cross-correlation of sound in the ocean,[1–6] a technique referred to as ocean acoustic interferometry (OAI). This concept has been applied across a broad range of physical scales in different media including ultrasonic noise,[7] ambient noise in a homogeneous medium,[8] seismic noise,[9–13] moon-seismic noise,[14] human skeletal muscle noise,[15] and general fluid-solid systems.[16] Noise interferometry is good at obtaining the travel time part of the Green's function, and it is common to estimate both group and phase velocities.[12,13] Green's function estimates can be useful for inferring information about the environment through which the acoustic transmission takes place.[9,14] Since noise is ever-present, noise interferometry is also potentially useful for the continuous monitoring of changes in the environment.[15] Green's function amplitudes from noise are more difficult to extract,[3] as will be shown in Sec. II; however, some progress has been made in this area.[17,18] Noise interferometry extracts the Green's function directly from the data, which requires time-averaging while the active or passive noise builds up. A stationary environment is therefore assumed. Media can, however, be inhomogeneous (including range dependence). In fact, a more complicated environment is beneficial, as the scattering tends to create a more isotropic field.

Based on theoretical and numerical analysis,[5,19,20] active source interferometry has been suggested as an alternative method for Green's function estimation. Although active OAI has similar difficulties with extracting amplitude information as passive techniques, it does present a number of advantages, including the use of higher frequencies (larger bandwidth gives sharper arrivals), controllability, and continuous monitoring. Greater knowledge of the contributing sources also means that more realistic simulated data can be produced.

To obtain an accurate representation of the Green's function, sound waves must propagate isotropically.[21] An active set of sources surrounding the receivers could potentially achieve this. However, only sources that emit acoustic paths passing through both receivers contribute to the Green's function.[2,3] These sources are all located in the end-fire plane, where *end-fire* is defined here as within the vertical plane containing both receivers (i.e., the plane defined by the vertical and horizontal arrays). Due to the technical complexity of surrounding the receivers completely, two simpler active source configurations are investigated here: (1) a source lowered through the water column at end-fire to a set of receivers, a configuration previously examined theoretically and through simulation;[1,5] (2) a source towed at a constant depth along the end-fire direction. A single empirical Green's function (EGF) is extracted between any two receivers in the array, thus there is no array processing performed. Active sources with similar geometries have been used elsewhere as a guide source to reduce the environment effects.[22] This is contrary to the approach taken here where unknown details of the propagation region are actually included in the EGF estimate.

To understand the salient features of EGF extraction in relation to the environment and source/receiver geometry, it is analyzed in Sec. II using ray theory and stationary phase. Although a homogeneous medium is assumed in the theory, the stationary phase argument does generalize to a heterogeneous medium, which may include varying sound speed profiles (SSPs),[19] as well as range dependent characteristics. Wavenumber integration[23] is used in Sec. III for the accurate modeling of the received waveforms in the experimental range independent environment.

[a]Also at the School of Mechanical Engineering, University of Adelaide, Australia. Present address: Institute of Geophysics, Victoria University of Wellington, New Zealand. Author to whom correspondence should be addressed. Electronic mail: laura.brooks@vuw.ac.nz

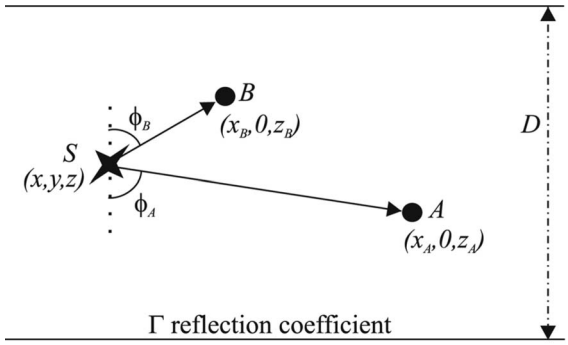


FIG. 1. Source-receiver geometry and notation. Receivers A and B define the $y=0$ plane, and source S is located within the waveguide of depth D, but is otherwise unrestricted.

The Shallow Water 2006 (SW06) experiments provided an opportunity to collect experimental data on a single array of ocean hydrophones for both active and passive source OAI, and as such, provide a unique set of data for analysis and comparison of the different source types. In Sec. III, EGFs obtained using the two active source configurations (source lowering and towed source) are compared and contrasted with EGFs of noise emitted by the ship from which the source was lowered, and also with EGFs from a noise field dominated by waves and shipping. The active source experiments were performed in seas with 2–2.5 m swell, residual effects from the passing of a tropical storm the day before.[6,24] Although conditions hindered controllability of the experiments, and the extracted EGFs may have fluctuated more than usual, meaningful results were still obtained. The relative merits of different source types and receiver locations are evaluated through examination of the experimental results, and by relating these findings to the theory.

II. BACKGROUND THEORY

Consider the isovelocity waveguide depicted in Fig. 1. The x, y, and z directions are defined as the horizontal axis, the axis in-and-out of the page, and the vertical axis, respectively. Cross-correlation of the signals received at A and B from a source at S yields

$$C_{AB}(\omega) = \rho_s^2 |S(\omega)|^2 G(\mathbf{r}_A, \mathbf{r}_S) G^*(\mathbf{r}_B, \mathbf{r}_S), \tag{1}$$

where ω is angular frequency, $S(\omega)$ is the source spectrum, ρ_s is the density of the medium, $G(\mathbf{r}_\psi, \mathbf{r}_S)$ is the Green's function between the source S, and receiver ψ, and * denotes the complex conjugate.

The sum of the cross-correlations over a set of sources is[5,19]

$$C_{AB}(\omega) = |\rho_s S(\omega)|^2 n \int G(\mathbf{r}_A, \mathbf{r}_S) G^*(\mathbf{r}_B, \mathbf{r}_S) dS, \tag{2}$$

where n is the number of sources per unit length (line source), area (planar source), or volume (volume source), and the integral is over the source line, plane, or volume.

In OAI, both the causal and acausal Green's functions, $G(\mathbf{r}_A, \mathbf{r}_B)$ and $G^*(\mathbf{r}_A, \mathbf{r}_B)$, respectively, are extracted from the cross-correlation, $C_{AB}(\omega)$:

$$G(\mathbf{r}_A, \mathbf{r}_B) - G^*(\mathbf{r}_A, \mathbf{r}_B) = f(C_{AB}(\omega)). \tag{3}$$

If the source is only on one side of both A and B, then only the causal Green's function is extracted. The relating function f depends on the physical properties of the medium and the source distribution, as well as the dimension of the wave propagation. Its origin and form are explored here from a stationary phase perspective for four different source types:

(1) active source lowered vertically over the depth of the waveguide at a location end-fire to the array, modeled as a vertical line of end-fire sources ($\int dS \sim \int dz$);
(2) active source towed along the end-fire direction at a constant depth z, modeled as a horizontal line of sources at end-fire ($\int dS \sim \int dx$);
(3) stationary ship source, modeled as an "extended" point source [$\int dS \sim \int \delta(x,y) dS$]; and
(4) ambient noise field, modeled as a horizontal plane of sources at a shallow depth z ($\int dS \sim \int\int dx dy$).

The first two cases, which are of main interest here, are described in Sec. II A. The two latter cases, which are included only for comparative purposes, are briefly described in Sec. II B.

A. Cross-correlations for active sources

The phase of the integral term in Eq. (2) oscillates rapidly relative to the amplitude, and hence the integrand averages to almost zero except at points where its phase term is stationary (i.e., where the phase term has an extremum). The integrand can therefore be evaluated via the method of stationary phase; the integral is estimated in the neighborhood of the points where the phase term is stationary (at these locations the source is said to be located at a stationary point) and the contributions are then summed over all the stationary points. Stationary phase solutions to Eq. (2) have been derived elsewhere for various source configurations,[5,19,21,25] and hence full derivations are not presented here. The stationary phase solution assuming three-dimensional (3D)-wave propagation, for a vertical line of sources at end-fire, source type 1, is[5]

$$C_{AB}^1(\omega) = in|S(\omega)|^2 \sum_{z_{\mathrm{st}}} \left(\frac{\Gamma^{b_{A,\mathrm{st}}+b_{B,\mathrm{st}}} \rho_s^2 G_f(R(z_{\mathrm{st}}))}{\sin \phi_{\mathrm{st}}} \sqrt{\frac{\xi(z_{\mathrm{st}})c}{-8\pi i\omega}} \right), \tag{4}$$

and for a horizontal line of sources at end-fire, source type 2, is[19,21]

$$C_{AB}^2(\omega) = in|S(\omega)|^2 \sum_{x_{\mathrm{st}}} \left(\frac{\Gamma^{b_{A,\mathrm{st}}+b_{B,\mathrm{st}}} \rho_s^2 G_f(R(x_{\mathrm{st}}))}{\cos \phi_{\mathrm{st}}} \sqrt{\frac{\xi(x_{\mathrm{st}})c}{-8\pi i\omega}} \right), \tag{5}$$

where z_{st} and x_{st} are the vertical and horizontal end-fire stationary points, respectively, Γ is the bottom reflection coefficient, $b_{\psi,\mathrm{st}}$ is the number of bottom reflections for the path between the source S, located at stationary point p_{st} (either z_{st} or x_{st}), and the receiver ψ (where $\psi=A$ or B), $R(p_{\mathrm{st}})$ is the difference in path lengths from the source, at p_{st}, to each of receivers A and B, $G_f(R)=e^{ikR}/4\pi R$ is the 3D Green's func-

tion within a homogeneous medium, ϕ_{st} is the acute angle between the ray path and the vertical (see Fig. 1), $\xi = (1/L_B) - (1/L_A)$, L_ψ is the length of the given path S and ψ, and c is the speed of sound in the medium. The 3D Green's function within a homogeneous medium, $G_f(R)$, differs from the true Green's function for a particular path between A and B, $C(R)$, in that it does not incorporate the path dependent amplitude reduction due to bottom reflections:[5]

$$G(R(p_{st})) = \Gamma^{b_{st}} G_f(R(p_{st})), \tag{6}$$

where b_{st} is the number of bottom reflections for the arrival between A and B corresponding to the stationary point p_{st}. The solutions for the line sources, Eqs. (4) and (5) differ only in the trigonometric function of the acute ray angle ($1/\sin\phi_s$ versus $1/\cos\phi_s$), and the locations of stationarity.

The cross-correlation sums in Eqs. (4) and (5) are achieved experimentally by lowering a source through the water column (C^1_{AB}), or towing it at end-fire to a hydrophone array (C^2_{AB}), and then summing the cross-correlations throughout the period of source movement. An approximation to the Green's function, the EGF, g^{emp}, is then obtained from the cross-correlations:

$$g^{emp} = \sqrt{i\omega} C^{obs}_{AB}, \tag{7}$$

where C^{obs}_{AB} is either C^1_{AB} (vertical) or C^2_{AB} (horizontal). The sources are only on one side of the array, and hence only the causal part of the Green's function is approximated. The constants (n, ρ_s, c, π, and numeric factors) and frequency dependent source term $S(\omega)$ of Eqs. (4) and (5) can also be accounted for when comparing g^{emp} with the true Green's function, but path dependent factors ($\Gamma^{b_A+b_B}$, ϕ_s, L_A, and L_B) are more difficult to account for and are therefore neglected. Hence, the EGF obtained will not give correct amplitudes but should provide accurate arrival times.

Spurious arrivals, defined as peaks in the cross-correlation function at times not corresponding to Green's function path travel times, can occur for each source geometry. For the horizontal line configuration, spurious arrivals will result due to stationary phase contributions from cross-correlations between waves that initially undergo a surface reflection and ones that do not (for an isovelocity water column, one wave departs at an angle of ϕ from the horizontal, and the other departs at an angle of $-\phi$).[3] If the source is close to the sea surface, the spurious peaks converge to the same time delay as the true Green's function paths; however, they are π out of phase and will therefore result in shading of the Green's function.[3] For the vertical line configuration, spurious arrivals will result when the line integral does not extend from the sea surface through the ocean and underlying sediments to the acoustic penetration depth.[5]

B. Stationary ship and ambient noise

For a point source (source type 3), Eq. (2) simplifies to

$$C^3_{AB}(\omega) = \frac{|\rho S(\omega)|^2 \Gamma^{b_A+b_B} e^{ik(L_A-L_B)}}{16\pi^2 L_A L_B}. \tag{8}$$

In general $L_A - L_B$ is less than the inter-receiver path length and therefore arrival times are underestimated. Although the stationary ship source is larger than a point source, the area of integration in Eq. (2) is small, and therefore it is not a true Green's function estimate. However, if the ship is close to a stationary path it may provide a good approximation of that path:

$$g^{emp} = C^{obs}_{AB}, \tag{9}$$

where C^{obs}_{AB} is C^3_{AB}. For a particular path p, as the ship approaches the corresponding stationary point, the component of the EGF corresponding to path p approaches the path p component of the Green's function between A and B.

Ship noise cross-correlations will only converge to the arrival structure of the Green's function when averaged over several ship tracks, hence the consideration of ship and wave dominated ambient noise (source type 4). For a horizontal plane of sources the stationary phase solution to Eq. (2) is (Refs. 3 and 6)

$$C^4_{AB}(\omega) = in|S(\omega)|^2 \sum_{\chi_{st}} \left(\frac{\Gamma^{b_{A,st}+b_{B,st}} c\rho_s}{2\omega \cos\phi_{st}} G_f(R(\chi_{st})) \right), \tag{10}$$

where χ_{st} are the horizontal planar stationary points and all other parameters are the same as for Eqs. (4) and (5).

Since this source distribution surrounds the hydrophones, both the causal and acausal Green's functions are approximated:

$$g^{emp} - g^{emp*} = \omega C^{obs}_{AB}. \tag{11}$$

III. EXPERIMENT

A. Data collection

The SW06 experiments were an Office of Naval Research sponsored set of low and medium frequency acoustic experiments conducted off the northeast coast of the United States. The acoustic data considered here were collected on an L-shaped array [geometry shown in Fig. 2(a)] located in water approximately 70 m deep. This allowed for many hydrophone pair geometries. Hydrophones 1–10 (H-1–H-10) constituted the vertical line array (VLA). They were evenly spaced at 5.95 m intervals, the lowest, H-10, being 4.65 m above the seafloor. H-13–H-32 constituted the 256.43 m long seafloor mounted horizontal line array (HLA). H-13 was located 7.795 m from the VLA. Spacing between subsequent HLA hydrophones decreased from 20.32 m at this end to 8.33 m at the array tail.

Data from four source types were recorded:

(1) 1200–2900 Hz one second duration continuous linear frequency modulated (LFM) signal emitted by an omnidirectional source lowered from 9.8–60 m at a constant rate of 1 m/min, at a location 466 m from the VLA, in the end-fire plane [see source lowering geometry and location in Figs. 2(a) and 2(b)];
(2) 1200–2900 Hz one second duration continuous LFM signal emitted by an omnidirectional source held at 10 m depth towed at 1 kn toward the array in the end-fire plane, from a distance of 1.5 km from the VLA, to a location mid-way between H-16 and H-30 [see towed source geometry and location in Figs. 2(a) and 2(b)];

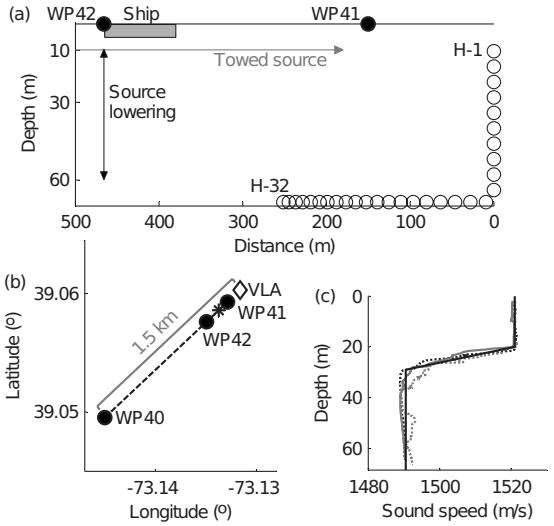


FIG. 2. (a) Source and receiver geometries: array hydrophones shown as circles, and ship location is during source lowering. (b) Plan view of array (VLA labeled, asterisk marks far end of HLA) and source geometries (source towed from WP40 to WP41, and lowered at WP42). (c) SSPs from CTDs 42–44 (black dotted, gray solid, and gray dotted lines, respectively), and assumed SSP for modeling (black solid line).

(3) 20–100 Hz noise generated by the deployment vessel *R*/*V* Knorr [location shown in Fig. 2(a)] during the source lowering experiment; and
(4) 20–100 Hz ship and wave dominated ambient noise.[6]

The ship and wave dominated noise data, source type 4, were collected during tropical storm Ernesto over the entire day of September 2, and data from source types 1–3 were collected on the afternoon of September 3, 2006. There was little wind, but there was a residual swell of 2–2.5 m, as well as strong inhomogeneity in the ocean due to the previous day's storm. These made it difficult to move the active source along the desired tracks. SSPs were recorded from three CTD-casts (conductivity, temperature, and depth) obtained before, during, and after the experiments on September 3, as shown in Fig. 2(c), along with a simpler SSP used for simulations. The September 2 storm would have increased mixing in the upper layer and hence the SSP would have changed throughout the day, the most noticeable aspect of which would have been the formation of a steeper thermocline. However, the adverse weather conditions meant that no CTD-casts could be obtained during this period. The September 2 data are therefore analyzed here using simulated travel times derived from the September 3 model SSP.

B. Data analysis

Active source data from all hydrophone pairs were bandpass filtered to 1200–2900 Hz, and ship and wave noise data were bandpass filtered to 20–100 Hz. The ship and wave dominated ambient noise data were then one-bit normalized in the time domain (i.e., amplitude was discarded but sign, or phase, of the waveform was retained), bandpass filtered once again, and spectrally whitened by inversely weighting the frequency domain data with a smooth version of their amplitude spectra within the bandpass frequency. The active source and stationary ship data did not require normalization since variations in the source amplitude and phase characteristics were negligible throughout each experiment.

The preprocessed data were cross-correlated over short time intervals, and then summed over the period of collection for each source type. As specified in Sec. II, a raw summed cross-correlation [see Eqs. (4), (5), and (10)] yields a phase and amplitude shaded Green's function approximation. Corrections to the phase shading and frequency dependent components for source types 1 and 2 were made. Although source type 3 also has a phase shift, it is geometry dependent, due to the length discrepancy in the exponential of Eq. (8), and therefore no correction factors were applied. Inclusion of the appropriate phase correction is, henceforth, implicit in the term "cross-correlation."

The cross-correlation sum corresponds to the EGF (see Sec. II). The normalized EGFs of the cross-correlations between H-30 and all other hydrophones are plotted in Fig. 3 overlying a pseudocolor plot of their envelopes for the four source types in Sec. III A. Simulated direct (D), surface reflected (S), surface-bottom reflected (SB, VLA only), and surface-bottom-surface reflected (SBS) path travel times, which were determined using OASES,[23] are overlaid as dash-dotted lines for comparison. The simulations use the simplified SSP of Fig. 2(c), and assume a 20 m deep sediment layer (c=1650 m/s, α=2.7 dB/λ, ρ=1.85 g/cm^3) overlying a basement (c=1900 m/s, α=2.0 dB/λ, ρ=2.0 g/cm^3). These geoacoustic parameters, inferred at the active source frequencies from geoacoustic inversion results at the array site,[26] correspond to a critical angle of 25°. Lower values of the sediment attenuation, which would be expected in the general region[27] or at lower frequencies, show small peaks in the Green's function corresponding to reflections from the sediment-basement interface (not shown).

The ship dominated ambient noise results, as shown in Fig. 3(d), have both causal and acausal components because the sound comes from all directions, though only the first 0.05 s of the acausal signal is shown here. Results from the other three configurations, shown in Figs. 3(a)–3(c), have sound traveling in one direction only, left to right from the perspective of Fig. 2(a), and therefore produce a one-sided EGF. The stationary ship EGF, (c), and ship and wave dominated ambient noise EGF, (d), show broader peaks than the active source results, (a) and (b), due to the lower frequencies of the ship (20–100 Hz compared to the 1200–2900 Hz active source frequencies).

The EGF envelopes for all source types, as shown in Fig. 3, exhibit distinct peaks at times agreeing with the simulated direct inter-hydrophone travel times. The times corresponding to these peaks are compared in Fig. 4(a). Simulated travel times are subtracted from these times, and the resulting time differences are shown in Fig. 4(b). Minimal variations are seen for all HLA hydrophone combinations, though the stationary ship peak time differences (c) are generally greater than the others, which is due to the discrete nature of the source location; no signals from the source pass through

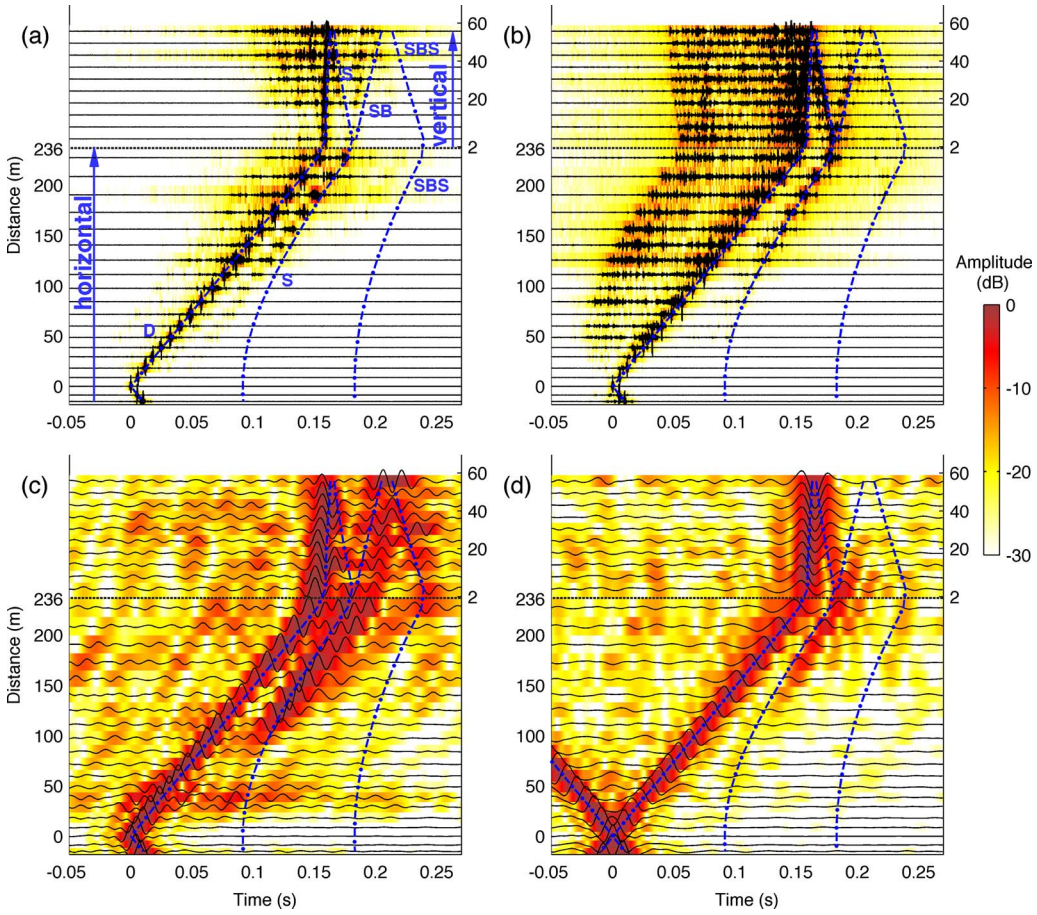


FIG. 3. (Color online) EGFs (black, normalized to their maximum value) between H-30 and all other hydrophones overlying pseudocolor plots of their envelopes (dB relative to maximum value) for (a) source lowering, (b) towed source, (c) stationary ship, and (d) causal ambient noise. The lower traces (below the dashed line) are from cross-correlations with HLA hydrophones; their distance from H-30 is on the left axis. The upper traces are from cross-correlations with VLA hydrophones; their vertical distance from the seafloor is on the right axis The dash-dot lines are simulated travel times.

the location of the first hydrophone, on their way to the second hydrophone, with a direct path in between.

The variation in the EGF peak times corresponding to the direct arrivals between the hydrophones are notably larger for the VLA. Due to their location, the VLA hydrophones are more sensitive to environmental variations and more susceptible to movement than their HLA counterparts.

The EGF peaks corresponding to the surface reflected arrivals in Fig. 3 show more variation than the direct path peaks, but due to the lower grazing angles, are consistently more accurate for hydrophones with a greater horizontal separation. The towed source and ambient noise results, shown in Figs. 3(b) and 3(d), respectively, show a surface reflection peak for all ranges larger than 40 m from H-30. The source lowering, shown in Fig. 3(a), exhibit peaks at slightly early times for ranges larger than 150 m from H-30. For ranges less than 150 m, peak times diverge from the simulated values. The stationary ship results, shown in Fig. 3(c), show an arrival peak for ranges greater than 100 m, but the VLA arrivals are less clear.

The amplitudes of the EGF peaks are greatest, relative to the background noise, for the active source cross-correlations. This is due to high levels of coherently propagating noise which result from the close proximity of the source and the even distribution of the source over the active source line integrals. Unlike the other source configurations, the towed source results in Fig. 3(b) show a peak in the EGF at an arrival time earlier than the direct path for all distances. The reason for this apparent early arrival is explained in Sec. III B 2.

To explain the features of the EGFs of Fig. 3, OAI data for one hydrophone pair, H-30 and H-5, will be examined in detail for each source type in the coming sections.

1. Source lowered vertically

The theoretical vertical line source description in Sec. II assumes a set of sources that is uniformly distributed along the line.[5] The single source used here was slowly lowered, but was only at one location at any time. The line source configuration was therefore obtained by cross-correlating data over short time intervals and summing these cross-correlations. Thus, while the cross-correlations over depth are described here, it is only the summed cross-correlations which are used to approximate the EGF. This EGF approximation could instead have been implemented as one large cross-correlation.

The geometry of the source lowering, as well as the stationary point travel paths and surface and bottom sources that converge to the stationary points for H-5 and H-30, are

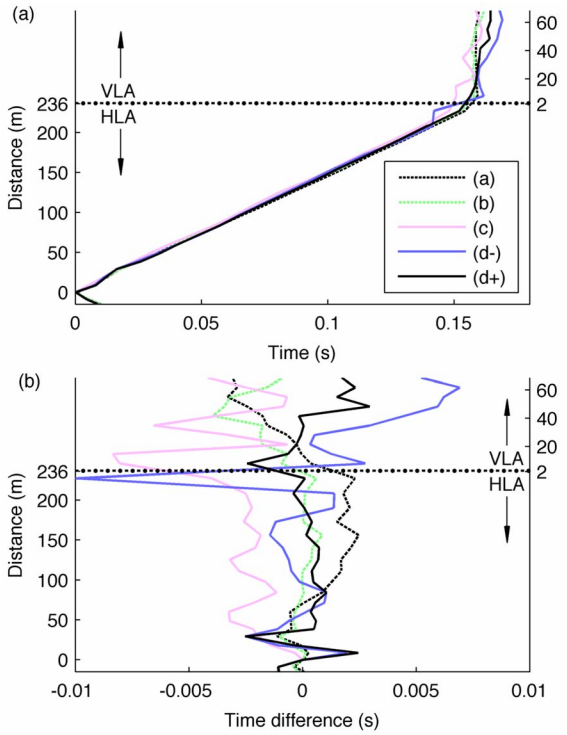


FIG. 4. (Color online) (a) Travel times corresponding to the direct path arrival EGF envelope peaks of Fig. 3 as a function of hydrophone number, and (b) differences between these and the simulated travel times. Legend: (a) source lowering, (b) towed source, (c) stationary ship, (d+) causal ambient noise, and (d−) acausal ambient noise.

shown in Fig. 5. Ideally a source should be lowered through the water column and sediment,[5,19] but due to experimental restrictions the source could only be lowered from 9.8–60 m (8.5 m above the seafloor).

Correlation gathers of 100 s duration between H-5 and

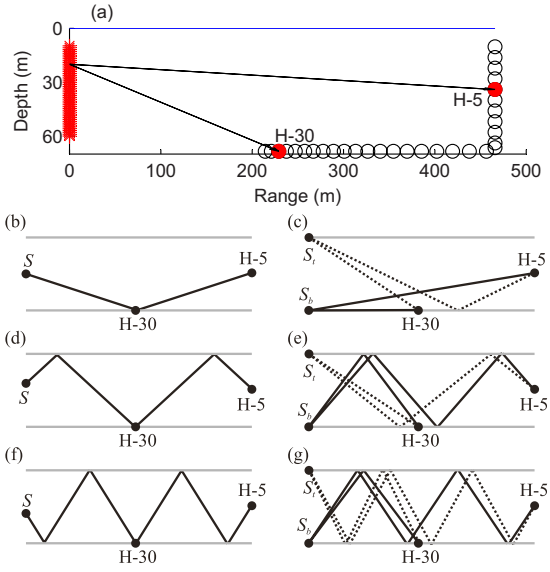


FIG. 5. (Color online) Source lowering. (a) Source (far left) is lowered from 9.8–60 m, and signals are recorded on H-5 and H-30 (solid circles). (b)–(g) Source-receiver geometry and stationary point paths for (b) direct, (d) surface, and (f) surface-bottom paths. (c), (e), and (g) are the surface (S_t) and bottom (S_b) source to receiver paths that converge to the stationary point paths in (b), (d), and (f), respectively.

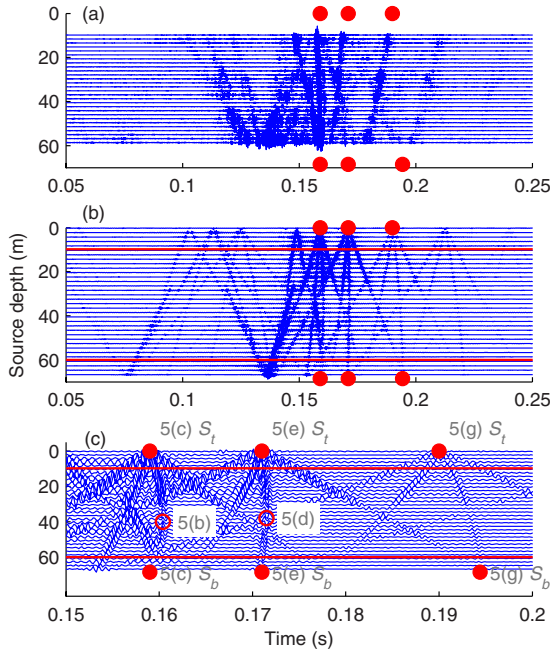


FIG. 6. (Color online) Correlation gathers for source lowering: (a) experimental, (b) simulated, and (c) extract of simulated data. The solid circles correspond to time differences between the surface and bottom source paths to H-5 and H-30 [Figs. 5(c), 5(e), and 5(g)], and the transparent circles in (c) correspond to the direct and surface reflection stationary points in Figs. 5(b) and 5(d).

H-30 are shown as a function of depth for experimental data in Fig. 6(a), and data simulated using OASES in Figs. 6(b) and 6(c). Cross-correlation peaks occur at the time differences between paths from the source to each hydrophone. The direct path between H-30 and H-5 is extracted from the cross-correlation of the direct path from the source to H-30 and the bottom reflected path to H-5 [see Figs. 5(b) and 5(c)]. The simulated time differences between these paths for sources at the top or bottom of the waveguide [Fig. 5(c)] are depicted in Fig. 6 as the first set of solid circles (0.16 s). The curve of maxima connecting these circles corresponds to the time difference between these paths for each source depth. The time difference increases to a stationary point at ~40 m depth [circle in Fig. 6(c)]. This stationary point occurs when the path to the second hydrophone (H-5) passes through the first hydrophone (H-30) [Fig. 5(b)]; i.e., the two paths have a travel time difference equivalent to the direct arrival between H-30 and H-5.

The surface and surface-bottom reflected arrivals between two hydrophones can be analyzed similar to the direct arrival as shown in Fig. 5 and marked with circles in Fig. 6.

Contributions to the EGF at the surface and bottom will generally cancel. Consider the four paths that converge at the surface of the waveguide to the direct path to H-30 and bottom reflected to H-5 [first circle (c) S_t in Fig. 6(c)], as shown in detail with schematics of the paths that converge to this point in Fig. 7. The cross-correlations of paths (b) and (e) are in phase (one surface reflection each) and their amplitudes are equal in amplitude at the convergence point. Since the path length difference of (b) increases towards the surface, and that of (e) decreases towards the surface, and their rates of change are also the same, the cross-correlation peaks due

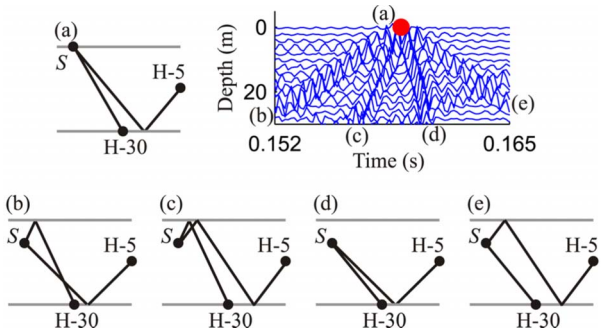


FIG. 7. (Color online) Close-up view of the simulated correlation gathers [Fig. 6(c)], showing the direct path to H-30 and bottom reflected to H-5 surface convergence point (a), and the four sets of paths that converge to this point (b)–(e).

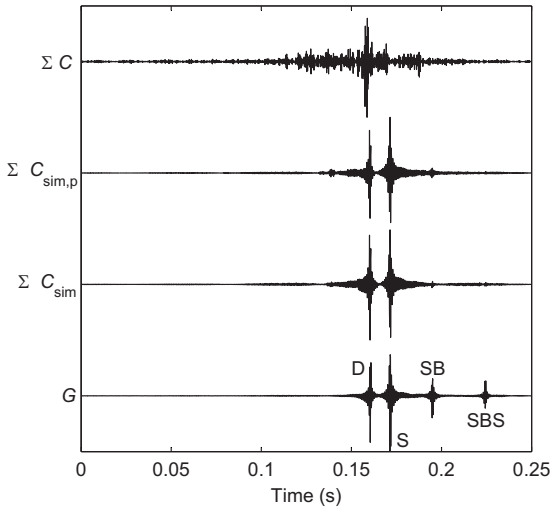


FIG. 8. H-5 to H-30 EGF from the source lowering experiment (ΣC) is compared to the simulated cross-correlations summed from 9.8–60 m ($\Sigma C_{\mathrm{sim},p}$), the simulated cross-correlations summed over the entire waveguide (ΣC_{sim}), and the simulated source shaded Green's function (G). Green's function shows direct (D), surface (S), surface-bottom (SB), and surface-bottom-surface (SBS) paths.

to these combinations transfer smoothly from one path combination to the other, and therefore there is no spurious arrival. A similar argument holds for other sets of paths at the surface and bottom [including paths (c) and (d) in Fig. 7], and therefore there is generally no peak in the summed cross-correlations at these convergence points. Exceptions to this generality can occur when the total number of surface and bottom reflections of the two converging paths are not identical. Since the amplitude of the surface and bottom reflections coefficients is generally less than unity, one of the converging paths will likely be stronger than the other, creating a residual peak in the cross-correlation at this point. The amplitude of this peak will often be negligible, but sometimes it will be large enough to have an impact on the results, as will be shown for two convergence points in the experimental data here.

The simulated and experimental data of Fig. 6 differ in three ways:

(1) The experimental data are less sharp, likely due to meter high waves, which caused both the source and waveguide depth to oscillate.
(2) There are variations in amplitudes for different path combinations, with some path combinations more affected than others. Likely reasons are that the bottom reflection coefficient, or sediment properties, of the simulation are only an estimation, and that due to ocean waves, the angle of the surface reflected signal (relative to horizontal) would be time dependent and the signal would be scattered. Most paths depend on some power of surface and bottom reflection, see Eqs. (4)–(10), and higher order paths are more sensitive to these reflection coefficients.
(3) Peak times differ slightly, likely due to slight mismatch in SSP and water column depth between the experimental and simulated environments.

The amplitude differences between the simulated and experimental data (item 2) could potentially be used to invert for surface and bottom reflection.

Having constructed the cross-correlation at each point in depth, the EGF can then be extracted by summing these contributions.[3,5,19] The cross-correlations were summed over depth and the resulting EGFs are compared with both simulated summed data and the source shaded Green's function in Fig. 8. The sum of the simulated data over the waveguide, ΣC_{sim}, shows direct, surface reflected, and surface-bottom reflected peaks at correct Green's function, G, time lags. The amplitudes are different, as explained in Sec. II. The significantly smaller amplitude of the surface-bottom reflected path in ΣC_{sim} when compared to the Green's function is due to losses from the many boundary interactions. The Green's function for this path has only one surface and one bottom reflection, but the two paths that are cross-correlated have three surface and four bottom reflections between them [see Eqs. (4) and (6)]. The sum of the simulated data from 9.8–60 m only, $\Sigma C_{\mathrm{sim},p}$, has small spurious peaks at ~0.14 s, which is earlier than the direct arrival. These are due to the 8.5 m gap in cross-correlations at the waveguide bottom.

While the experimental data matches well the direct arrival, both the surface and surface-bottom reflected arrivals appear early in Figs. 3 and 8. This is because in the experimental cross-correlation [Fig. 6(a)], complete cancellation at the near surface cross-correlation endpoints (second and third solid circles) does not occur. The amplitude of the surface reflection coefficient is likely less than unity, resulting in one arrival dominating and therefore contributions near the surface convergence points (circles in Fig. 6).

2. Source towed horizontally

Following the same reasoning as with the source lowering (Sec. III B 1), short time cross-correlations were summed as the source was towed from WP40 to WP41. Tapering was incorporated to cross-correlations near WP40 to minimize end-effects. Ideally the ship would have continued past WP41, over the VLA and 1.5 km further, with tapering at both ends; however, the ship had to stop at WP41 as it could not sail above the VLA. Tapering at this endpoint is difficult as this reduces the amplitudes of nearby stationary points.

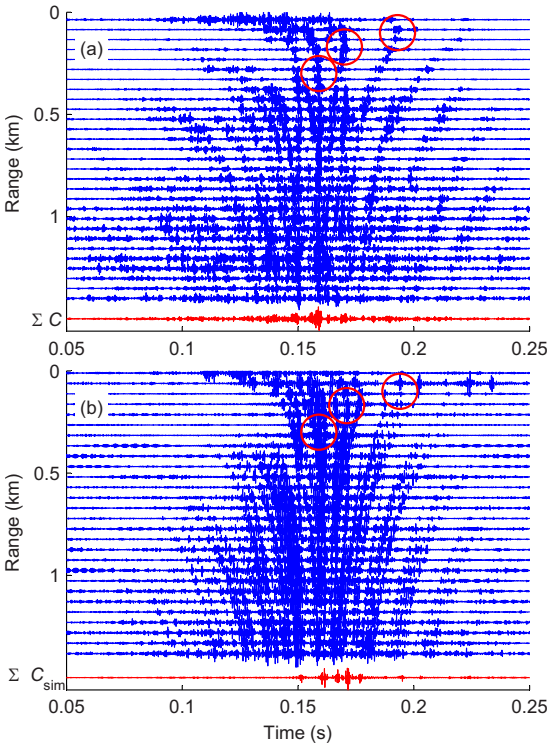


FIG. 9. (Color online) Towed source correlation gathers: (a) experimental and (b) simulated data, as a function of range (0–1.4 km) from H-30. The summed cross-correlations ΣC and ΣC_{sim} are at the bottom of each plot.

The cross-correlations at this endpoint are visible in Fig. 3(b) as peaks in amplitude at times less than 0.05 s corresponding to the path length difference to each hydrophone at this point. These times are significantly less than the expected arrival times of D, S, and B.

The H-5 and H-30 100 s long correlation gathers are shown in Fig. 9 as a function of range for experimental data, and data simulated using OASES. The direct, surface reflection, and surface-bottom reflection stationary points (the turning points where the travel time for an arrival is a maximum) are circled. The corresponding stationary phase paths are shown in Figs. 10(a)–10(c). The stationary points occur within the first few hundred meters; the cross-correlation peak times increase rapidly in range to these points (Fig. 9),

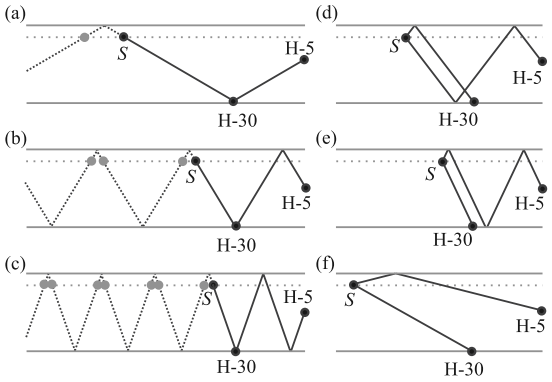


FIG. 10. (a)–(c) Source-receiver geometry and stationary point paths for (a) direct path, (b) surface reflection path, and (c) surface-bottom reflection path. The gray circles represent weaker stationary points. (d)–(f) Stationary phase geometries that yield spurious arrivals.

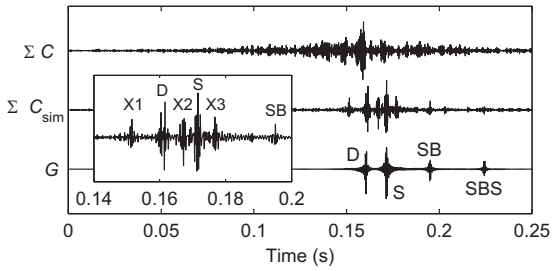


FIG. 11. H-5 to H-30 EGF from the toward source experiment (ΣC) is compared to the simulated summed cross-correlations (ΣC_{sim}), and the simulated source shaded H-5 to H-30 Green's function (G). Green's function shows direct (D), surface (S), surface-bottom, (SB) and surface-bottom-surface (SBS) paths. An enlarged view of ΣC_{sim} from 0.14–0.2 s, showing inter-hydrophone (D, S, and SB) and spurious (X1–X3) arrivals, is inset.

and then asymptote towards a far-field fixed value. Higher multiples [gray in Figs. 10(a)–10(c)] also yield stationary points, but due to increased boundary interactions, their amplitudes are much smaller, and are not visible in Fig. 9.

The H-5 to H-30 EGF for the source towed horizontally (the experimental cross-correlations summed over range), ΣC, is compared in Fig. 11 with the simulated summed cross-correlations, ΣC_{sim}, and the simulated source shaded Green's function, G. The simulated cross-correlation sum shows direct (D), surface reflected (S), and surface-bottom (SB) arrival peaks at correct lag times. The experimental data have stationary points [Fig. 9(a)] that yield arrival peaks (Fig. 11) at times slightly less than the simulated direct and surface reflected arrivals. This is likely due to mismatch between the experimental and simulated water depths and SSPs. The experimental summed cross-correlation also has a higher noise level, which is likely due to convergence difficulties near zero range, where the data are sensitive to tapering and the chosen physical endpoint.

Both the experimental and simulated cross-correlation sums exhibit numerous high amplitude spurious arrivals. For example, consider the two spurious arrivals (X2) and (X3) that are visible in the summed simulated cross-correlations (Fig. 11) before and after the surface reflected arrival (S). These are the result of stationary points corresponding to non-Green's function arrivals [paths in Figs. 10(d) and 10(e)], as explained in Sec. II.

The arrivals and stationary points that create these peaks are visible in Fig. 9(b). Peaks corresponding to the time difference in the direct path to H-30 and the bottom-surface reflected path to H-5 (with the surface reflection stationary point at 0.17 s circled) are flanked by a set of arrivals at slightly earlier and later times. These additional arrivals, which are due to the cross-correlation of a wave which initially undergoes a surface reflection with one that does not, have stationary points corresponding to the geometry of Figs. 10(d) and 10(e), and hence the spurious arrivals (X2) and (X3) are apparent in the summed cross-correlations of Fig. 11.

A significant peak (X1) is apparent in both the experimental and simulated cross-correlations at 0.15 s, prior to the direct path (D) arrival. This spurious arrival is due to a stationary phase contribution from the cross-correlation of the direct path to H-30 and surface reflection to H-5 [Fig. 10(f)].

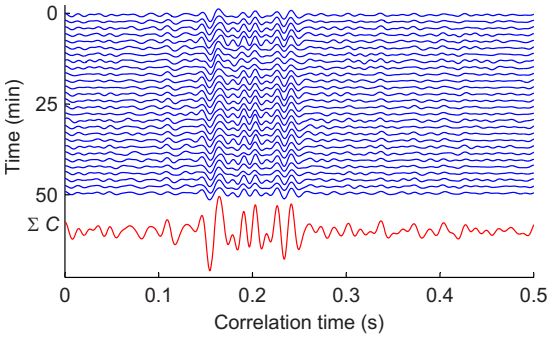


FIG. 12. (Color online) H-30 to H-5 EGFs from stationary ship data as a function of time during the source lowering event. The summed EGF (ΣC) is shown underneath.

The peaks in the simulated and experimental data exist only in this varying SSP environment. Simulated cross-correlations (not shown) for an isovelocity (1500 m/s) waveguide with the same geometry do not show this peak, because such a stationary phase geometry does not exist when considering straight line paths only [the schematic of straight line paths in Fig. 10(f) comes close to, but does not satisfy, the equal departure angle requirement; the path to H-5 always departs the source at an angle closer to the horizontal than the path to H-30].

3. Stationary ship noise and ambient noise

The cross-correlated data from the stationary ship varied little with time (see Fig. 12). Due to the crude approximation, a good estimate of the Green's function is not expected. Cross-correlation of the ship data during this time yields a multi-path result that looks similar to the Green's function, however the arrival *structure* will have path dependent inaccuracies. The times tend to be a little early due to stationary phase paths not all being sampled by the small ship source volume.

The structure of ship cross-correlations will only converge to that of the true Green's function if either the ship moves along the end-fire direction or the cross-correlations are averaged over many ship tracks that intersect the end-fire direction. Ship dominated ambient noise, which has been investigated in detail for the same data elsewhere,[6] will not be discussed further here.

IV. CONCLUSION

It has been demonstrated theoretically and experimentally that the EGF between two receivers can be extracted from active sources. The EGF is determined by summing the cross-correlations between the receivers over all source positions. Here a source was lowered vertically through the water column, and also towed horizontally at shallow depth.

EGFs determined from two active source configurations, a vertically lowered source and a horizontally towed source, were compared with EGFs from a stationary ship at a single end-fire location, and EGFs from a ship/wave dominated ambient noise field. It has been shown that the EGFs from all source configurations yield direct arrival time estimates that match well with the simulated travel times. However, the stationary ship arrival times were slightly early because it is only an extended point source.

The horizontal tow results exhibited high noise levels at times prior to the direct arrival. This noise would be greatly reduced if the towed source were extended past the array. For the source lowered vertically had lower noise levels, but surface and surface-bottom paths were not as well determined. Lowering over the entire water column depth would improve the result, as would calmer sea conditions; however, due to difficulties sampling near the surface and bottom as well as with penetrating the sediment, higher order paths would still not be retrieved as well as they are for the horizontally towed source, which shows more overall potential. If instead of a single source that is moved along a line, or a line of multiple sources (vertical or horizontal) were used simultaneously, then the problem of a changing environment could be ameliorated. This could potentially be advantageous over the ambient noise methods, which although they are sound in most other aspects, suffer from needing data collection over a time window larger than the time-scale of some environmental changes, which results in them providing an estimate of a time-averaged rather than instantaneous Green's function.

Active source OAI can potentially be used to construct new propagation paths between the receivers, which can then be used for inferring the medium between these paths. To make this practical, it is important to understand the sources of error and to take these into account when designing future experiments. In particular, elimination of spurious arrivals and reducing arrival time bias are important for developing active source methods further.

ACKNOWLEDGMENTS

Work supported by the Office of Naval Research under Grant No. N00014-05-1-0264. L.A.B. is appreciative of support from a Fulbright Postgraduate Award in Science and Engineering, sponsored by Clough Engineering, as well as support from the Defence Science and Technology Organisation, Australia.

[1]P. Roux and M. Fink, "Green's function estimation using secondary sources in a shallow water environment," J. Acoust. Soc. Am. **113**, 1406–1416 (2003).
[2]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[3]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[4]K. G. Sabra, P. Roux, A. M. Thode, G. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[5]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[6]L. A. Brooks and P. Gerstoft, "Green's function approximation from cross-correlations of 20–100 Hz noise during a tropical storm," J. Acoust. Soc. Am. **125**, 723–734 (2009).
[7]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[8]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).

[9]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[10]P. Gerstoft, K. G. Sabra, P. Roux, W. A. Kuperman, and M. C. Fehler, "Green's functions extraction and surface-wave tomography from microseisms in southern California," Geophysics **71**, SI23–SI31 (2006).
[11]K. Wapenaar and J. Fokkema, "Green's function representations for seismic interferometry," Geophysics **71**, SI33–SI46 (2006).
[12]F. Lin, M. H. Ritzwoller, J. Townend, S. Bannister, and M. K. Savage, "Ambient noise Rayleigh wave tomography of New Zealand," Geophys. J. Int. **170**, 649–666 (2007).
[13]Y. Yang, M. H. Ritzwoller, A. L. Levshin, and N. M. Shapiro, "Ambient noise Rayleigh wave tomography across Europe," Geophys. J. Int. **168**, 259–274 (2007).
[14]E. Larose, A. Khan, Y. Nakamura, and M. Campillo, "Lunar subsurface investigated from correlation of seismic noise," Geophys. Res. Lett. **32**, L16201 (2005).
[15]K. G. Sabra, S. Conti, P. Roux, and W. A. Kuperman, "Passive *in vivo* elastography from skeletal muscle noise," Appl. Phys. Lett. **90**, 194101 (2007).
[16]O. A. Godin, "Retrieval of Green's functions of elastic waves from thermal fluctuations of fluid-solid systems," J. Acoust. Soc. Am. **125**, 1960–1970 (2009).
[17]G. A. Prieto and G. C. Beroza, "Earthquake ground motion prediction using the ambient seismic field," Geophys. Res. Lett. **35**, L14304 (2008).
[18]S. R. Taylor, P. Gerstoft, and M. C. Fehler, "Estimating site amplification factors from ambient noise," Geophys. Res. Lett. **36**, L09303 (2009).
[19]R. Snieder, K. Wapenaar, and K. Larner, "Spurious multiples in seismic interferometry of primaries," Geophysics **71**, SI111–SI124 (2006).
[20]D. Mikesell, K. van Wijk, A. Calvert, and M. Haney, "The virtual refraction: Useful spurious energy in seismic interferometry," Geophysics **74**, A13–A17 (2009).
[21]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[22]M. Siderius, D. R. Jackson, D. Rouseff, and M. R. Porter, "Multipath compensation in shallow water environments using a virtual receiver," J. Acoust. Soc. Am. **102**, 3439–3449 (1997).
[23]H. Schmidt, OASES Version 3.1 User Guide and Reference Manual, Department of Ocean Engineering, Massachusetts Institute of Technology (2004).
[24]J. Traer, P. Gerstoft, P. D. Bromirski, W. S. Hodgkiss, and L. A. Brooks, "Shallow-water seismoacoustic noise generated by tropical storms Ernesto and Florence," J. Acoust. Soc. Am. **124**, EL170–EL176 (2008).
[25]K. G. Sabra, P. Gerstoft, P. Roux, and W. A. Kuperman, "Extracting time-domain Green's function estimates from ambient seismic noise," Geophys. Res. Lett. **32**, L03310 (2005).
[26]D. P. Knobles, P. S. Wilson, J. A. Goff, and S. E. Cho, "Seabed acoustics of a sand ridge on the New Jersey continental shelf," J. Acoust. Soc. Am. **124**, EL151–EL156 (2008).
[27]Y.-M. Jiang, N. R. Chapman, K. Yang, and Y. Ma, "Estimating marine sediment attenuation at low frequency with a vertical line array," J. Acoust. Soc. Am. **125**, EL158–EL163 (2009).

Analyzing lateral seabed variability with Bayesian inference of seabed reflection data

Jan Dettmer[a)]
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

Charles W. Holland
Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania 16804-0030

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

(Received 11 February 2009; revised 7 May 2009; accepted 8 May 2009)

This paper considers Bayesian inversion of seabed reflection-coefficient data for multi-layer geoacoustic models at several sites, with the goal of studying lateral variability of the seabed. Rigorous uncertainty estimation is carried out to resolve lateral variability of the sediments from inherent inversion uncertainties. The uncertainty analysis includes Bayesian model selection, comprehensive quantification of data error statistics, and a Markov-chain Monte Carlo approach to transforming data uncertainties to model uncertainties. Model selection is addressed using the Bayesian information criterion to ensure parsimony of the parametrizations. Data error statistics are quantified by estimating full covariance matrices from data residuals, with posterior statistical validation. A Metropolis–Hastings sampling algorithm is used to compute posterior probability densities. Four experiment sites are considered along a track located on the Malta Plateau, Mediterranean Sea, and the inversion results are compared to cores taken at each site. Differences between profile marginal-probability distributions at adjacent sites are quantified using the Bhattacharyya coefficient. Differences that exceed the estimated geoacoustic uncertainties are interpreted as spatial variability of the seabed. The results are compared to an interpretation of geologic features evident in a chirp sub-bottom-profiler section.
 [DOI: 10.1121/1.3147489]

PACS number(s): 43.30.Pc, 43.30.Ma, 43.60.Pt [AIT] Pages: 56–69

I. INTRODUCTION

Knowledge of the uncertainty and spatial variability of geoacoustic properties of the seabed is important for many sonar applications in shallow water. Spatial variability is a measure of the inherent heterogeneity in an environmental property. In principle, variability is a property of the environment and is independent of measurement techniques. In practice, however, variability is typically estimated from measured data. Hence, variability estimates are limited by the resolving power of the data, and improved experiment or data-analysis techniques may improve the ability to estimate variability. In geoacoustic inversion, vertical variability is commonly addressed by resolving vertical seabed structure in terms of discontinuous layers and/or gradients. Quantifying vertical variability requires estimating the amount of vertical structure resolved (supported) by the data, which can be addressed using model selection techniques. While much work has been directed toward resolving vertical seabed structure, relatively little effort has been applied toward the challenging problem of quantitatively determining the lateral variability, which is the focus of this paper.

In contrast to variability, uncertainty is a measure of the state of knowledge of an environmental parameter and is ideally quantified as a probability distribution. Uncertainty is a property of existing knowledge of the environment, and hence uncertainty can be reduced by improved experiment or data-analysis techniques. The uncertainty does not quantify the variability of the environmental parameter over the region (e.g., an accurate estimate of the average could be obtained even for a highly variable property). However, meaningful uncertainty estimates are a prerequisite to resolving environmental variability: differences in parameter estimates must exceed their uncertainties to be interpreted as lateral variability.

It is important to note that uncertainties from geoacoustic inversion quantify the accuracy of parameter estimates for the particular model (i.e., parametrization) adopted to represent the environment. The error due to an inappropriate choice of model parametrization can be considered part of the theory error. Therefore, environmental variability can cause theory error due to under-parametrization, which affects parameter uncertainties. Further, under-parametrization generally leads to underestimating uncertainty and over-parametrization leads to overestimating uncertainty since the amount of data space accessible by the model increases with the number of parameters. Hence, a quantitative approach to model selection (i.e., selecting an appropriate model parametrization) must be included in the inversion methodology to obtain meaningful uncertainty estimates whenever the parametrization is not simple to address.

[a)]Author to whom correspondence should be addressed. Electronic mail: jand@uvic.ca

This paper applies Bayesian inference, including model selection and comprehensive uncertainty estimation, to single-bounce reflection data collected at several sites along a 20-km track on the Malta Plateau to resolve spatial variability of the upper seabed sediments. The experiment measured the single-bounce seabed response at a single hydrophone for source-receiver ranges from several meters to 1 km and frequencies from several hundred to several thousand hertz. The data are processed as reflection coefficients as a function of frequency and angle. These data laterally average environmental properties over an approximately 100-m seabed footprint. The parameter estimates via inversion are uncertain due to the effects of data errors, including measurement errors (e.g., instrument effects and ambient noise) and theory errors (due to the simplified propagation model and idealized parametrization). Collecting more data (e.g., at more grazing angles or frequencies) or improving data quality by reducing errors (improving signal to noise ratio, reducing theory error) will generally reduce the parameter uncertainty and improve the ability to resolve variability between measurement sites.

The differences between the recovered marginal posterior probability distributions for sediment profiles at the sites can be interpreted in terms of the environmental variability. In this case, rigorous uncertainty estimation for the individual inversions is essential to determine whether observed differences are due to actual environmental variability or simply result from uncertain parameter estimates. To this end, model selection is considered by assessing several models with increasing numbers of sediment layers using the Bayesian information criterion (BIC), with each layer defined by four parameters consisting of layer thickness, sound velocity, density, and attenuation.[1] Data-error statistics are estimated from data residuals by computing full nonstationary data covariance matrices[2] in initial inversions, which are subsequently applied in estimating the posterior probability density (PPD). The validity of the statistical assumptions is examined using posterior statistical tests for Gaussianity and randomness of standardized residuals, as well as qualitative plots of residual histograms and autocorrelation functions.[2]

Four experiment sites are considered along the track and plane-wave reflection-coefficient inversion is carried out here for the upper 4 m of the seabed sediments. Data at three sites indicate several reflectors in the upper 4 m. For these sites, model selection studies are carried out to determine the appropriate number of sediment layers based on the BIC. The use of Bayesian model selection to determine appropriate parametrization for layering structure that is below the pulse length of the acoustic source has been considered previously by the authors for a different experiment location.[1]

One site contains no discernible reflectors in the upper 4 m; hence, model selection for a layered parametrization is not carried out, rather the sediments are modeled using continuous sound-velocity and density gradients.[3] The inversion results for this site have been previously published[3] but are reproduced here in a new way to be consistent with the results in this paper. While the fine structure of this transition layer (gradients in density and velocity) is interesting and often observed in fine-grained sediments, it is not discussed in detail in this paper.

In the present study, sediment cores taken at all experiment sites provide an independent estimate of the seabed structure for comparison with the results from the Bayesian reflectivity inversion.

Inversion results for all sites are considered in the context of spatial variability along the track. Differences and similarities between adjacent sites are examined by measuring the amount of overlap between profile marginal distributions with the Bhattacharyya coefficient (BC).[4] The interpretation of the spatial variability is supported by a basic interpretation of the geologic features of a chirp bottom-profiler section that was collected along the track.[5]

II. BAYESIAN INFERENCE

This section gives a brief overview of the Bayesian inference applied in this paper; more complete treatments can be found elsewhere.[6–11] Let $\mathbf{d} \in \mathbb{R}^N$ be a random variable of N observed data containing information about a physical system. Further, let $\mathcal{I}$ denote the model specifying a particular choice of physical theory, model parameterization, and error statistics to explain that physical system. Let $\mathbf{m} \in \mathbb{R}^M$ be a random variable with $M(\mathcal{I})$ free parameters representing a realization of the model $\mathcal{I}$. Bayes' rule can then be written as

$$P(\mathbf{m}|\mathbf{d},\mathcal{I}) = \frac{P(\mathbf{d}|\mathbf{m},\mathcal{I})P(\mathbf{m}|\mathcal{I})}{P(\mathbf{d}|\mathcal{I})}, \tag{1}$$

where the conditional probability $P(\mathbf{m}|\mathbf{d},\mathcal{I})$ represents the PPD of the unknown model parameters given the observed data, prior information, and choice of model $\mathcal{I}$. The conditional probability $P(\mathbf{d}|\mathbf{m},\mathcal{I})$ describes the data error statistics. Since data errors include measurement and theory errors (which cannot generally be separated), the specific form of this distribution is often unknown. To interpret Eq. (1) quantitatively, some particular form must be assumed for this distribution. In practice, mathematically-simple distributions, such as multivariate Gaussian distributions, are commonly used; the validity of such assumptions should be checked *a posteriori* using statistical tests (described later).[2,3,12] The general multivariate Gaussian distribution for real data is given by

$$P(\mathbf{d}|\mathbf{m},\mathcal{I}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}_d|^{1/2}} \times \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{d}(\mathbf{m}))^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{d}(\mathbf{m}))\right), \tag{2}$$

where $\mathbf{C}_d$ is the data covariance matrix and $\mathbf{d}(\mathbf{m})$ are the modeled data. The covariance matrix $\mathbf{C}_d$ is often unknown since the source of errors may be poorly understood but can be estimated non-parametrically using data residuals from an initial inversion (i.e., the difference between modeled and measured data).[1,2] In inverse theory, $P(\mathbf{d}|\mathbf{m},\mathcal{I})$ is interpreted as the likelihood function $\mathcal{L}(\mathbf{m}|\mathcal{I})$ of $\mathbf{m}$ for fixed (observed) data $\mathbf{d}$. The term $P(\mathbf{m}|\mathcal{I})$ in Eq. (1) represents the model prior distribution. In this paper, prior distributions are taken

to be bounded, uniform distributions of the form

$$P(\mathbf{m}|\mathcal{I}) = \begin{cases} \prod_{i=1}^{M(\mathcal{I})} (m_i^+ - m_i^-)^{-1} & \text{if } m_i^- \leqslant m_i \leqslant m_i^+, \\ i = 1, M(\mathcal{I}), & \\ 0 & \text{else.} \end{cases} \tag{3}$$

To estimate the PPD for a fixed choice of model $\mathcal{I}$, Markov-chain Monte Carlo (MCMC) sampling methods are usually applied.[8–10,13,14] The PPD quantifies the state of information about the model parameters given the data, prior information, and parametrization, and can be used to obtain model parameter estimates [e.g., the maximum *a posteriori* (MAP) estimate] and uncertainty estimates (marginal distributions). MAP estimates can be determined using numerical optimization methods, such as a parallel implementation of adaptive simplex simulated annealing.[1,15] Uncertainty estimates require numerical integration of the PPD. The PPD is computed here by MCMC sampling that applies an adaptive Metropolis–Hastings algorithm.[13,14,16]

The conditional probability $P(\mathbf{d}|\mathcal{I})$ is commonly referred to as the *evidence* or *marginal likelihood* of $\mathcal{I}$. It describes how likely a parametrization $\mathcal{I}$ is given the observed data and prior. Since the evidence $P(\mathbf{d}|\mathcal{I})$ normalizes Eq. (1), it can be written as

$$\mathcal{Z}(\mathcal{I}) = P(\mathbf{d}|\mathcal{I}) = \int_{\mathcal{M}} P(\mathbf{d}|\mathbf{m},\mathcal{I}) P(\mathbf{m}|\mathcal{I}) d\mathbf{m}. \tag{4}$$

Bayesian evidence, Eq. (4), is the basis for model selection and brings a natural parsimony to the model selection problem.[10] Evidence quantifies the level of model complexity (e.g., the number of model parameters) that is supported by the data and discriminates against excessive model complexity. Due to the high computational demand of the forward and inverse problems considered in this paper, an asymptotic point estimate (for the maximum-likelihood model vector $\mathbf{m}^{\mathrm{ML}}$) is used to carry out model selection, consisting of the BIC,[1,17] defined as

$$\mathrm{BIC}(\mathcal{I}) = -2 \log_e \mathcal{L}(\mathbf{m}^{\mathrm{ML}}|\mathcal{I}) + M(\mathcal{I}) \log_e N. \tag{5}$$

Since the BIC is based on the negative log likelihood, the model with the smallest BIC is selected as the preferred model. The value of the BIC cannot be directly associated with a probability and cannot yield the significance of the selection.

The Bayesian formulation of the inversion is based on the fundamental assumption of random errors that are distributed according to a certain distribution (e.g., Gaussian). However, the actual data errors are generally not separable from the data. Therefore, data errors are commonly approximated by the data residuals. To obtain meaningful inversion results, the residuals should be Gaussian distributed and satisfy the assumption of randomness. These assumptions are often not satisfied by raw data residuals $\mathbf{d}-\mathbf{d}(\hat{\mathbf{m}})$ due to serial correlations in the data errors. These correlations can be (approximately) quantified by estimating the data covariance matrix from the data residuals.[2,3,12] This matrix can then be used to calculate standardized data residuals $\mathbf{C}_d^{-1/2}[\mathbf{d}-\mathbf{d}(\hat{\mathbf{m}})]$, which should represent an uncorrelated Gaussian random process to satisfy the assumptions.

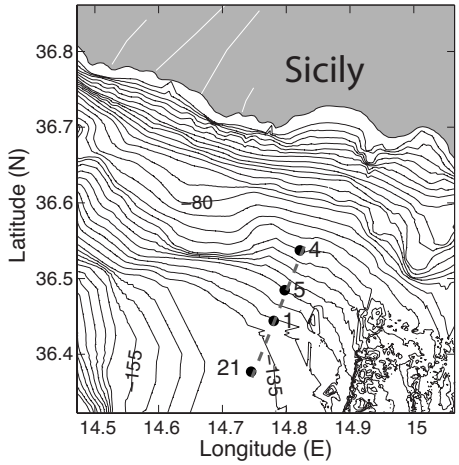


FIG. 1. Bathymetry of the Malta Plateau with the locations of sites 4, 5, 1, and 21. The dashed line indicates the chirp bottom-profiler track of ~20 km length.

There are several qualitative and quantitative approaches to examining standardized residual statistics. Quantitative tests include the runs test (one sided median-delta test) for randomness and the Kolmogorov–Smirnov (KS) test for the distribution form (e.g., Gaussian).[2,3,12] Passing these tests indicates no evidence against the assumptions of Gaussian-distributed errors with the estimated covariance. However, it is important to note that these tests indicate only whether a deviation from the assumptions can be detected statistically but do not provide a measure of the significance of the deviation. Hence, in case of failed tests it is useful to qualitatively examine the residual distribution and correlation to determine whether the deviation is significant. For example, small deviations from Gaussianity are usually benign;[18] however, large outliers can substantially distort inversion results and may require alternative error distributions to be considered (e.g., Laplace distribution). It is possible that no obvious violations of assumptions exist but standardized residuals still fail quantitative statistical tests, since the error statistics of measured data may not be completely explained by mathematically-simple distributions. In such cases, alternative error distributions may not be easily found and qualitative tests should be carried out to ensure that no serious violations exist. Useful qualitative tests include examining the auto-correlation function and histogram plots of the standardized residuals. The auto-correlation function is commonly used to visualize serial correlations: A narrow peak at zero lag indicates short correlation lengths and largely random errors. Residual histogram plots can illustrate differences from the theoretical Gaussian distribution and reliably detect outliers.

III. EXPERIMENT AND MODELING

A. Experiment

Seabed reflection data were recorded at four sites along a track on the Malta Plateau, south of Sicily (Fig. 1), as part

TABLE I. Experiment information including location, water depth, name of the experiment, date, and source type. The GeoAcoustics 5813B Geopulse boomer and EG&G 265 Uniboom sources are abbreviated as GB and EU, respectively.

Site	East (deg)	North (deg)	Depth (m)	Experiment	Date	Source
1	14.7804	36.4441	130	Advent99	May 10 1999	GB
4	14.8216	36.5375	102	SCARAB98	Apr 20 1998	EU
5	14.7979	36.4846	120	Advent99	May 10 1999	GB
21	14.7450	36.3771	137	Boundary04	May 15 2004	GB

of three experiments carried out in different years (Table I). At each site, the experiment consisted of recording signals from a ship-towed controlled source at a bottom-moored hydrophone. The source is towed along a radial track moving toward the hydrophone, resulting in seismo-acoustic recordings at ranges typically between 10 and 1000 m. The water depth increases from north to south along the track, from 102 m at site 4 to 138 m at site 21. Coordinates and water depth for all sites are given in Table I. Inversion results for site 4 have been published previously;[3] the earlier results are repeated in this paper but are illustrated here in a new way to be consistent with the presentation of the other three sites.

The water-column sound-velocity profiles measured for each site are given in Fig. 2. The largest change in sound velocity is $\sim$10 m/s over the full water column. The seismo-acoustic recordings were generated with two different electro-mechanical impulsive boomer sources (depending on the experiment, see Table I) with a short pulse length ($<$1 ms) and a broad bandwidth (0.5–10 kHz). Data were recorded at a single receiver that was part of a vertical line array. The hydrophones used in this study were situated in the lower third of the water column with depths between 60 and 126 m. The source was towed at 0.3-m depth for all sites.

The seismo-acoustic recordings were used to compute reflection-coefficient data as a function of frequency and

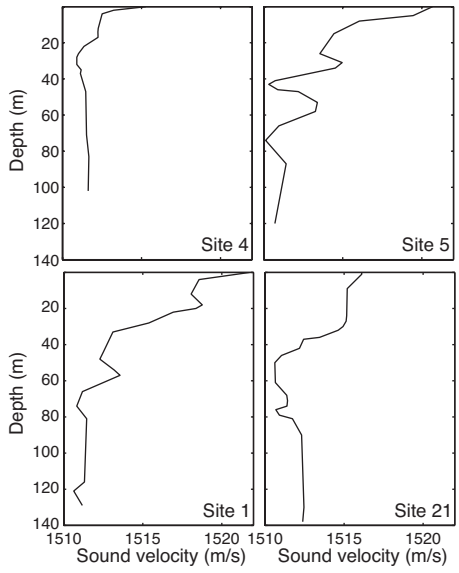


FIG. 2. Water column sound-velocity profiles used in the inversion for sites 4–21 (sites are ordered north to south from left to right).

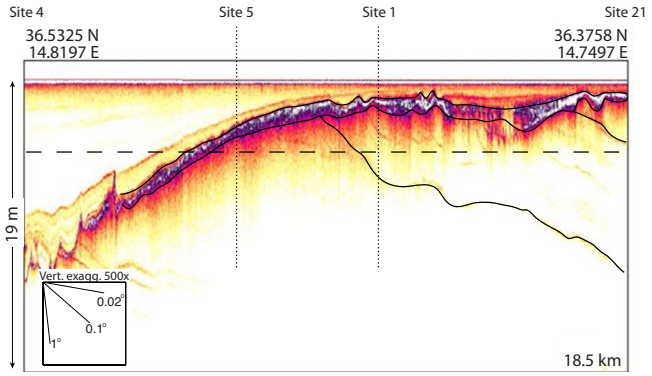


FIG. 3. (Color online) Chirp sub-bottom-profiler section with bathymetry removed (Ref. 5). Sites 4 and 21 are close to the left and right ends of the section, at water depths 102 and 137 m, respectively. Sites 5 and 1 are indicated by dotted lines. The dashed line indicates approximately 4-m depth, the limit of the reflectivity inversions. Heavy solid lines indicate a basic interpretation of geologic features.

angle using the methods of Ref. 19. All time-domain recordings were time windowed to represent returns from the upper approximately 4 m of the sediment. For each site, reflection-coefficient data were computed for several frequency bands with a fractional bandwidth of 1/10 applying a Gaussian frequency average.[20] This bandwidth was found to retain structure in the reflection-coefficient data while reducing noise and allowing a computationally feasible inversion.[1] The number of frequency bands used in the inversions and the total bandwidth considered for each data set were chosen guided by an assessment of the reflection-coefficient data quality (discussed in Sec. IV). The bandwidth for sites 4, 5, and 1 is approximately 300–2500 Hz. Site 21 appeared to have more layering structure in the upper-most part of the sediment and good quality high-frequency data, so frequency bands up to 4000 Hz were used. The reflection-coefficient data are interpolated onto a uniform spacing in angle; data points with a signal to noise ratio of less than 6 dB are excluded from the inversion. Further, interpolated data that fall into recording gaps (due to data transfer and recording during the experiment) are excluded. This results in approximately 90 data at each frequency with an angular range from 12° to 81° (55° at site 5 due to short-range clipping of the recorded arrivals).

B. Chirp bottom-profiler section

Figure 3 shows a chirp bottom-profiler section taken along a portion of the track (see Fig. 1) with the bathymetry removed (originally published in Ref. 5). Figure 3 also shows the approximate site locations as well as an interpretation of five main geologic features. A sediment wedge is apparent in the upper left part of the figure. Below the wedge is a layer with higher impedance that can be traced throughout the section. Further, a layer can be seen that pinches out between sites 5 and 1. Finally, another layer pinches out between sites 1 and 21. The inversion results in this paper will be compared to this section.

The upper silty-clay wedge in Fig. 3 was deposited in the Holocene, 6–7 kabp (kilo-annum before present) and is composed of both terrigeneous and biogenic fines. During

this time period, sea level was approximately constant, within a few meters. Given the depositional conditions (and in the absence of other processes), the properties of this layer should exhibit modest lateral variability. However, large storms can easily suspend surficial fine-grained sediments, with the fines carried in bottom currents seaward toward the shelf-break. Thus, the upper part of this layer may exhibit gradients that vary as a function of distance from the coast.

The thin, ~0.5 m, layer appears to be fairly continuous between sites 5 and 1. Between sites 1 and 21, the layer appears discontinuous for some distance. This layer below the sediment wedge (bounded by black lines) was deposited during rising sea level (from 17 to 7 kabp) in a high-energy environment. At the beginning of this time period, sea level was roughly 130 m lower than present, and thus all of the area experienced near-shore (wave-breaking) conditions at some time, and some of the area was sub-aerially exposed. The erosional channels near site 1 were cut during this time period. Sands and broken shell material are expected (and found) in this layer. Substantial lateral variability is expected in this layer.

Sediments below the thin layer were deposited during the last Ice Age when the sea level was roughly constant at ~130 m lower than present. These sediments are expected to be heterogeneous and coarse, ranging from clay to coarse sand.

C. Core measurements

Sub-bottom cores were collected at all experiment sites to provide independent estimates of the sound-velocity and density structure of the upper sediments. The method of collecting cores and measuring core properties differed between experiment sites. The highest quality cores were gravity cores; however, gravity coring has a limited ability to penetrate the sediment and typically only samples the uppermost tens of centimeters. Piston cores penetrate deeper (up to several meters), but are more likely to suffer from systematic errors since they must be stored horizontally on ship (due to their length) between taking the sample and measuring velocity and density, which can result in biased property measurements due to settling processes.

The uncertainties given for most core profiles in this paper indicate two standard deviation measurement errors of the core logger [Geotek Multi-Sensor Core Logger (MSCL)] but do not include errors due to sampling. For instance, disturbance of the sediment sample due to core penetration, sphincter closing, core recovery and capping, and core storage can lead to significant systematic errors.[3] The density estimates of the cores at sites 5 and 1 differ from the other sites, since they were measured by manual sampling (samples extracted and weighed) and not gamma ray decay (used for the MSCL measured cores). Uncertainty estimates are therefore difficult to establish for these cores. It is also important to note that the cores typically sample over a very small lateral scale with core diameters of 10 cm; in contrast, the reflection measurements applied in this paper average laterally over a spatial scale of ~100 m. Hence, core profiles may contain small-scale local structure not present in reflection inversions.

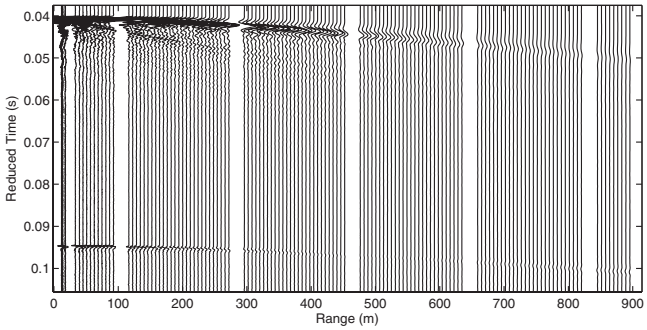


FIG. 4. Seismo-acoustic recordings (in reduced time, with 1512-m/s reducing velocity) collected at site 4 on the Malta Plateau.

D. Forward model and prior

All inversions were carried out using a plane-wave reflection-coefficient forward model that approximates the seabed as a layered lossy fluid.[2] Since this study focuses on the upper-most ~4 m of the sediment, spherical-wave effects (which increase with penetration depth) are small, and the plane-wave reflection-coefficient approximation was found to be sufficient through a forward modeling study.[21] Further, shear velocities in fine-grained sediments are low,[22] and earlier inversion studies[23] that treated the seabed as a half-space showed that the reflection coefficient is relatively insensitive to shear properties and that ignored shear properties do not significantly affect estimates for the other physical properties. The replica reflection-coefficient data for each frequency band are computed using the same frequency averaging as applied to the measured data. However, to address limited computational time, the number of frequencies in the average is limited to 12 frequencies per band. The inversions for sites 5, 1, and 21 were carried out in terms of reflection coefficient (V) data. The inversion for site 4 was carried out in terms of bottom loss ($\mathrm{BL}=-20\log_{10}|V|$) in decibels, since this site features a prominent angle of intromission. The main difference in carrying out inversions in bottom loss is that data errors are represented differently. Constant errors in decibels are equivalent to errors in reflection coefficients that scale with the data (e.g., small data have small errors). This helps emphasize the small reflectivity values near the angle of intromission in the inversion. A detailed account of the bottom-loss inversions for site 4 can be found in Ref. 3.

The prior information in this study was chosen to be largely non-informative, so that the data predominantly determine the posterior. In particular, wide uniform priors are applied for all parameters. The time-domain reflection recordings (e.g., Fig. 4) at short range were used to form a rough initial estimate of layer thickness on which the priors of these parameters are based.[1] Finally, parameter contrasts across layer boundaries are constrained so that the sound velocity and density changes both have the same sign.

IV. DATA, MODEL SELECTION, AND INVERSION RESULTS

This section considers the reflection data, model parametrization, and inversion results for the four experiment sites, ordered from north to south along the track (i.e., progressing away from Sicily into deeper water, as shown in Fig. 1). The variability along the track derived from the inversion results is considered in Sec. V.

A. Site 4

The results for site 4 were published previously,[3] but are presented here in a new form. Figure 4 shows good-quality recordings with low noise. In this case, the hydrophone was farther from the seabed than for the other sites, which reduces the effects of ringing and reflections off the ship's hull on the recordings, but increases the potential of water-column multiples interfering with the bottom-response. The direct arrival occurs at 0.04 s at the shortest range and the bottom-response starts at 0.095 s. The short range data of seismo-acoustic recordings is essentially equivalent to vertical incidence sub-bottom-profiler data. These recordings show clear angle-of-intromission effects (extremely weak reflection) between 480 and 650 m range.

Site 4 is located in an area with a thick (24 m) layer of low sound-velocity sediments. The thickness of the sediment is evident in the seismo-acoustic recordings (Fig. 4), where no reflectors appear past the water-sediment interface. Because of the lack of internal reflectors, a model selection study was not carried out for site 4. However, sound-velocity and density gradients are often observed in this type of sediment; therefore, the model parametrization was chosen to represent general sediment gradients.[3] This parametrization is discussed in detail in Ref. 3 and allows for a wide range of shapes in the density gradient. The sound-velocity for this site is parametrized as a linear gradient, which is sufficient for the type of transition-layer structure expected at those sites. Attenuation is considered constant with depth, which is consistent with the limited ability to resolve attenuation in non-layered sediments. While these gradients are an interesting feature of this type of sediment layer, they are not discussed further in this paper.

The reflection data (Fig. 5) show a prominent angle of intromission which is constrained by many data. The fit of the inversion results to the data is very good.

Figure 6 shows the inversion results in terms of marginal-probability profiles, which exhibit a fairly constant sound-velocity structure with a slight negative gradient over the upper 1.7 m. The density-profile marginals show a distinct positive gradient.

The parameter estimates from two cores taken at site 4 (Fig. 6) agree remarkably well with the inversion results. Both cores indicate a slight negative sound-velocity gradient which agrees with the inversion results to within uncertainties. The core density measurements agree closely with the inversion results. However, the core densities show a stronger gradient in the upper-most few centimeters, which is consistent with the depth-resolution limit of the frequency bandwidth considered in the inversion. The attenuation is fairly well constrained at $\sim 0.4\ \mathrm{dB}/\lambda$.

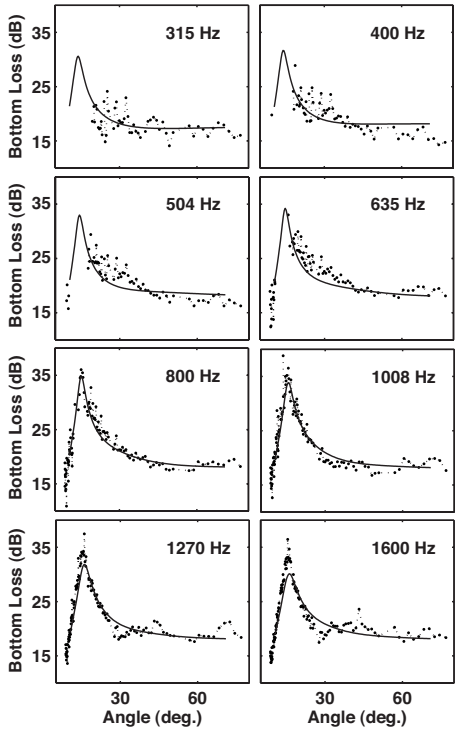


FIG. 5. Measured data (dots) and inversion fit (line) for site 4.

B. Site 5

The recordings at site 5 (Fig. 7) were clipped at short ranges (<100 m) and could not be used for this study. The missing short-range recordings translate to a loss of high-angle reflection data which limits the information content for this site.

The reflection section in Fig. 7 indicates multiple sub-bottom reflectors (described below); however, the actual layering structure is not necessarily evident from the section due to the possible presence of layers with thicknesses below the acoustic source pulse length.[1] Hence, a model selection study using the BIC [Eq. (5)] was carried out for this site (as well as sites 1 and 21, considered later). In the parametrization

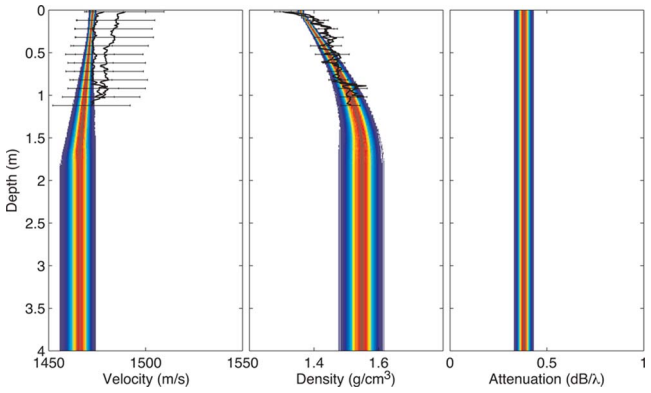


FIG. 6. (Color online) Marginal-probability depth distributions for site 4. Core measurements indicated by line with error bars on every fifth point.

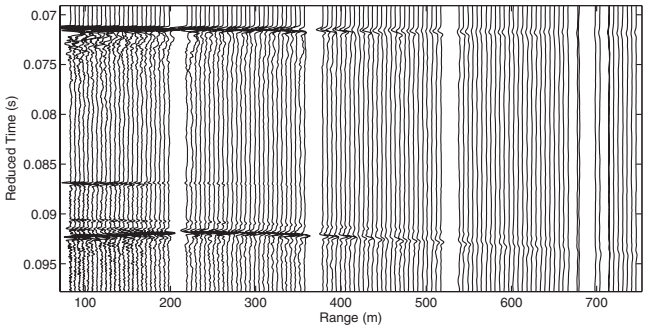


FIG. 7. Seismo-acoustic recordings (in reduced time, with 1512-m/s reducing velocity) collected at site 5 on the Malta Plateau.

adopted here, each layer is represented by four unknowns: layer thickness, sound velocity, density, and attenuation. The model selection study results are shown in Fig. 8 in terms of the BIC values and data misfit as a function of the number of layers included in the inversion. Plotting the misfit is required to check that meaningful BIC values are obtained. The misfit should decrease at least slightly with each additional layer; an increase in misfit indicates that the optimization algorithm did not converge to a global minimum. This check is particularly important for models with many layers, where optimization becomes challenging.

The time-domain data of site 5 (Fig. 7) indicate three clear reflections below the water-sediment interface at 0.087 s and two closely-spaced reflectors near 0.91 s. Hence, the sediment likely contains at least two resolvable layers plus a third layer that will be treated as the half-space in the inversion. The model selection study was carried out for four different models containing from one to four layers plus a half-space to include the most likely parametrizations for this site. The BIC and misfit results in Fig. 8 show that the two layer model is the preferred parametrization. This analysis is consistent with the reflectors visible in the seismo-acoustic recordings (Fig. 7), and does not suggest additional layering structure that is not obvious in the time-domain recordings.

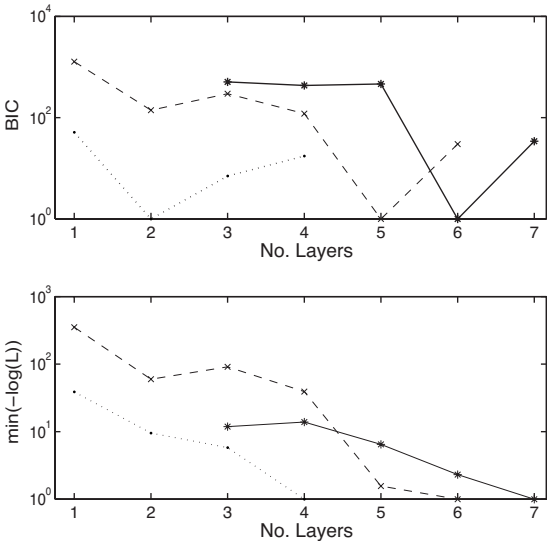


FIG. 8. BIC and misfit for sites 1 (dashed line), 5 (dotted line), and 21 (solid line). BIC and misfit values are plotted relative to the minimum for each site for display purposes.

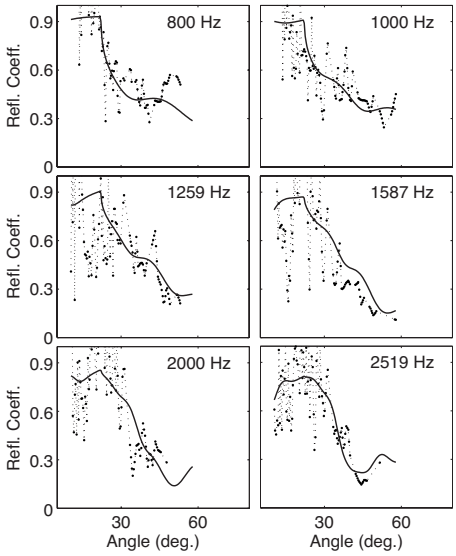


FIG. 9. Measured data (dots) and inversion fit (line) for site 5.

The reflection-coefficient data for site 5 (Fig. 9) are perhaps the most challenging in this study, due to the limited angular range and the high noise level. Almost no signs of interference patterns are obvious. However, the critical angle is clearly visible at the higher frequencies (~30° at 2519 Hz). The profile marginals in Fig. 10 indicate a low sound-velocity layer of approximately 2.3-m thickness. The second layer has significantly higher velocity (~1750 m/s), with a velocity decrease in the lower half-space. The density-profile marginal shows the same pattern although the upper-most density seems poorly constrained and peaks at the lower bound. The density in the lower half-space is not well resolved but is clearly lower than that of the second layer. Attenuation is low and well resolved in the upper-most layer, but not well resolved below that.

The core (Fig. 10) covers the upper 2.2 m of the sediment and shows low sound velocities (largely <1500 m/s), in excellent agreement with the inversion result. Some small-scale structures in the core (e.g., at 1.65-m depth) are not resolved in the inversion result. The density estimates from the core were obtained by hand sampling in this case, so that uncertainties are difficult to estimate. The inversion results

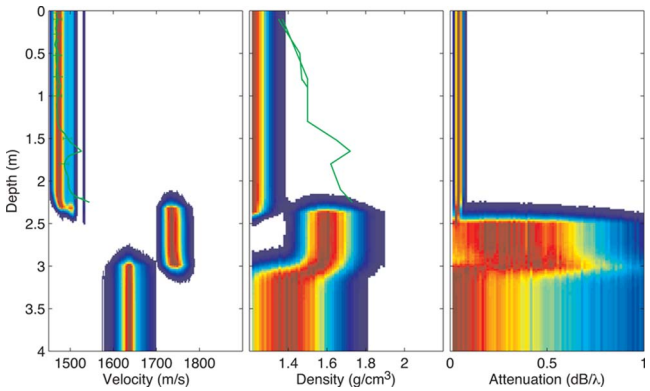


FIG. 10. (Color online) Marginal-probability depth distributions for site 5. Core measurements indicated by line with error bars on every fifth point. Density cores do not include error estimates due to hand sampling.

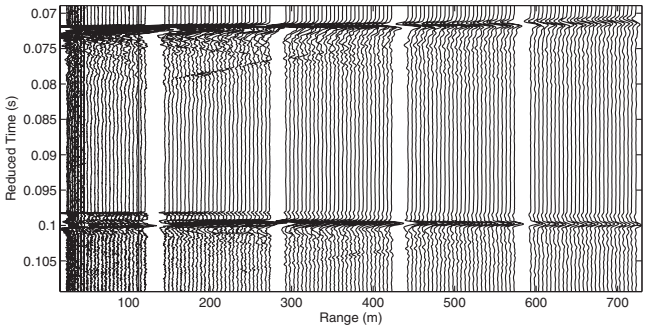


FIG. 11. Seismo-acoustic recordings (in reduced time, with 1512-m/s reducing velocity) collected at site 1 on the Malta Plateau.

agree with the core density measurements near the surface (~10-cm depth) and at the top of the second layer (~2.3-m depth). It is worth noting that the reflection data did not support the ability to resolve a gradient over the upper 2 m of sediment in the model selection study.

C. Site 1

The direct arrival and bottom-response at site 1 (Fig. 11) are at 0.071 and 0.099 s, respectively. The time-domain recordings for this site (Fig. 11) shows two reflectors close to the water-sediment interface (0.098 s) and a complicated signal at later times. The model selection study was carried out for six different models including one- to six-layer models. The model selection results in Fig. 8 support the five-layer model as the best choice. The optimization for the six-layer model was found to be particularly challenging, involving 27 unknowns, and required a very slow cooling schedule[15] but did yield a slightly lower misfit than the five-layer model. Hence, according to the model selection study, the data support additional layers that are not easily discernible in the seismo-acoustic time series.

The reflection-coefficient data for site 1 (Fig. 12) are fairly noisy but show a critical angle at some frequencies (e.g., 45° at 1600 Hz) and recognizable interference patterns. The inversion result matches the data well but misses some smaller-scale structure. The profile marginals shown in Fig. 13 indicate a thin (~25 cm) layer with low sound velocities (<1500 m/s) below the sediment-water interface. Below this layer, a higher sound velocity of ~1550 m/s extends to 1-m depth. Below this is a 50-cm high-velocity layer (~1800 m/s) that also indicates high densities (>2.0 g/cm^3); however, the densities peak at the upper prior bound. The lower half-space has a sound velocity between 1500–1550 m/s and a density between 1.4–1.8 g/cm^3. The attenuation value in the upper meter is relatively well resolved with fairly low values.

The core sound-velocity measurements (Fig. 13) only extend to 0.8-m depth and indicate sound velocities mostly lower than water velocity with a generally positive gradient. The core also indicates a strong increase in density between 0.5- and 1.5-m depth. The density then remains high for the remainder of the core (to 2.8-m depth).

The inversion sound-velocity results in the upper layer agree well with the surficial core measurements, but increase to slightly higher velocities than the core in layers 2 and 3.

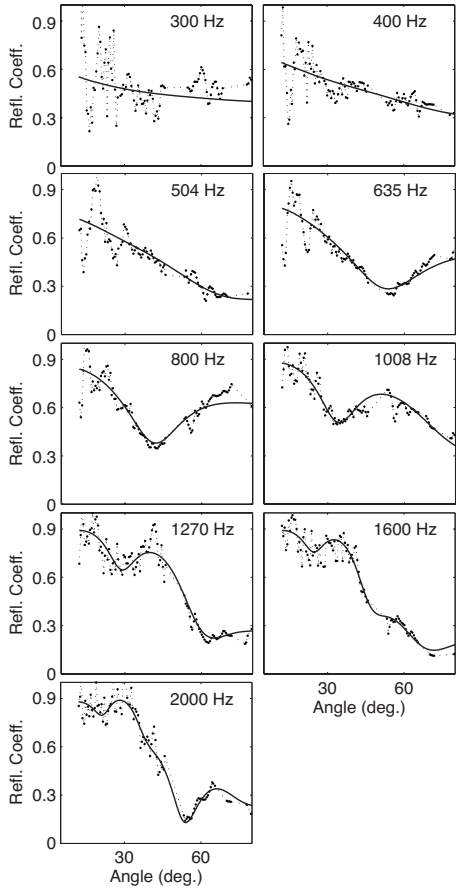


FIG. 12. Measured data (dots) and inversion fit (line) for site 1.

Inversion estimates for density agree fairly well with the core measurements. The upper two layers exhibit slightly higher density estimates than the core but show a similar structure. While the model selection study allowed for approximating a gradient with a few homogeneous layers, the inversion results do not show a clear gradient. Rather, the density is fairly high at around 1.6 g/cm^3 in the upper meter of the sediment. The third and fourth layers agree with the core from 0.8–1.5-m depth. However, the basement density estimate from the inversion is significantly lower than the core estimate. The reason for the discrepancy is not clear, in particular, since no core sound-velocity estimate exists for the

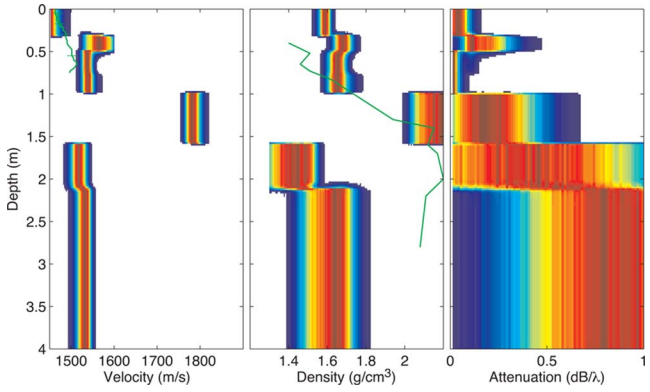


FIG. 13. (Color online) Marginal-probability depth distributions for site 1. Core measurements indicated by line with error bars on every fifth point. Density cores do not include error estimates due to hand sampling.

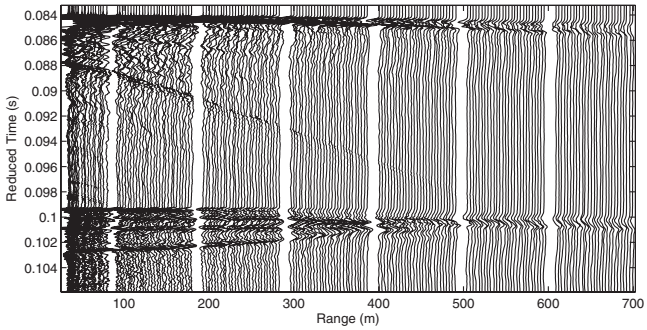


FIG. 14. Seismo-acoustic recordings (in reduced time, with 1512-m/s reducing velocity) collected at site 21 on the Malta Plateau.

lower half-space. The density estimates for the core were hand sampled, so core error estimates are difficult to assess.

D. Site 21

Figure 14 shows the seismo-acoustic recordings for site 21 (direct arrival at 0.084 s, bottom-response at 0.099 s), which include a strong reflection off the ship's hull starting at 0.088 s and running into the bottom reflection at about 500-m range. The seismo-acoustic data for site 21 (Fig. 14) show a particularly complicated bottom-response (>0.099 s). This indicates that several layers close to the water-sediment interface may be required to sufficiently model the data. In fact, the model selection study (Fig. 8) favors a six-layer model over all other models considered, providing the appropriate number of layers in a case where this is not at all obvious from the seismo-acoustic recordings.

The reflection-coefficient data for site 21 (Fig. 15) show the most complicated angular dependence of all sites considered in this study, indicating a complicated layering structure of the seabed. The data are fairly noisy at low angles and frequency bands, but improve for the higher bands. Therefore, a higher bandwidth from 500 to 4000 Hz was chosen for this site. Figure 15 shows that the interference structure of the data is matched well by the inversion results across all frequencies. The profile marginals in Fig. 16 show a complicated layering structure. In particular, the first 0.5 m of the sediment show three layers starting with a thin layer of low velocity and density. Below that, sound velocity increases to about 1-m depth, with a low-velocity layer following. Below these five relatively thin layers is a thicker ~1-m layer with the highest sound velocity. The lower half-space shows sound velocities between 1500 and 1600 m/s.

The densities of the site appear to be generally high. After the thin first layer, density increases rapidly over the upper-most meter. Beyond 1 m, the density remains high until it decreases in the lower half-space. Attenuation is generally low and most constrained in the thickest layer. There is virtually no sensitivity to attenuation in the lower half-space.

The core estimates in Fig. 16 show a strong density gradient over the first 25 cm of the core followed by a fairly constant interval. The velocity structure appears more complicated with a thin layer of fairly constant sound velocity followed by a gradient and finally a fairly constant interval. The overall agreement between the inversion results and the core estimates is excellent, and the inversion results provide a good approximation to the gradients indicated by the core.

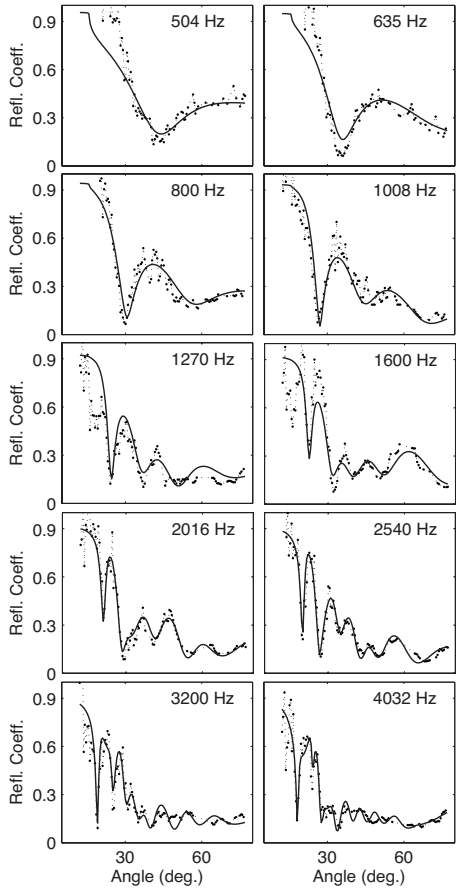


FIG. 15. Measured data (dots) and inversion fit (line) for site 21.

E. Posterior residual statistics

To ensure meaningful data error estimates for the inversions, posterior statistical tests were carried out for raw residuals $\mathbf{d}-\mathbf{d}(\hat{\mathbf{m}})$ and for standardized residuals $\mathbf{C}_d^{-1/2}[\mathbf{d}-\mathbf{d}(\hat{\mathbf{m}})]$ to examine the assumptions of Gaussian random errors. The results for the tests are summarized in Table II and indicate that the estimated covariance effects accounted for in the standardized residuals lead to a profound improvement

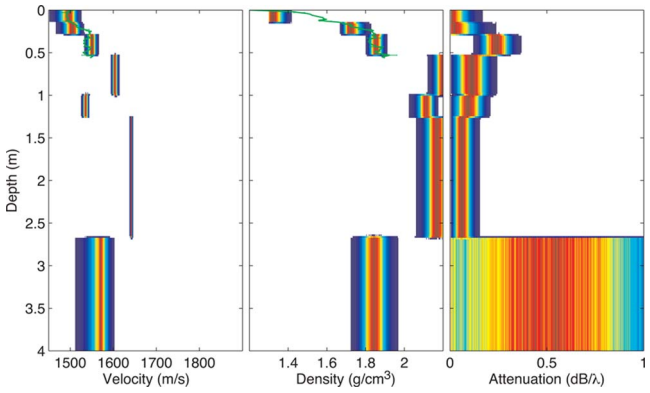


FIG. 16. (Color online) Marginal-probability depth distributions for site 21. Core measurements indicated by line with error bars on every fifth point.

TABLE II. Results of KS and runs tests for all experiment sites. The values give the numbers of frequency bands where residuals passed the tests (at a 0.05 level) out of the total number of frequency bands included in the inversion for both raw and standardized residuals.

	KS test		Runs test	
Site	Raw	Standardized	Raw	Standardized
4	5/8	4/8	1/8	5/8
5	1/6	6/6	0/6	4/6
1	4/9	9/9	0/9	6/9
21	2/10	7/10	0/10	8/10

over the raw residuals. However, some test results for the standardized residuals still appear to be potentially unsatisfactory.

One of the most problematic test results involve the assumption for randomness at site 1, where the standardized residuals fail the runs test at six out of nine frequency bands. To consider this case in more detail, Fig. 17 shows auto-correlation functions for raw and standardized residuals at all frequencies for this site. The wide center peaks for the raw-residual auto-correlation functions indicate strongly correlated data errors and long correlation lengths. The auto-correlation functions for the standardized residuals show narrow central peaks (one point wide) with a decrease in correlation by an order of magnitude to the neighboring points. This means that even though the runs test detects the presence of correlation in the standardized residuals, the level of correlation is very low and is likely not a concern for practical purposes. Table II also indicates unsatisfactory KS test results for site 21. Figure 18 compares histograms of the raw and standardized residuals for each frequency to a standard Gaussian distribution. The histograms indicate roughly Gaussian distribution shapes and no significant outliers; rather, several of the histograms appear somewhat more peaked than the Gaussian distribution. Therefore, a Gaussian distribution seems to be a reasonable approximation for the

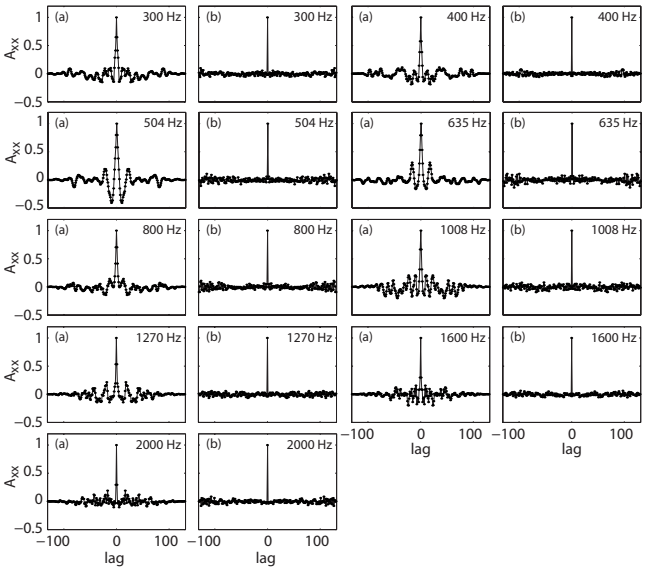


FIG. 17. Auto-correlation function for (a) raw and (b) standardized residuals at site 1.

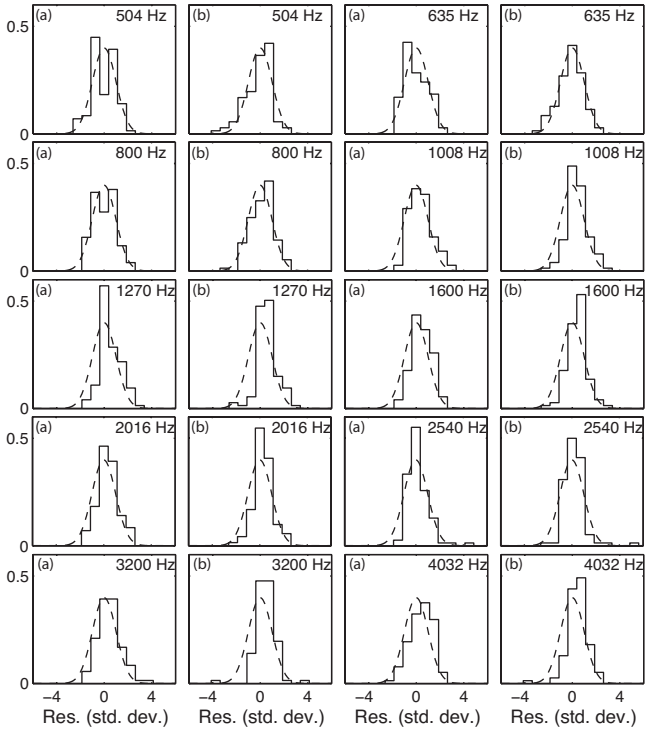


FIG. 18. Residual histograms for (a) raw and (b) standardized residuals at site 21 compared to a Gaussian distribution (dashed line).

residuals. Auto-correlation and histogram plots for other sites that failed the runs and/or KS test (Table II) are similar to Figs. 17 and 18 and are not shown here. Overall, these results suggest that data errors are characterized reasonably well, providing confidence in the inversions.

V. SPATIAL VARIABILITY ALONG THE TRACK

This section discusses and compares the reflection inversion results to study sediment spatial variability along the experiment track. Figure 19 shows the sound velocity and density posterior mean models plotted along the track from north to south. To give an impression of the uncertainty associated with the mean profiles, the width of the profiles corresponds to the 90% highest probability density credibility interval. Figure 19(c) shows part of a chirp sub-bottom profile (see Fig. 3) taken between sites 4 and 21 scaled to approximately fit the geometry of the sites. The basic geologic interpretation identifying five main features in the full chirp section (Fig. 3, discussed in Sec. III B) is indicated by heavy solid lines in Figs. 19(a)–19(c). The layer thickness results of the inversion profiles can be compared to the geologic interpretation where the dashed lines indicating site position intersect the heavy lines of the interpretation.

To quantitatively compare the inversion results along the track, profile marginals for adjacent sites are compared statistically by computing the BC (Ref. 4) as a function of depth. The BC is defined as

$$\mathrm{BC}(p,q) = \int \sqrt{p(x)q(x)}dx, \tag{6}$$

where p and q are probability distributions over the same variable x. The BC is one method to measure the degree of

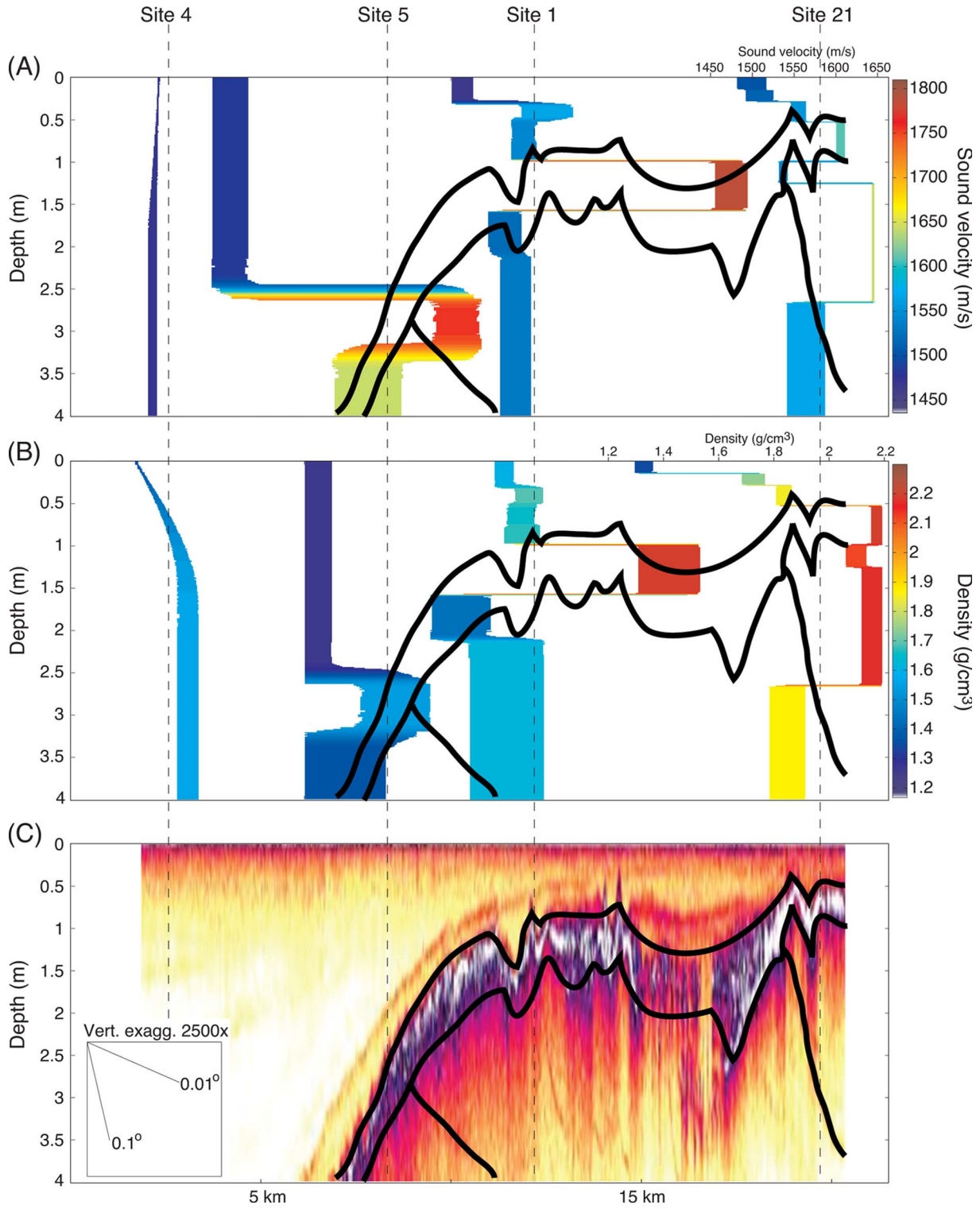


FIG. 19. (Color online) Mean sound velocity (a) and density (b) profiles compared for all sites. Widths of shaded areas correspond to 90% highest-probability density credibility intervals. Panel (c) shows a crop of the chirp bottom-profiler section from Fig. 3 with the solid lines indicating a basic geologic interpretation.

overlap of two distributions by integrating over the geometric mean of the distributions. The coefficient is 0 if no overlap exists and 1 if the distributions are identical. Figure 20 shows several examples of distributions with different amounts of overlap and the corresponding BCs for velocity profile marginal distributions for pairs of adjacent sites along the track. For each pair of sites, four discrete depths are shown (as indicated in Fig. 21) and the corresponding BC values are also given. Although there is no clear cut-off, values of BC <0.3 indicate relatively little overlap, suggesting the underlying quantities likely differ. Values of BC >0.4 indicate substantial overlap, suggesting similar quantities. How similar distributions are interpreted depends on the state of information of the underlying geoacoustic properties.

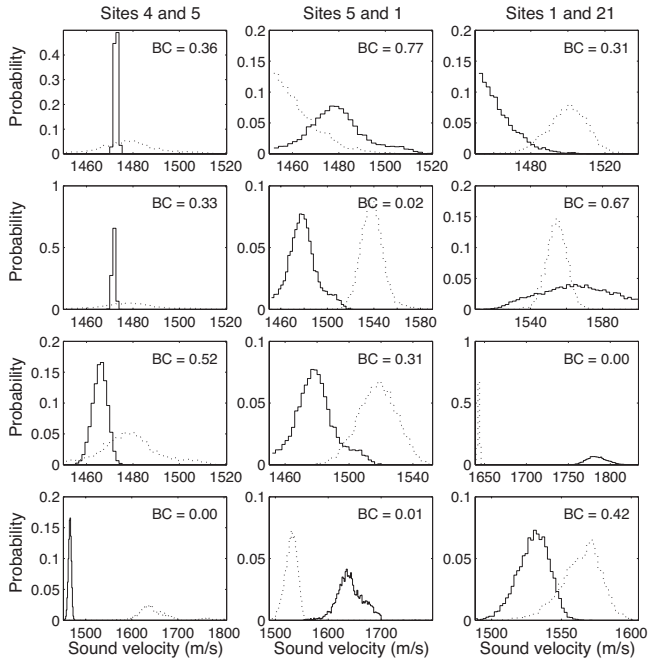


FIG. 20. Marginal-probability distributions for selected depths (see Fig. 21) for each pair of adjacent sites. Solid lines indicate the northern site of the pair, and dotted lines indicate the southern site. In some panels, plot bounds are chosen smaller than prior bounds for graphical purposes.

Similar but very wide uncertainty distributions (e.g., approaching the width of the prior bounds) provides little information regarding whether the underlying properties are similar or different. Similar narrow distributions provide reasonable confidence that underlying properties are similar.

Figure 21 shows the BC for sound-velocity profile marginals for pairs of adjacent sites as a function of depth. Comparing the two most northerly sites, sites 4 (Fig. 6) and 5 (Fig. 10), the low-velocity sediment of site 4 can be found at site 5 (and all other sites), as evident in Fig. 19. With increasing water depth and distance from the coast of Sicily (southwards), the low-velocity layer decreases in thickness. While the layer thickness is greater than 4 m at site 4, it decreases to 2.25 m at site 5 (Fig. 10), 0.5 m at site 1 (Fig. 13), and

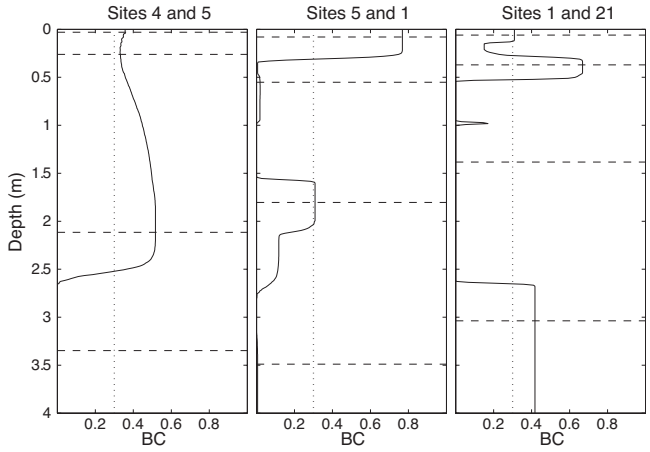


FIG. 21. BC of sound velocity as a function of depth for each pair of adjacent sites (solid line). Dotted lines indicate BC=0.3. Particular depths for which marginal distributions are compared in Fig. 20 are indicated as dashed lines.

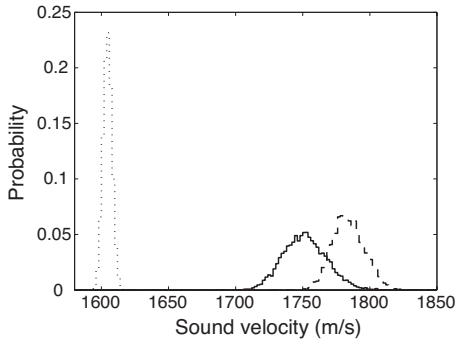


FIG. 22. Velocity profile marginal distributions at single depths for the high-velocity layer at sites 5 (solid), 1 (dashed), and 21 (dotted).

only a few centimeters at site 21 (Fig. 16). Quantitatively, Fig. 21 shows a high degree of overlap between the upper low-velocity sediment at all sites, quantifying the presence of the low-velocity sediment wedge. Figure 21 also shows a high degree of overlap between the upper low-velocity sediment at all sites, with BC values of 0.3–0.8. This, together with the relatively narrow marginal distributions shown in Fig. 20, suggests similar sediments corresponding to the low-velocity wedge that are present at all sites. Sites 1 and 21 show more complicated layering in the upper-most part of the sediment than the other sites (Fig. 19).

Below the low-velocity sediment layer, sites 5 and 1 both show a distinct high-velocity layer. From the geologic interpretation of the chirp section (Fig. 19), this high-velocity layer should also be present at site 21. Site 21 shows fairly complicated layering and the inversion results (Fig. 16) indicate strong velocity and density gradients in the upper-most sediment. To quantitatively examine the continuity of the high-velocity layer, the layers connected by the geologic interpretation can be compared using BC values for sound velocities. The BC value for sites 5 (at 3-m depth) and 1 (at 1.3-m depth) is 0.49, indicating substantial overlap and, hence, strong evidence for the same sediment velocity. However, sites 1 and 21 show a value of BC=0.00, indicating no overlap. Figure 22 illustrates these results with sound-velocity marginals taken at single depth in the high-velocity layer at each site. Marginals for sites 5 and 1 overlap substantially, but significantly lower velocities are indicated for site 21. Since the chirp section shows complicated structure between sites 1 and 21, it is not surprising that the sound velocity changes in this layer between these sites.

Results for site 21 also show a high-velocity layer that terminates at 2.7-m depth, which is not present at site 1. The geologic interpretation of the chirp profile indicates a layer pinching out between these sites that appears at a similar depth ($\sim$3 m) in the vicinity of site 21.

The results for the lower half-space at all sites are not as well determined as the upper layers, since the half-space is not constrained by reflection information from a lower interface. Nevertheless, the sound velocities for sites 1 and 21 appear to be similar at around 1550 m/s (Fig. 19). This observation is also supported by a BC of 0.42 (Fig. 21). Densities are also similar around 1.65–1.85 g/cm^3 (Fig. 19). The half-space at site 5 shows a somewhat higher velocity of 1650 m/s. The BC between sites 5 and 1 is 0.01, indicating

dissimilar distributions (Fig. 21), and this is fully supported in the chirp profile [deepest black line in Figs. 19(c) and 3] which indicates that the lowermost sediments at site 5 are distinct from those at sites 1 and 21.

VI. SUMMARY AND DISCUSSION

This paper developed a general approach to examine lateral variability of seabed sediments from a series of one-dimensional inversion results. To meaningfully analyze variability between measurement sites requires addressing several important issues: (i) determining a geoacoustic parametrization for each site that represent only structure supported by the data (carried out here using Bayesian model selection), (ii) rigorous geoacoustic uncertainty estimation for each site so that lateral variability can be differentiated from inherent inversion uncertainties (nonlinear Bayesian inference), and (iii) a quantitative measure of parameter differences that accounts for uncertainties (e.g., the BC).

Reflection inversions were carried out for four sites along a track on the Malta Plateau using Bayesian inference. Model selection based on the BIC was applied to determine appropriate parametrizations in terms of the number of sediment layers comprising the seabed model. Determining the appropriate number of layers is important to reasonably represent the uncertainty of the inversion results, since under-parametrization can lead to underestimating uncertainty and over-parametrization can lead to overestimating uncertainty. Residual analysis was applied to estimate full data covariance matrices at multiple frequency bands, and posterior statistical and qualitative tests were carried out to examine the validity of the statistical assumptions and establish confidence in the inversion results. The sediment sound-velocity and density profiles computed via reflection inversion agreed well with core measurements at each site.

The inversion results for the four sites were used to infer information about the spatial variability of the seabed along the track. Pairs of adjacent sites were compared qualitatively and examined for common features. Four main features were identified in the inversion results. First, a low-velocity sediment layer is present at all sites and decreases in thickness from north to south. Below this sediment wedge, a high-velocity layer appears at site 5 (approximately 13 km along the track) and is also present at site 1. Below the high-velocity layer, a change in basement sound velocity between sites 5 and 1 indicates a potential additional layer that pinches out between these two sites. Finally, a prominent high-velocity layer present only at site 21 indicates an additional layer pinching out between sites 1 and 21.

The continuity of these main features were examined quantitatively by measuring the overlap of sound-velocity profile marginal-probability distributions with the BC. The BC clearly quantifies the presence of the low-velocity wedge as well as the change in the sediment half-space between sites 5 and 1. The BC also indicated significant overlap of the half-space sediment properties between sites 1 and 21. Finally, the BC also indicated substantial overlap between the high-velocity layers of sites 5 and 1.

The inversion results were compared to the geologic interpretation of a chirp sub-bottom profile which identified the low-velocity sediment wedge, the high-velocity layer, a layer that pinches out between sites 1 and 21, and a change in lower half-space velocity between sites 5 and 1. Evidence for all geologic features was present in the inversion results. However, the high-velocity layer identified in the inversion results at sites 5 and 1 was not obvious in the inversion results at site 21. Nonetheless, the chirp section clearly connected this layer with a layer at site 21 and therefore indicates a significant change of the geoacoustic properties of this layer along the track.

Several previously published results[24–27] for this area produced velocities at the half-space depth (2–4 m) which are in good agreement with the values presented in this study. In particular, Siderius *et al.*,[25], Fallat *et al.*,[26], Holland,[24] and Dosso[27] found 1552, 1566, 1600, and 1526 m/s, respectively, from inversion of matched-field, reflection, and reverberation data.

Attenuation estimates for the shallow low-velocity wedge agree well at three (5, 1, 21) of the four sites, as shown in Figs. 10, 13, and 16. Only the attenuation at site 4 differs significantly (Fig. 6). In general, attenuation values are not as easy to interpret as velocity and density, since the attenuation values inferred from inversion are likely effective values that account not only for the intrinsic attenuation of the sediment, but likely also for other effects such as roughness or scattering.

The results presented in this paper indicate that inversion of reflection-coefficient data can provide high-resolution geoacoustic profiles with uncertainties suitable for interpretation of lateral variability between measurement sites.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Office of Naval Research postdoctoral fellowship (N000140710540) and the Ocean Acoustics Program (ONR OA Code 321). Some of the data were collected under the Boundary Characterization Joint Research Project including the NATO Undersea Research Centre (NURC, La Spezia, Italy), Pennsylvania State University—ARL-PSU (State College, PA), Defence Research and Development Canada—DRDC-A (Halifax, CAN), and the Naval Research Laboratory—NRL (Washington, DC). The chirp sonar data were processed by Altan Turgut and the authors wish to acknowledge helpful geologic insight from Allen Lowrie.

[1]J. Dettmer, S. E. Dosso, and C. W. Holland, "Model selection and Bayesian inference for high resolution seabed reflection inversion," J. Acoust. Soc. Am. **125**, 706–716 (2009).
[2]J. Dettmer, S. E. Dosso, and C. W. Holland, "Joint time/frequency-domain inversion of reflection data for seabed geoacoustic profiles," J. Acoust. Soc. Am. **123**, 1306–1317 (2008).
[3]C. W. Holland, J. Dettmer, and S. E. Dosso, "Remote sensing of sediment density and velocity gradients in the transition layer," J. Acoust. Soc. Am. **118**, 163–177 (2005).
[4]A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by probability distributions," Bull. Calcutta Math. Soc. **35**, 99–109 (1943).
[5]C. W. Holland, R. Gauss, P. Hines, P. Nielsen, D. Ellis, J. Preston, K. D. LePage, C. H. Harrison, J. Osler, R. Nero, D. Hutt, and A. Turgut,

"Boundary characterization experiment series overview," IEEE J. Ocean. Eng. **30**, 784–806 (2005).
[6]A. F. M. Smith, "Bayesian computational methods," Philos. Trans. R. Soc. London, Ser. A **337**, 369–386 (1991).
[7]A. F. M. Smith and G. O. Roberts, "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods," J. R. Stat. Soc. **55**, 3–23 (1993).
[8]*Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics, edited by W. R. Gills, S. Richardson, and D. J. Spiegelhalter (Chapman and Hall, London/CRC, Boca Raton, FL, 1996).
[9]M. Sambridge and K. Mosegaard, "Monte Carlo methods in geophysical inverse problems," Rev. Geophys. **40**, 3-1–3-29 (2002).
[10]D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003).
[11]A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Siam, Philadelphia, 2005).
[12]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).
[13]N. Metropolis, A. Rosenbluth, M. Rosenbluth, and A. T. A. E. Teller, "Equations of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[14]W. K. Hastings, "Monte Carlo sampling methods using markov chains and their applications," Biometrika **57**, 97–109 (1970).
[15]S. E. Dosso, M. J. Wilmut, and A.-L. S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[16]S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simulataneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2008).
[17]G. Schwartz, "Estimating the dimension of a model," Ann. Stat. **6**, 461–464 (1978).
[18]W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in Fortran 77*, 2nd edition (Cambridge University Press, Cambridge, 1997).
[19]C. W. Holland, "Seabed reflection measurement uncertainty," J. Acoust. Soc. Am. **114**, 1861–1873 (2003).
[20]C. H. Harrison and J. A. Harrison, "A simple relationship between frequency and range averages for broadband sonar," J. Acoust. Soc. Am. **97**, 1314–1317 (1995).
[21]J. Dettmer, S. E. Dosso, and C. W. Holland, "Full wave-field reflection coefficient inversion," J. Acoust. Soc. Am. **122**, 3327–3337 (2007).
[22]M. D. Richardson, *Shallow Water Acoustics* (China Ocean, Beijing, 1997).
[23]S. E. Dosso and C. W. Holland, "Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data," IEEE J. Ocean. Eng. **31**, 657–671 (2006).
[24]C. W. Holland, "Geoacoustic inversion for fine-grained sediments," J. Acoust. Soc. Am. **111**, 1560–1564 (2002).
[25]M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).
[26]M. R. Fallat, S. E. Dosso, and P. L. Nielsen, "An investigation of algorithm-induced variability in geoacoustic inversion," IEEE J. Ocean. Eng. **29**, 78–87 (2004).
[27]S. E. Dosso, P. L. Nielsen, and C. H. Harrison, "Bayesian inversion of reverberation and propagation data for geoacoustic and scattering parameters," J. Acoust. Soc. Am. **125**, 2867–2880 (2009).

Temporal and spatial coherence of sound at 250 Hz and 1659 km in the Pacific Ocean: Demonstrating internal waves and deterministic effects explain observations

John L. Spiesberger
Department of Earth and Environmental Science, University of Pennsylvania, 240 South 33rd Street, Philadelphia, Pennsylvania 19104-6316

(Received 16 September 2008; revised 21 April 2009; accepted 21 April 2009)

The hypothesis tested is that internal gravity waves explain temporal and spatial coherences of sound at 1659 km in the Pacific Ocean for a signal at 250 Hz and a pulse resolution of 0.02 s. From data collected with a towed array, the measured probability that coherence time is 1.8 min or longer is 0.8. Using a parabolic approximation for the acoustic wave equation with sound speeds fluctuating from internal waves, a Monte-Carlo model yields coherence time of 1.8 min or more with probability of 0.9. For spatial coherence, two subsections of the array are compared that are separated by 142 and 370 m in directions perpendicular and parallel to the geodesic, respectively. Measured coherence is 0.54. This is statistically consistent with the modeled 95% confidence interval of [0.52, 0.76]. The difference of 370 m parallel to the section causes spatial coherence to degrade deterministically by a larger amount than the effect of internal waves acting on the 142 m separation perpendicular to the section. The models are run without any tuning with data. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3133243]

PACS number(s): 43.30.Re [RCG] Pages: 70–79

I. INTRODUCTION

The temporal and spatial coherences of sound are estimated from a towed receiving-array with signals originating from an acoustic source over the Hoke seamount at 1659 km distance [Fig. 1(e)]. The signal is centered at 250 Hz with a bandwidth of 50 Hz (1/50=0.02 s pulse resolution). We test the hypothesis that coherence scales are accurately modeled by fluctuations of internal gravity waves obeying the Garrett–Munk spectrum.[1,2] The test uses standard methods in acoustics and oceanography with no effort to tune the models with data. A Monte-Carlo technique yields time-varying impulse responses at the receiver by evolving the internal wave field using a linear dispersion relation. If the temporal scales from the data and model agree, it would be the fourth time agreement is found using the models in this paper (Table I, Fig. 1). Another comparison with data is inconclusive because acceleration of the instruments is not accounted for even though it appears that acceleration significantly affects coherence.[6] If the spatial scales of coherence from the towed array are consistent with measurements, it would also be the fourth time that the standard spectrum for internal waves could account for such phenomena.[7–10] One of these analyses uses the data to set a model parameter to fit the data.[9] It appears that model does not compare well without tuning with data.

There are two reasons for again testing the ability of models to predict coherence. First, coherence is important for acoustic communication, signal processing, acoustical oceanography, and theoretical studies. Second, the author does not believe enough comparisons with data have been made for enough frequencies, distances, and oceans to state with certainty when modeled coherence yields accurate predictions. It appears justifiable to form a strong scientific background comprised of hundreds of papers published by numerous investigators. In this paper, comparison of temporal coherence is made at higher frequency and shorter distance than before. Previous frequencies were near 75 and 133 Hz,[3–5] and involved distances between 3000 and 4000 km.

II. EXPERIMENT AND DATA PROCESSING

A Hydroacoustics HLF-5 source was deployed over the Hoke seamount in the North Pacific in 1999 by Chiu and co-workers.[11,12] It was located at 32.105 33° N 233.088 83° E at a depth of 673 m. It is tautly moored 104 m above its anchored position on the seamount. We concern ourselves with a single transmission at 00:00 Greenwich Mean Time on 14 September, 1999. The transmission consists of 11 periodic linear shift register *M*-sequences lasting 135.036 s. Each period lasts 12.276 s and consists of 1023 digits. Each digit consists of three cycles of carrier at 250 Hz, and is encoded by modulating the phase of the carrier by $\pm 88.209\ 22°$. The sequence law is 2033_8. The source level is 192 dB re 1 μPa at 1 m (132 W). The pulse resolution is about 0.02 s. Bathymetry of the Hoke Seamount was measured using a Knudsen echo-sounder from the R/V POINT SUR in May 1999.[11]

The signal was received on a towed array at 171 m depth near 46.9023° N and 230.3542° E. The location is written with greater precision than its accuracy of 1 km so that others can reproduce the model results of this paper. The ship was heading at 12° true at a speed of about 1.7 m/s. Since the bearing angle from the source is about −9° true, the signal arrives near endfire [Fig. 1(e)]. Data were separately processed from two parts of the array to investigate spatial coherence. Going perpendicular to the geodesic between the source and receiver, the ends were separated by 142 m. In a

 0001-4966/2009/126(1)/70/10/$25.00

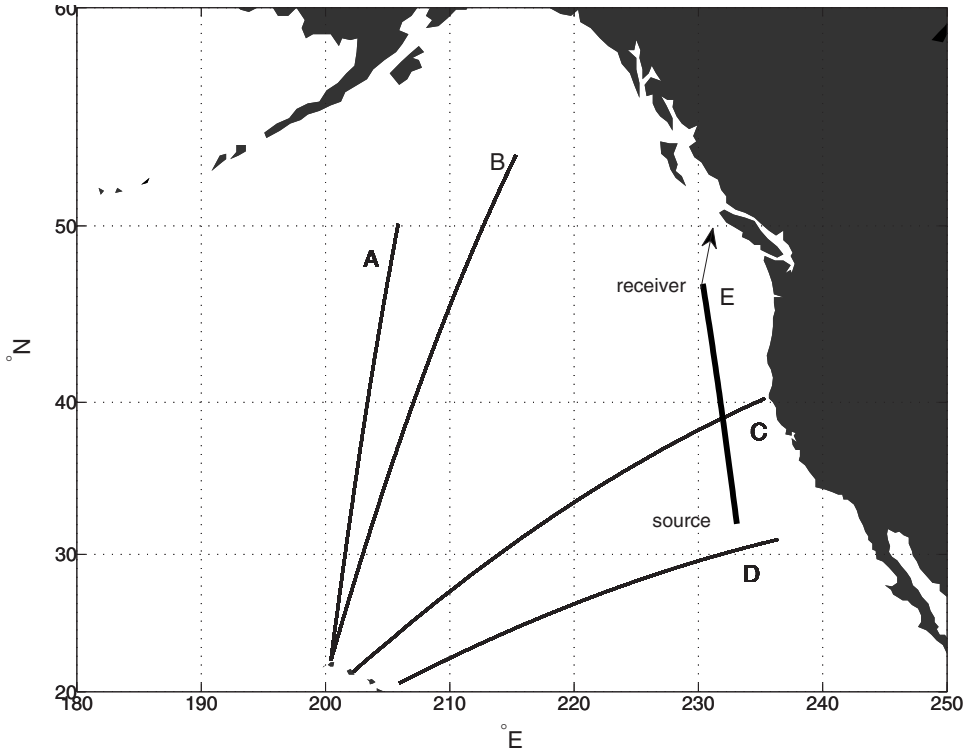


FIG. 1. Five sections where blind predictions of coherence time have been made. (A) 3115 km between a bottom-mounted source on Kauai (75 Hz, 0.03 s resolution) and a towed receiver, (B) 3683 km between the same source and a towed receiver, (C) 3709 km between a bottom-mounted source at Kaneohe Bay, Oahu (133 Hz, 0.06 s resolution) to SOSUS station mounted on the bottom, (D) 3250 km transmission between source dangled from R/V Flip (75 Hz, 0.03 s resolution) and a vertical array, and (E) 1659.32 km transmission reported in this paper between a source moored over the Hoke seamount (250 Hz, 0.02 s resolution) and a towed array. Heading of the towed array is 12° true (arrow).

direction parallel to the geodesic, the ends are separated by 370 m. These are called the "cross" and "along" geodesic separations, respectively.

A beam is steered toward the source from each end using a standard non-adaptive time-domain beamformer. The beam is much wider than any variation in signal direction. Data are Doppler corrected for each separate M-sequence period to yield the largest output of a matched filter. The signal-to-noise ratio (SNR) of the highest peak in each period was about 29 dB. In this paper, the level of noise used in computing the SNR is computed from a portion of the impulse response where signal does not occur. Using the best Doppler correction for each period, a coherent average was applied to 9 of the 11 periods, yielding a peak SNR of 38 dB [Fig. 2(a)]. 2 of the 11 periods were unprocessed to avoid end-effect sidelobes that occur when match filtering with an M-sequence.[1] Coherence times up to nine periods, or $12.276 \times 9 = 110.484$ s, can be investigated here. The ship traveled $1.7\ \mathrm{m/s} \times 110.484 = 188$ m during this interval.

TABLE I. Summary of five experiments where a Monte-Carlo technique is used to see if modeled and measured coherence times of sound are consistent. Section letter refers to Fig. 1. Analysis of sections A, B, and C are from Refs. 3–5, respectively. Results from section D are inconclusive because observations of coherence time may not quite be complete (Ref. 6). This paper concerns section E.

Section	Distance (km)	Frequency (Hz)	Pulse resolution (s)	Data-model agree?
A	3115	75	0.03	Yes
B	3683	75	0.03	Yes
C	3709	133	0.06	Yes
D	3250	75	0.03	Inconclusive
E	1659	250	0.02	Yes

Coherent averages are constructed by weighting records according to their noise variance. The result for complex demodulate d_i is

$$\overline{d}_i(J) = \frac{\sum_{j=1}^{J} d_i/\sigma_j^2}{\sum_{j=1}^{J} 1/\sigma_j^2}, \tag{1}$$

where J is the number of coherent averages, and the variance of the noise in record j is σ_j^2. The variance of the noise is estimated from a portion of the impulse response without signal.

Because location of the array is only known within 1 km, it is not possible to compare with models the absolute time of signal propagation between the source and receiver.

III. MODELS

Models for the oceanic environment and the propagation of sound are described next.

A. Environment

As there were no *in-situ* environmental measurements, climatological archives of ocean properties were used in the modeling. They are almost always sufficiently accurate to derive an acoustic impulse response that looks like day-long averages of the measured response.[13] The speed of sound along the section is computed with Del Grosso's algorithm[14] and Levitus' climatological averages[15] of temperature and salinity for summer. The depth of minimum speed varies from 560 m at the source to 430 m at the receiver. Since the acoustic models use Cartesian coordinates, the sound-speed profiles are translated to Cartesian coordinates using the Earth-flattening transformation.[16]

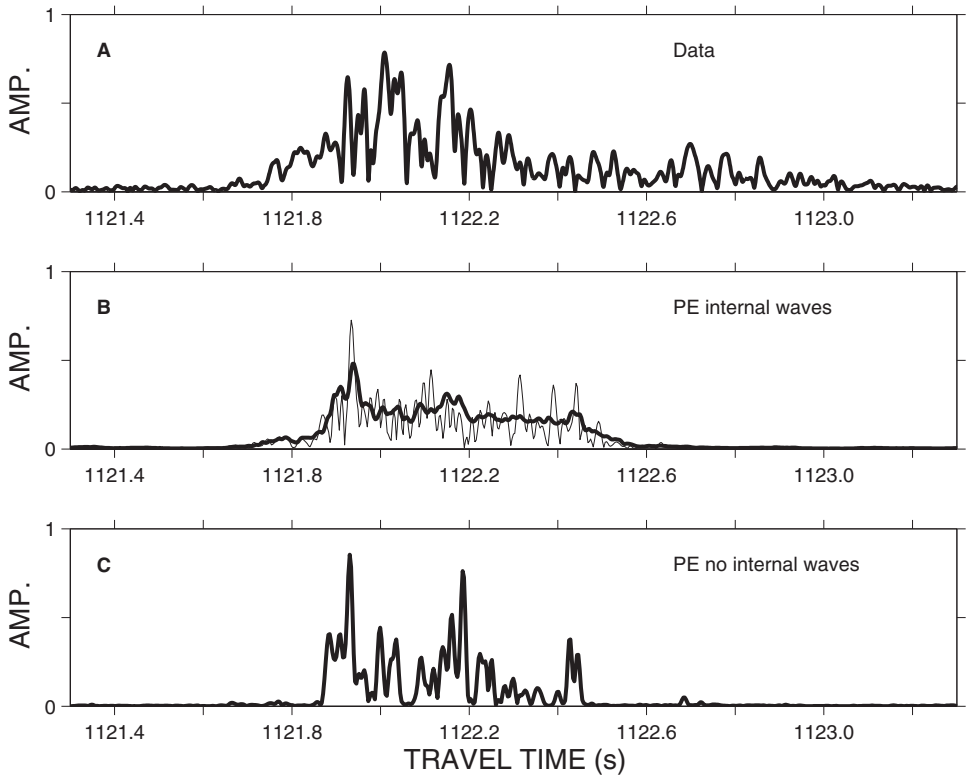


FIG. 2. Coherent average of impulse response from Hoke source (A) compared with two models [(B) and (C)]. Travel time of the data is adjusted to approximately coincide with the models. Only panels (B) and (C) have amplitudes that can be compared. PE is the parabolic approximation. Panel (B) shows an incoherent average (thick line) from many realizations of the impulse response computed at different geophysical times through an evolving field of internal waves. The thin line shows one of those realizations. Panel (C) is the impulse response computed through a climatological average of sound speed without perturbations from internal waves.

Internal waves are modeled with the Garrett–Munk[1,2] spectrum, with details published elsewhere.[17] Currents are ignored, being two orders of magnitude less than sound-speed perturbations arising from adiabatic vertical displacements of water in the upper ocean. The perturbations are added to the climatology of sound speed described above. The complete set of internal wave modes is precomputed and retrieved as needed at range intervals of 80 km to account for changes in water depth, buoyancy frequency, and sound speed. Vertical displacements of these modes are set to zero at the surface and bottom. For each 80-km interval, a three-dimensional field of internal waves is computed in a box of 80 km $\times$ 80 km $\times D$ m where D is the average depth of the ocean in that interval. A vertical slice through the box gives the vertical displacements along the geodesic for any desired section. Temporal evolution of the field is governed using the linear dispersion relation. The energy of the internal wave field is taken to be that specified by Garrett and Munk.[1,2]

Bathymetry along the section consists of a steeply sloping region at the seamount, followed by an abyssal region until 900 km (Fig. 3). This is followed by a region marked by ridges and seamounts over a bottom with depths between 2500 and 3000 m. Older and newer bathymetric databases are shown to indicate substantial differences (Fig. 3). Although the model uses the newest data, it is not clear if it is accurate enough to yield an accurate impulse response of the acoustic field.

The parabolic approximation of the acoustic wave equation requires parameters to describe acoustic propagation in the bottom of the ocean. These are provided to make it easier for others to replicate our model. The thickness of the sediment, and the ratio of the sediment to water density are taken from the Laske-Masters database at 50-km intervals.[18] The thickness varies non-monotonically along the section within the interval 155–407 m. The thickest sediments occur at distances between 550 and 1300 km from the source. The sound speed at the top of the sediment divided by that at the bottom of the water column is 1.02. The density of the sediment varies from 1.8 to 1.7 gm cm^{-3}. The attenuation in the sediment is

$$\alpha(f) = \alpha_0 f^p \quad (\text{dB m}^{-1}), \tag{2}$$

where f is the frequency in kHz, $p=1$, and $\alpha_0 = 0.02$ dB m^{-1} kHz^{-1}. The speed in the sediment is taken to increase with depth as 1 s^{-1}. The speed in the basement divided by that at the bottom of the sediment layer is 2. The density of the basement layer is 2.5 gm cm^{-3}. The attenuation in the basement is given by Eq. (2) except $\alpha_0 = 0.5$ dB m^{-1} kHz^{-1} and $p=0.1$. While all these geoacoustic parameters may not match those along the section, coherence of modeled multipath is probably insensitive to their values. They would likely change their amplitudes, but this does not seem to be important for comparing measured and modeled probability distributions of coherence in time nor modeled values of spatial coherence as long as most observed paths are present in the model.

B. Acoustic models

This parabolic approximation[21] outputs a two-dimensional field of sound along the geodesic from 0- to 8000-m depth. Tests suggest that travel times of pulses are computed with an accuracy of a few milliseconds.[21] The result is insensitive to reasonable variations in a reference speed of sound, which is why it is called the sound-speed insensitive approximation. The impulse response is com-

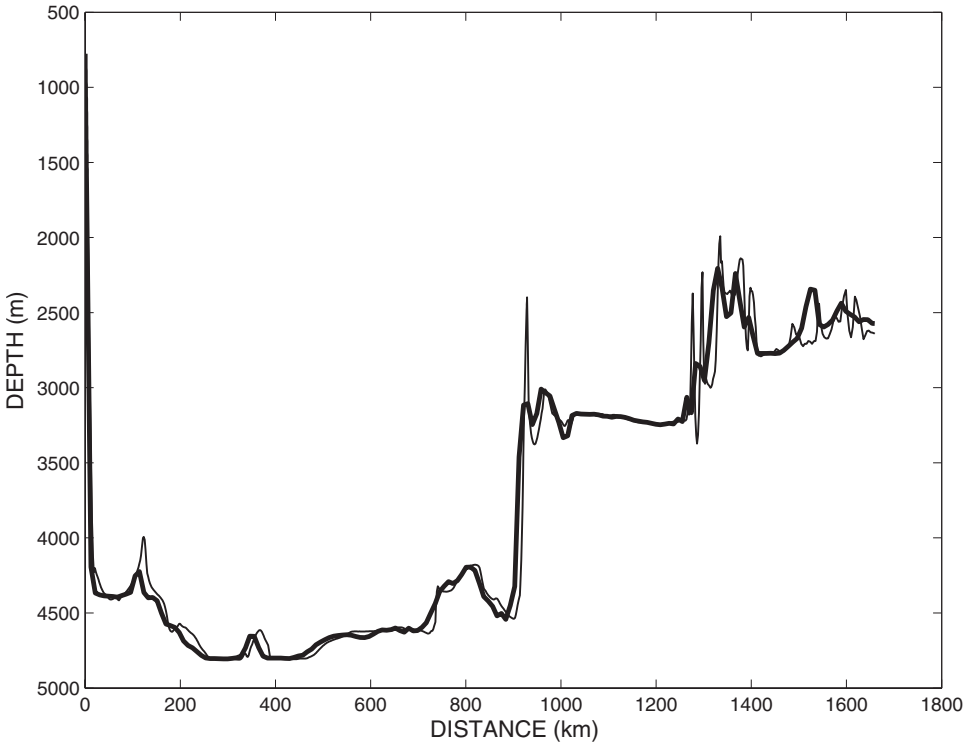


FIG. 3. Two estimates of bathymetry along section E in Fig. 1. Both estimates incorporate experimental measurements of bathymetry within 2.8 km of the acoustic source moored over the Hoke seamount (Ref. 11). The thicker line is from a 1987 database of depth (Ref. 19). The thinner is from a 2006 database (Ref. 20).

puted by applying an inverse Fourier transform to many single-frequency computations. Calculations of horizontal coherence of sound at the receiver are made assuming sound propagates without effects due to diffraction, refraction, and scattering in the horizontal coordinate. Instead, computations are made by approximating the solution of the acoustic wave equation with two-dimensional vertical slices through the modeled ocean. Despite the ubiquitous use of this vertical slice approximation, a rigorous justification has apparently not been published for the frequency considered here (250 Hz). Vertical slices of sound speed are obtained from the three-dimensional field of internal waves (Sec. III A). The convergence of the parabolic approximation is found by halving the grid sizes until the answers do not change significantly. We use a vertical grid spacing of 1.95 m. The grid separation in the horizontal dimension varies between 10 and 50 m. A separation of 10 m is used when the bathymetry is steep, such as near the Hoke seamount.

IV. IMPULSE RESPONSE

The SNR increases monotonically with the number of M-sequence periods coherently averaged. As mentioned in Sec. II, the peak SNR from each processed period is 29 dB. If each period was perfectly coherent and the noise was uncorrelated from period-to-period, a coherent average of all nine periods would have a SNR of $29+10\log_{10}9=38.5$ dB. This is about the same as the 38 dB measured from the coherent average [Fig. 2(a)]. The coherent averaging scheme that uses variable Doppler for each M-sequence period appears to yield a SNR close to the best that can be expected from theory.

The impulse response starts near 1121.75 s and ends near 1123.5 s [Fig. 2(a)]. This is aligned by eye to the best model result available in this paper [Fig. 2(b), thick line]. [Comparison of absolute times is impossible because location of the receiver is uncertain within 1 km (Sec. II).] This model is an incoherent average of 61 impulse responses computed through internal waves at 3-h intervals. Each impulse response is synthesized from 512 single-frequency runs of the parabolic approximation.[21] We find that separate incoherent averages from the first 31 and last 30 impulse responses are similar (not shown). Therefore, the incoherent average converged. 1 of the 61 impulse responses is shown [Fig. 2(b), thin line] to give an idea of how much the incoherent average smoothes a typical impulse response. Note the energy lasts longer in the data than the model by about 1 s [Fig. 2(a)]. This could be due to errors in bathymetry or too much attenuation in the bottom for later-arriving multipath. Another possibility is that the acoustic energy undergoes an extension in time due to a bias incurred from oceanic mesoscale. This hypothesis has been discussed, but not confirmed definitely.[17] Our environmental models do not include a mesoscale. The author believes it unlikely that uncertainty of energy in the internal wave spectrum would lead to an extension of 1 s, but this possibility cannot be excluded with certainty without further modeling. This is beyond the scope of this paper.

Without internal waves, the impulse response is shorter [Fig. 2(c)]. One may question why the jagged nature of the measured impulse response seems to better resemble the model in panel (C), without internal waves, than the thick line in panel (B), which includes internal waves. The reason is due to the smoothing of the impulse response that created the thick line in panel (B) through its incoherent averaging of 61 separate impulse responses. A single modeled impulse response through internal waves is usually less smooth [thin line, panel (B)]. The model in panel (C) is not an average, and neither is the measurement in (A). From past

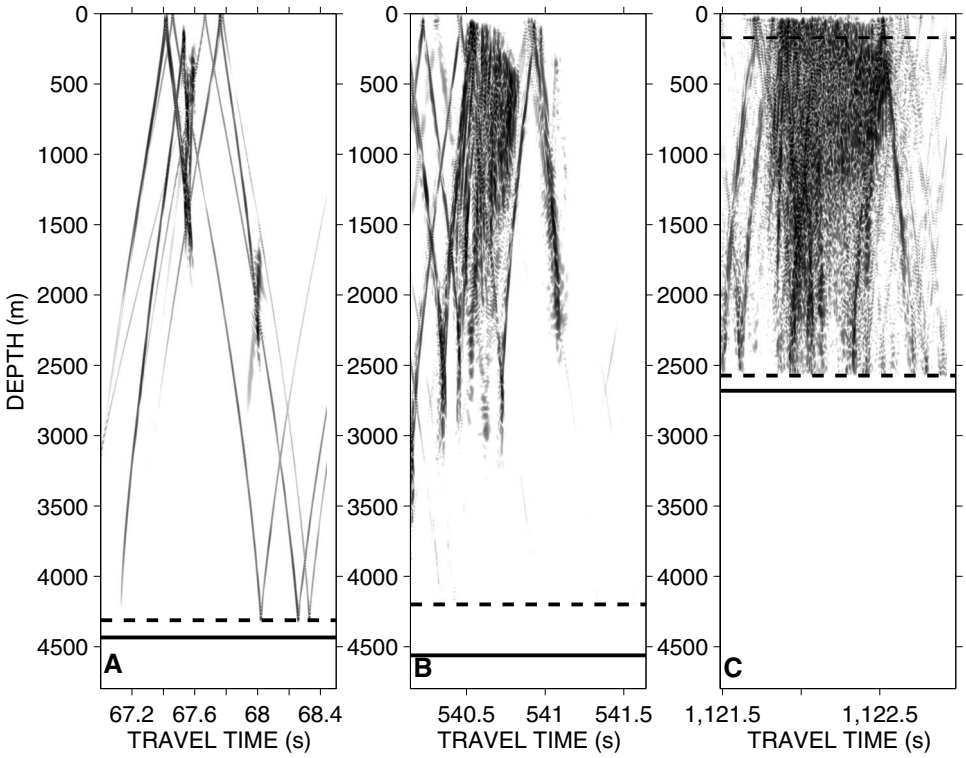


FIG. 4. Time fronts at distances of 100, 800, and 1659 km from source in panels (A), (B), and (C), respectively. Levels shown are in upper 38 dB at each distance. Depths of the water and basement are dashed and solid lines, respectively. Each panel shows 1.5 s of acoustic travel time. Top dashed line in panel (C) is at depth of the receiver. Time fronts are modeled with the parabolic approximation (Ref. 21) for a snapshot of sound-speed fluctuations due to internal waves added to a climatological background.

experience,[13] we believe that a daily or longer incoherent average of measured impulse responses would better resemble the thick line in panel (B) than in panel (C).

A. Interactions with surface and bottom

Time fronts indicate that the signal interacts with the surface and bottom of the ocean (Fig. 4). The top 38 dB are shown at each range because this is the SNR in the impulse response derived from the coherent average [Fig. 2(a)]. At distances exceeding about 1000 km, interactions with the surface appear to occur more frequently (not shown) because the acoustic waveguide rises toward the surface in the northern cold water. Reports from NOAA/NODC buoys and the volunteer ship observing program indicate crest to trough wave heights between 3 and 4 m along this section on 14 September 1999. The standard deviation of wave height is about 1 m.

The Rayleigh parameter, $P \equiv 2kh \sin\theta$, is useful for estimating effects of surface waves on sound, where the acoustic wavenumber is k, rms displacement of the surface is h, and grazing angle of sound with respect to the surface is θ.[22] For the center frequency of 250 Hz, and a rather large grazing angle of 5°, $P \sim 0.17$. This extreme case is much larger than predicted from ray traces (not shown), so the actual values for P would be less. For P much less than 1, the scattered wave can be thought of as specularly reflecting from the surface with rms variation of radian acoustic phase given by P.[22] The calculation is a ray approximation where the distance of a path is modified by a single interaction with the rough surface. At finite frequency and finite bandwidth, the region that influences each multipath expands from a point to a finite horizontal region. The region expands with decreasing frequency and increasing distance of transmission.[23] For example, at 2500 Hz and a transmission distance of 600 km, the region of influence is about 10 km (Fig. 9 in Ref. 23). The region would be larger in this experiment. We approximate the net effect of sound interacting with the rough surface over one 10-km region as follows. Since a typical crest-to-crest distance is about 50 m, there are $n \equiv 10\,000/50 = 200$ waves that interact with sound in 10 km. This reduces the rms value of P for a single wave by the factor $1/\sqrt{n}$. The rms phase variation of 0.17 rad for one interaction is reduced to $0.17/\sqrt{200} = 0.012$ rad. At 1659 km range, and an acoustic interaction with the surface every 50 km, there are at most $1659/50 \sim 33$ encounters of sound with the surface. At 50-km spacings, the net effect of phase with each surface interaction is statistically independent. Therefore, the net rms phase from 50 interactions is a random-walk process that increases by $\sqrt{33}$ the rms phase change from 0.012 to $0.012\sqrt{33} = 0.07$ rad. This is a negligible variation in multipath phase at the receiver. Effects of surface waves are too small to affect measurements of coherence.

Bottom spectra are less well known than the spectrum of surface waves. However, it does not appear that acoustic interaction with the bottom during the 135-s long transmission affects arrival structure. For the transmission, the ship moves 370 m away from the source (Sec. II). The scale of influence for the bottom is probably similar to that for the surface, i.e., 10 km or more. Since $370/10\,000 \ll 1$, it seems unlikely that the bottom significantly affects coherence.

V. TEMPORAL CORRELATION

In Sec. IV, we found that all nine impulse responses could be coherently averaged to increase the SNR over that for any individual impulse response. In this section, we increase the degrees of freedom by computing coherence time in small time windows from each impulse response.

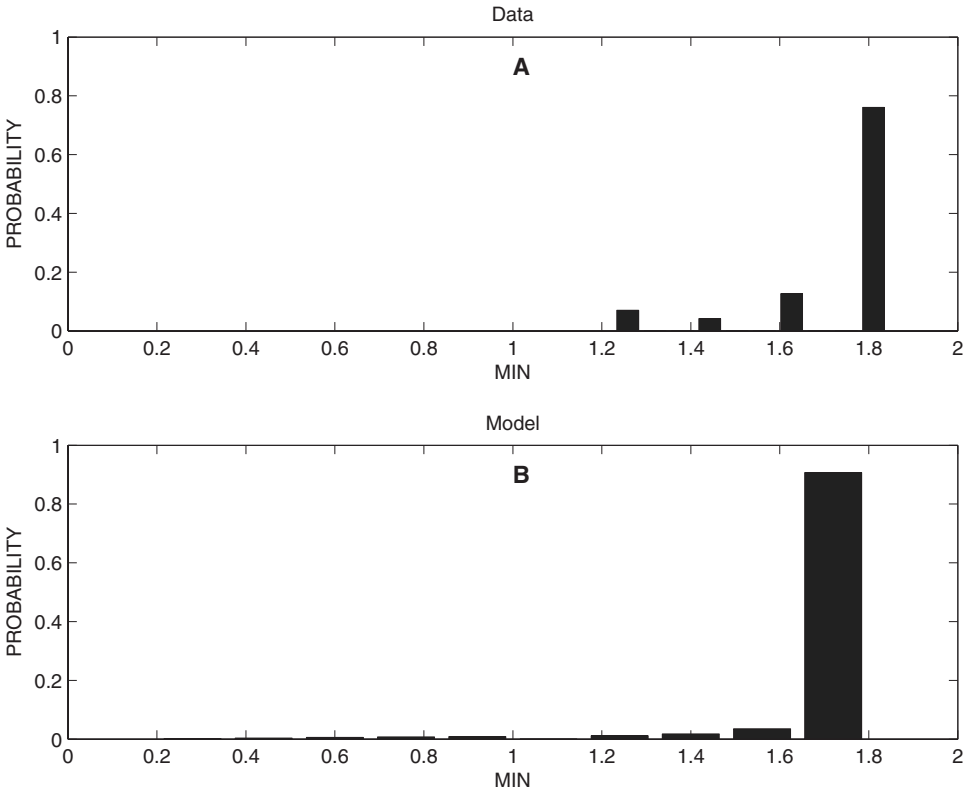


FIG. 5. Probability distribution of coherence time from section E (Fig. 1) for the data (A) and model (B). Data come from a single transmission of the Hoke source covering 1.833 min. Coherence time of the data is discretized to 12.276 s, the period of each of the nine transmitted M-sequences. Coherence time of the model is discretized at 12 s intervals. Coherence times from data and model are analyzed on the same basis. Actual coherence times may extend past 1.833 min, but cannot be explored with a transmission of 1.833 min. The sum of the probabilities is one for each panel separately.

A. Data

Data are Doppler corrected as described in Sec. II. The resulting impulse response for each of 9 periods is subdivided into 71 windows of travel time of duration 0.02 s each. The windows are chosen to cover energetic arrivals lasting 1.42 s. Even though the peak SNR of the entire impulse response increases monotonically with the number of added M-sequence periods (Sec. IV), this is not necessarily what happens when subsections of the impulse response are considered separately. For each window, we compute the peak SNR for each coherent average via Eq. (1) starting from periods 1 to 9. Letting m denote the period yielding maximum SNR, coherence time for that window is computed using $T=m12.276/60$ min. An empirical probability distribution is plotted for these 71 values [Fig. 5(a)]. The most likely coherence time is 1.8 min. It occurs with probability 0.8. Lesser coherence times are distributed between 1.2 and 1.6 min. The probability distribution is insensitive to modest changes in window duration. For example, when the duration is changed from 0.02 to 0.05 s, the distribution looks almost the same (not shown).

We now turn to the question as to whether the observed variation of 0.1 m/s of effective Doppler during 1.8 min is caused by the ocean or the instruments. Our previous experience with tautly moored sources in the Pacific and Atlantic is consistent with maximum speed, v_0, of a few cm/s near semi-diurnal and diurnal periods. Therefore, change in source speed in geophysical time interval τ has maximum value, $\delta v=2\pi\tau v_0/T$, where T is the period. For $v_0=0.02$ m/s, $\tau=1.8$ min, and $T=12$ h, $\delta v=3.1\times10^{-4}$ m/s. This is too small to explain a 0.1 m/s change in Doppler. On the other hand, a change in 0.1 m/s is entirely possible for the towed receiver. Changes in barotropic currents or other short-term fluctuations of currents and sound speeds, including internal waves following a Garrett–Munk spectrum, are more than a factor of 10 too small to account for the observed Doppler change of 0.1 m/s (Sec. VA of Ref. 4). Acceleration of the receiver is the only mechanism we can think of that could cause the observed variation in Doppler. It is likely that the variable Doppler correction merely removes effects from receiver acceleration and does not contaminate the measured estimates of coherence time.

B. Model

The parabolic approximation yields the impulse response for 122 records at 12 s intervals. (Internal waves evolve by 12 s between computations.) This allows a comparison of coherence time near the same resolution as the data (12.276 s).

The impulse response of each record is subdivided into $W\equiv41$ adjacent windows of width 0.02 s. This covers the modeled impulse response lasting 0.8 s [Fig. 2(b)]. White Gaussian noise with mean zero and variance σ^2 is added to each record. The SNR of each realization is set to be the same as the data in the following sense. Let the peak amplitude of record j be $\hat{a}_j$, $j=1,2,3,\ldots,122$. The record-averaged peak amplitude is $\bar{a}=122^{-1}\Sigma_{j=1}^{122}\hat{a}_j$. The variance is determined by solving for σ^2 in $29=10\log_{10}(\bar{a}^2/\sigma^2)$ (dB). This ensures that the record-average peak SNR is the same as measured.

A bootstrap scheme is used to estimate coherence time for each of 41 windows. First, we select at random $B=3000$ different starting records among 122 possibilities. The direction of the coherent average is selected at random to go forward or backward in time with respect to the starting

record. Nine total records are added together in the randomly chosen direction. End point problems are handled by choosing a direction that would not extend below 1 or above 122. With nine records, we are exploring coherence times up to 9 records $\times$ 12 s/record/(60 s/min) = 1.8 min. For each starting record, coherence time is computed by selecting the number of records, n, yielding the largest SNR where n can go from 1 to 9. The coherence time is $12n/60$ min. Letting coherence time for bootstrap b of travel time window k be T_{bk}, there are $BW = 3000 \times 41 = 123\,000$ estimates of T_{bk}. An empirical probability distribution is computed from these [Fig. 5(b)]. It is similar to the data. The most likely coherence time is 1.8 min, occurring with a probability of 0.9. Histogram-bars have slightly different centers for the model and data because the data and model are available at 12.276 and 12 s intervals, respectively.

VI. SPATIAL CORRELATION

A. Data

Coherent averages from 9 M-sequence periods were computed from two sub-arrays whose cross- and along-geodesic separations are 142 and 370 m, respectively (Sec. II). Each coherent average is computed by beamforming, using a matched filter with the emitted waveform, and by using a variable Doppler scheme for each period to optimize average SNR. The peak SNR of each coherent average is 38 dB. The energetic portion of each coherent average is about 1.5 s. A normalized cross-correlation coefficient is computed between the single coherent average from one array with the single coherent average from the other array. The value of the correlation coefficient is 0.54. The SNR is so high that virtually none of this degradation in coherence is explained by noise.

B. Models

Degradation of spatial coherence in the presence of internal waves is computed using our model that places a horizontal array at fixed distance from the acoustic source. In other words, it does not have the flexibility of letting the array be anything except perpendicular to the geodesic between source and receiver. The array is, however, not perpendicular to the geodesic. For convenience, we therefore divide the modeling of spatial coherence into effects due to cross-geodesic and along-geodesic components. Dividing analysis into two components allows identification for independent causes for de-coherence.

1. Cross-geodesic separations

We estimate the extent to which a horizontal separation of 142 m (perpendicular to the section) can explain the measured correlation coefficient of 0.54. At 1659 km distance, a 10-km horizontal array is placed perpendicular to the geodesic with elements at 10 m spacing. Vertical sections of sound speed are taken from the three-dimensional field of internal waves between the source and each element on the array. No attempt is made to model effects due to horizontal coupling between the vertical sections. A similar approach has been discussed elsewhere.[7] The approximation has only been shown to be valid up to a frequency of 75 Hz.[24] It might be valid at higher frequencies, but a direct numerical confirmation apparently awaits future investigation.

The acoustic field at 250 Hz only is computed at each array element for eight geophysical times at 8.4 h intervals. An 8.4 h interval is more than enough to yield uncorrelated impulse responses for this model. Using the bootstrap, we find the normalized correlation coefficient for cross-geodesic separation falls to e^{-1} at 0.5 km (Fig. 6). Note the tight bounds on correlation coefficients at the 95% confidence limit. Since there are eight uncorrelated realizations of the acoustic field on a 10-km long array, there are about $8 \times 10/0.5 = 160$ degrees of freedom. Reading from the figure, we see that the correlation coefficient is between 0.904 and 0.918 at 142 m. We conclude that a cross-geodesic separation of 142 m cannot explain the measured correlation coefficient of 0.54. We will see next that another mechanism does explain a coefficient of 0.54 when combined with the values between 0.904 and 0.918 here.

2. Along-geodesic separation

We estimate coherence of the signal between two points on the geodesic separated by 370 m. The parabolic approximation is used to compute the impulse response through the same fluctuating internal wave field as before [Fig. 7(a)], except the final range is decreased by 370 m. Comparing 61 impulse responses separated by 370 m at 3-h intervals, the mean and standard deviation of the normalized cross-correlation coefficient are 0.73 ± 0.079. The 95% confidence limits are in the interval [0.57, 0.83]. The lower limit is close to the measured value of 0.54.

There appear to be three hypotheses for degradation of modeled coherence in the along-geodesic component. (1) Acoustic signals are affected by different components of the internal wave field. (2) Travel times of multipath are sensitive to interactions of sound with bathymetric features in the presence of internal wave fluctuations. (3) The travel time change for energy arriving at different inclination angles is differentially affected for along-geodesic displacements in the absence of internal waves. The first hypothesis does not explain the degradation because we compute about the same degradation when internal waves are absent. The second hypothesis does not explain the degradation because we obtain the same answer when the bathymetry is changed to be flat at 5 km of depth. The third hypothesis does appear to explain the degradation. Using realistic bathymetry, but not internal waves, the computed correlation coefficient is 0.7. This value is within one standard deviation of the correlation coefficient computed with internal wave fluctuations reported above (i.e., 0.73 ± 0.079).

An analytical calculation seems to confirm that degradation of correlation is primarily explained by the third hypothesis. The change in acoustic phase for ray i at frequency f due to a receiver horizontally displaced by δx along a geodesic is

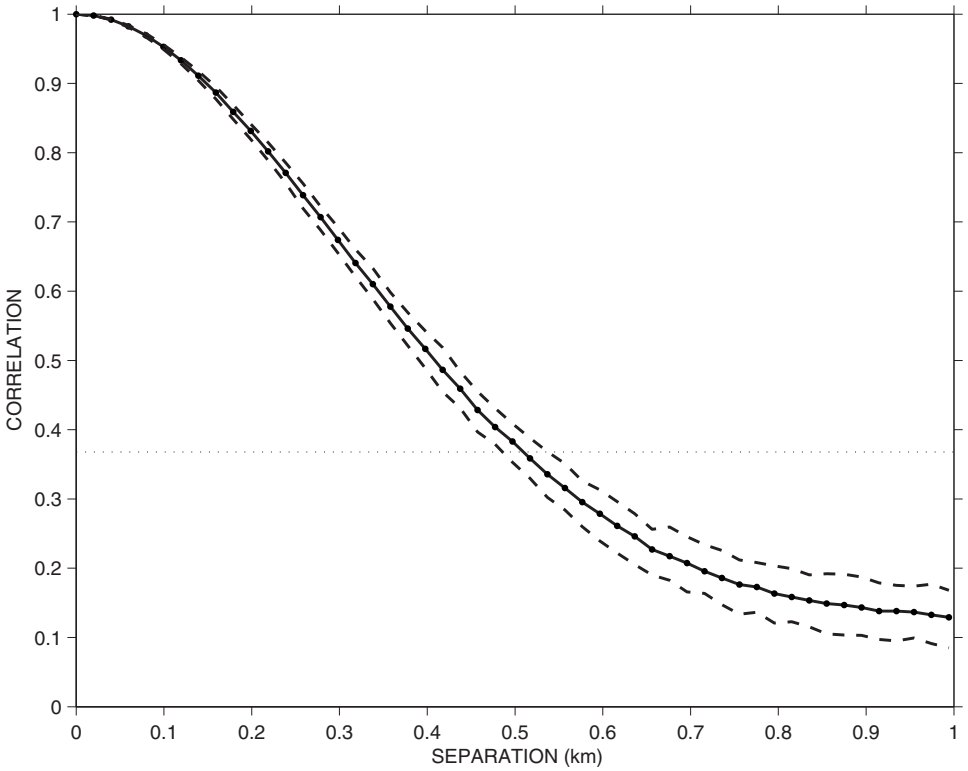


FIG. 6. Modeled estimates of the normalized horizontal correlation coefficient of the acoustic field for section E in Fig. 1 at 250 Hz with 95% confidence limits shown. The dotted line is at exp(−1). The correlation decreases with separation because of the effects of internal gravity waves.

$$\delta\phi_i \approx \delta x f \cos\,\theta_i/c \quad \text{(cycles)}, \tag{3}$$

where θ_i is the inclination angle of the ray at the receiver, measured positive up from the horizontal. The equivalent change in travel time is $\delta t_i = \delta\phi_i/f$. The result is based on the linearized assumption that the ray angle does not change significantly for horizontal displacement δx. Consider I temporally-resolved arrivals with amplitudes a_i, $i = 1,2,3,\ldots,I$ that have a simple time series

$$b(t) = \sum_{i=1}^{I} a_i \cos[\omega(t - T_i)]\Pi\left[\frac{t - T_i}{3\Delta/2}\right], \tag{4}$$

where the boxcar function, Π, equals unity when its argument has absolute value of 1 or less, and is 0 otherwise. The travel time of ray i is T_i, the period of the sinusoid is $\Delta = 2\pi/\omega$, and the speed of sound at the receiver is c. The boxcar is unity for three periods, which is the same as the emitted signal for this experiment. Thus, each arrival is represented by three cycles of carrier. When the receiver is moved by δx along the geodesic, the predicted time series is

$$q(t) = \sum_{i=1}^{I} a_i \cos[\omega(t - T_i - \delta t_i)]\Pi\left[\frac{t - T_i - \delta t_i}{3\Delta/2}\right], \tag{5}$$

assuming a_i are unchanged. To calculate the maximum value of the cross-correlation coefficient between $b(t)$ and $q(t)$, we approximate the maximum lag to occur at δt_1. We further assume that the cross-correlation coefficient is primarily degraded due to changes in phase between corresponding paths once corresponding paths in $b(t)$ and $q(t)$ are approximately lined up at lag δt_1. This approximation neglects degradation due to the fact that the boxcar envelopes for corresponding paths will not quite line up due to differential effects of travel time caused by various values of θ_i. Noting that the time-mean values of $b(t)$ and $q(t)$ are zero, the normalized correlation coefficient has maximum value

$$\rho = K/J, \tag{6}$$

where

$$J = \int b^2(t)dt = \int q^2(t)dt = \sum_{i=1}^{I} a_i^2 \int_{-3\Delta/2}^{3\Delta/2} \cos^2\frac{2\pi t}{\Delta}dt$$

and

$$K = \int b(t)q(t + \delta t_1)dt.$$

Then,

$$K \approx \sum_{i=1}^{I} a_i^2 \cos(\delta\phi_i - \delta\phi_1) \int_{-3\Delta/2}^{3\Delta/2} \cos^2(\omega t)dt.$$

Substituting K and J into ρ we get

$$\rho \approx \frac{\sum_{i=1}^{I} a_i^2 \cos[\delta x f(\cos\,\theta_i - \cos\,\theta_1)/c]}{\sum_{i=1}^{I} a_i^2}. \tag{7}$$

For the simple case of $a_i = 1$,

$$\rho \approx \frac{1}{I}\sum_{i=1}^{I} \cos[\delta x f(\cos\,\theta_i - \cos\,\theta_1)/c] \quad \text{for } a_i = 1, \tag{8}$$

so all the degradation is due to differences in arrival angle. For the simple case of two arrivals, we solve for δx

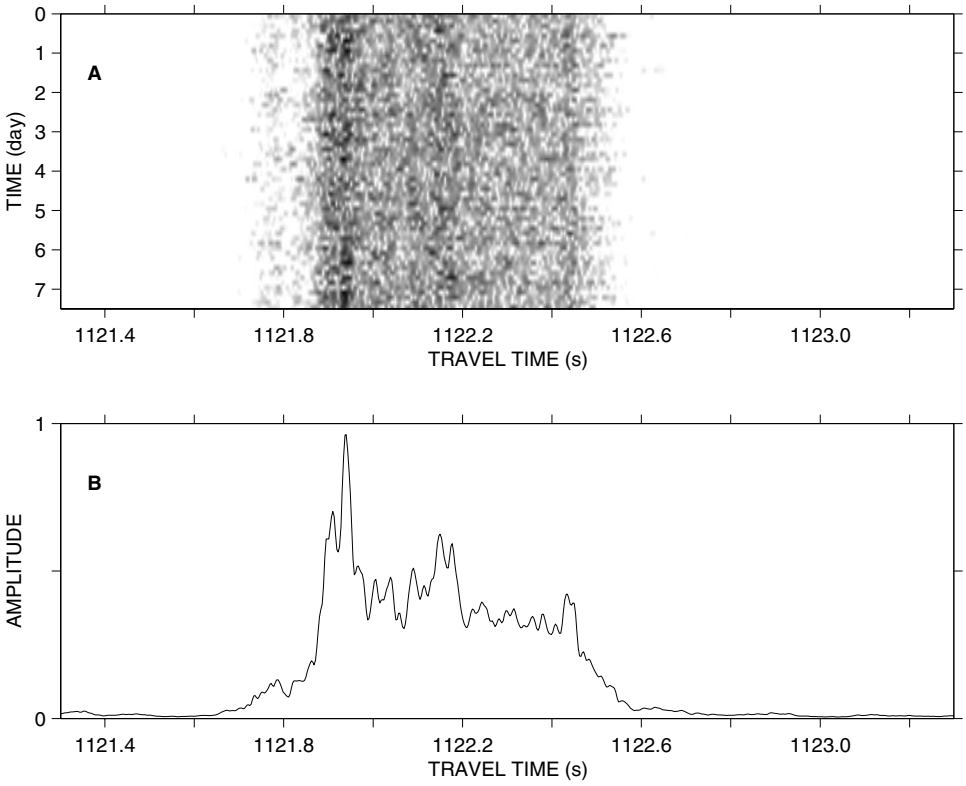


FIG. 7. (A) Contours of top 30 dB of 61 impulse responses, at 3-h intervals, computed from the parabolic approximation model for section E in Fig. 1. Modeled variations are due to the time evolution of a standard internal wave field. (B) Incoherent average of the 61 impulse responses used in (A) [same as thick line in Fig. 2(b) except amplitudes are scaled with a different value].

$$\delta x \approx \frac{c \cos^{-1}(2\rho - 1)}{f(\cos\ \theta_2 - \cos\ \theta_1)}. \tag{9}$$

Using a modeled value of $\rho=0.7$, and for typical arrival angles of eigenrays, $\theta_1=2°$ and $\theta_2=9°$, the needed along-geodesic displacement, δx, is 610 m. This is less than a factor of 2 from the measured displacement of 370 m and confirms this hypothesis for explaining decorrelation in the along-geodesic direction.

3. Combined along- and cross-geodesic effects

We seek 95% confidence limits for the cross-correlation coefficient due to the combined effects of along- and cross-geodesic separations. It is likely that effects from along- and cross-geodesic separations are approximately statistically independent. So, it is likely that $\rho \approx \rho_a \rho_b$ where ρ_a and ρ_b are the coefficients due to along- and cross-geodesic separations, respectively. The formal method for estimating confidence limits for ρ is to empirically draw pairs of model realizations of ρ_a and ρ_b, then form their product to obtain a realization of ρ. From these, 95% confidence limits are obtained empirically. We sidestep this procedure because the 95% confidence limits for cross-geodesic separation are narrow. We approximate $\rho_b=0.91$, its mean value at a separation of 142 (Fig. 6). Since the 95% confidence interval for ρ_a is [0.57, 0.83], we approximate the 95% interval for ρ as $0.91 \times [0.57, 0.83] = [0.52, 0.76]$. This is statistically consistent with the measured value of 0.54. We did not attempt to find a confidence limit for the measured value because only one realization of spatial coherence is available and the SNR is so high that effects from noise are negligible.

VII. CONCLUSIONS

We tested the hypothesis that the Garrett–Munk spectrum of internal gravity waves accounts for observations of temporal coherence of sound for a 1659 km section in the Pacific Ocean [Fig. 1(e)]. Sounds emanated from a 250 Hz source following a phase-coded signal with 0.02 s resolution. Without any tuning of the oceanographic or acoustic models to fit the data, we obtain similar modeled and measured probability distributions of temporal coherence.

The model for spatial coherence is statistically consistent with the measured value of 0.54. The model needs two components to yield this result. The first is the decorrelation of the signal due to internal waves due to displacements perpendicular to the section. The second is a deterministic effect due to the difference in distance between the source and two arrays at either end of the complete towed array. The array is not perpendicular to the section. In this experiment, the deterministic effect leads to a larger loss of spatial coherence than the effects from internal waves.

Modeled degradation of spatial coherence due to internal waves is computed assuming negligible interaction of sound between separate vertical slices of the acoustic field. An indirect confirmation of this approximation appears to come from the statistical consistency between measured and modeled values of spatial coherence.

Varying model parameters to test sensitivity of the results does not seem critical in predicting coherence in light of the resemblance with data using archival parameters. This paper is not a study in sensitivity analysis. It simply seeks to determine whether the models are reliable predictors of coherence, and they are. This is an important finding. Another study might investigate modifying the spectrum for internal

waves. However, variations should be done realizing the spectrum from the literature was fitted to myriads of hydrographic data collected world-wide.

It is important to continue comparing with models coherence from other experiments since coherence is important for numerous applications (Sec. I). Comparison here is made at higher frequencies and shorter ranges than before (Sec. I).

It is possible that the probability distribution for coherence time could exceed 1.8 min. We did not address this question because the hypothesis is untestable with our data. What seems to be important is that the modeled probability distribution looks like that derived from data when both are analyzed in the same way.

Finally, the Monte-Carlo impulse responses are run on a supercomputer. Others are working on faster methods for implementing Monte-Carlo methods.[25]

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research Contract No. N00014-06-C-0031 and by a grant of computer time from the DOD High Performance Computing Modernization Program at the Naval Oceanographic Office. I thank Curt Collins, Chris Miller, and Ching-Sang Chiu, of the Naval Postgraduate School, for providing bathymetry of the Hoke seamount and providing answers to my inquiries on their experiment with this source. I thank Eric Freeman (NOAA) for helping me obtain surface wave data from ships. I thank the editor, Roger Gauss, and reviewers for their valuable and detailed suggestions.

[1]We might have been able to use ten *M*-sequences to avoid sidelobes. However, the matched filter output contains sidelobes in the time domain if the replica is correlated with data containing no signal (i.e., before signal arrival and after signal termination). To be safe we analyzed nine interior *M*-sequence periods.

[1]C. Garrett and W. Munk, "Space-time scales of internal waves," Geophys. Fluid Dyn. **3**, 225–264 (1972).
[2]C. Garrett and W. Munk, "Space-time scales of internal waves: A progress report," J. Geophys. Res. **80**, 291–297 (1975).
[3]J. L. Spiesberger and D. Green, "Statistical characterization of very low frequency communication channels at ocean basin-scales," Proceedings of the MTS-IEEE Oceans '06 Conference, Boston, MA (2006), pp. 1–6, Paper No. 060322-06.
[4]J. L. Spiesberger, "Comparison of measured and modeled temporal coherence of sound near 75 Hz and 3683 km in the Pacific Ocean," J. Acoust. Soc. Am. **124**, 2805–2811 (2008).
[5]J. L. Spiesberger, F. Tappert, and A. R. Jacobson, "Blind prediction of broadband coherence time at basin-scales," J. Acoust. Soc. Am. **114**, 3147–3154 (2003).
[6]J. L. Spiesberger, "Numerical prediction of coherent integration time at 75 Hz, 0.03 temporal resolution at 3250 km," Proceedings of the MTS-IEEE Oceans '06 Conference, Boston, MA (2006), pp. 1–4, Paper No. 060323-02.
[7]M. Vera and M. Dzieciuch, "Horizontal coherence in the NPAL experiment," J. Acoust. Soc. Am. **115**, 2617 (2004).
[8]V. E. Ostashev, A. G. Voronovich, and the NPAL Group, "Spatial coherence of acoustic signals measured during the 1998–1999 North Pacific Acoustic Laboratory (NPAL) experiment," J. Acoust. Soc. Am. **116**, 2608–2609 (2004).
[9]R. K. Andrew, B. M. Howe, J. A. Mercer, and the NPAL Group, "Transverse horizontal spatial coherence of deep arrivals at megameter ranges," J. Acoust. Soc. Am. **117**, 1511–1526 (2005).
[10]A. G. Voronovich and V. E. Ostashev, "Horizontal refraction of acoustic signals retrieved from the North Pacific Acoustic Laboratory billboard array data," J. Acoust. Soc. Am. **117**, 1527–1537 (2005).
[11]T. Rago, C. S. Chiu, C. S. Collins, P. F. Worcester, and C. Castro, "Oceanographic data from Sur Ridge (36.3° N, 122.4° W) to Hoke Seamount (32,1° N, 126.9° W)," Technical Report No. NPS-OC-00-002, Naval Postgraduate School, Monterey, CA, 1999.
[12]S. K. Han, C. Collins, C. Miller, C. S. Chiu, and P. Worcester, "Mapping the regional variability of the California Current acoustically using a waveform inversion method," J. Acoust. Soc. Am., **107**, 2862 (2000).
[13]J. L. Spiesberger, "An updated perspective on basin-scale tomography," J. Acoust. Soc. Am. **109**, 1740–1742 (2001).
[14]V. A. Del Grosso, "New equation for the speed of sound in natural waters with comparisons to other equations," J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[15]S. Levitus, "Climatological atlas of the world ocean," NOAA Professional Paper 13 (U.S. Government Printing Office, Washington, DC, 1982).
[16]A. Ben-Menahem and S. J. Singh, *Seismic Waves and Sources* (Springer-Verlag, New York, 1981), p. 1108.
[17]M. A. Wolfson and J. L. Spiesberger, "Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations," J. Acoust. Soc. Am. **106**, 1293–1306 (1999).
[18]G. Laske and G. Masters, "A global digital map of sediment thickness," EOS Trans. Am. Geophys. Union **78**, F483 (1997).
[19]National Geophysical Data Center, "5 minute gridded world elevations and bathymetry—A digital database," ETOPO5, Boulder, CO, 1987.
[20]Naval Oceanographic Office Data Model, Oceanographic and Atmospheric Master Library (OAML), "Digital bathymetric database variable resolution (DBDB-V)," version 5.1, April, 2006.
[21]F. Tappert, J. L. Spiesberger, and L. Boden, "New full-wave approximation for ocean acoustic travel time predictions," J. Acoust. Soc. Am. **97**, 2771–2782 (1995).
[22]L. M. Brekhovskikh and Y. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York, 1982).
[23]J. Spiesberger, "Regions that influence acoustic propagation in the sea at moderate frequencies, and the consequent departures from the ray-acoustic description," J. Acoust. Soc. Am. **120**, 1842–1850 (2006).
[24]J. L. Spiesberger, "Comparison of two and three spatial dimensional solutions of the wave equation at ocean-basin scales in the presence of internal waves," J. Comput. Acoust. **15**, 319–332 (2007).
[25]A. K. Morozov and J. A. Colosi, "Stochastic differential equation analysis for ocean acoustic energy scattering by internal waves," J. Acoust. Soc. Am. **119**, 3344 (2006).

Underwater acoustic beam dynamics

Francisco J. Beron-Vera[a] and Michael G. Brown
RSMAS/AMP, University of Miami, Miami, Florida 33149

(Received 30 August 2008; revised 5 March 2009; accepted 24 April 2009)

Ray- and mode-based theoretical predictions of the spreads of directionally narrow beams are presented and compared to parabolic-equation-based simulations in deep-ocean environments. Both the spatial and temporal spreads of beams are considered. The environments considered consist of a range-independent deep-ocean background sound channel on which a highly structured sound-speed perturbation, associated with either internal waves or homogeneous isotropic single-scale turbulence, is superimposed. The simulation results are shown to be in good agreement with simple theoretical expressions which predict that beam spreading, in both the unperturbed and perturbed environments, is largely controlled by a property of the background sound channel—the ray-based *stability parameter* α or the asymptotically equivalent mode-based *waveguide invariant* β. These results are consistent with earlier results showing that wavefield structure and stability are largely controlled by α (or β). © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3139901]

PACS number(s): 43.30.Re, 43.30.Bp, 43.30.Cq, 43.30.Ft [JAC] Pages: 80–91

I. INTRODUCTION

In this paper the spreading, both spatial and temporal, of directionally narrow beams of sound in deep-ocean environments is considered. Both types of spreading are shown to be controlled by a property, the *stability parameter* α (defined below), of the background sound-speed profile. The work presented represents the convergence of four seemingly unrelated lines of research. First is ray-based work[1–15] in which it is shown that various measures of wavefield structure and stability in both unperturbed (range-independent) and perturbed (e.g., by internal waves) sound channels are controlled by α. Second is mode-based work[16–30] in which it is shown that various mode-based descriptors of wavefield structure and stability are largely controlled by the mode-based *waveguide invariant* β (defined below). An important connection between the ray- and mode-based analyses was established in Ref. 31 by showing that when β is evaluated asymptotically (making use of a WKB analysis), $\beta=\alpha$. Third is work involving the propagation of beams in the study of finite frequency wavefields that, in the ray limit, exhibit chaotic behavior.[32,33] Fourth is work[34–37] on "weakly divergent beams," which have been shown to have special properties. The connection between the weakly divergent beams and work on α and β is that the former are defined by a condition which is shown below to be equivalent to $\alpha=0=\beta$. We do not in this paper consider the construction of solutions to the wave equation using a summation of Gaussian beams (see, e.g., Ref. 38 and references therein, and Ref. 39).

The terms "beam" and "ray bundle" were used loosely in Refs. 34–37 inasmuch as these papers considered wavefields produced by point sources which excite energy at all angles, i.e., a continuum of beams were excited. In contrast, in this paper, we consider directionally narrow beams whose generation requires an array with nonzero aperture. Specifically, we consider sound fields generated by a vertical array, forming a fairly narrow beam in the range-depth plane. The advantage of considering directionally narrow beams is that, because the ray stability parameter α is in general a function of ray launch angle, narrow beams are better suited to elucidate the role of α than are directionally broad beams, which generally contain rays corresponding to a broad range of α-values. (Similar statements can be made if one adopts a mode-based description of the propagation and replaces α by β.) We consider both the spatial spread of beams (in either cw or transient wavefields) and the temporal spread of beams in transient wavefields. Both types of spread are shown to be controlled by α (or β if a modal wavefield description is adopted).

In Sec. II, relevant theoretical results are presented and some numerical details are described. In Sec. III, numerical results are presented and discussed. In Sec. IV, we summarize our results, returning to the four lines of research mentioned above, and provide a brief discussion of the broader implications of our results.

II. THEORETICAL BEAM SPREAD ESTIMATES

The three subsections that follow focus on, respectively, ray- and mode-based background theoretical material, spatial beam spreads, and temporal beam spreads.

A. Preliminary theoretical results

We begin with a discussion of ray-based theoretical results. Extensive use is made of the action-angle description of ray motion as this description provides the most concise statement of ray-based estimates of beam spreads, and because action-angle variables provide a direct link with mode-based theoretical considerations that will follow. The material in this subsection is closely related to the background material presented in many papers including Refs. 4–6, 8, 14, and 30.

[a]Author to whom correspondence should be addressed. Electronic mail: fberon@rsmas.miami.edu

 0001-4966/2009/126(1)/80/12/$25.00

We consider propagation in the vertical plane (z,r) and make use of the one-way description of ray motion. The ray/travel time equations have Hamiltonian form

$$\frac{dp}{dr} = -\frac{\partial H}{\partial z}, \quad \frac{dz}{dr} = \frac{\partial H}{\partial p}, \quad \frac{dT}{dr} = p\frac{dz}{dr} - H, \tag{1}$$

with Hamiltonian

$$H(p,z,r) = -\sqrt{c^{-2}(z,r) - p^2}. \tag{2}$$

Here c is sound-speed, and $p=p_z$ and $-H=p_r$ are the vertical and horizontal components of the slowness vector $\mathbf{p}$; $\sigma\mathbf{p}=\mathbf{k}$ where $\sigma=2\pi f$ is the acoustic frequency and $\mathbf{k}$ is the wavenumber vector. The relationship between p and ray angle ϕ, measured relative to the horizontal, is $cp=\sin\phi$.

We focus here on the background range-independent, $c=c(z)$, problem. The phase space variables (p,z) can be replaced by a more convenient set of variables, so-called action-angle variables, (I,θ). To simplify our discussion, we shall assume that $c(z)$ has a single minimum so that rays have two (upper and lower) turning points. The canonical transformation from (p,z) to (I,θ) involves a generating function $G(z,I)$ which can be taken to be (there is some flexibility in the choice of integration limits which is tied to the choice of where, along a ray, $\theta=0$)

$$G(z,I) = \pi I \pm \int_{\check{z}(H)}^{z} \sqrt{c^{-2}(\xi) - H^2}\, d\xi, \tag{3}$$

where

$$I(H) = \frac{1}{\pi}\int_{\check{z}(H)}^{\hat{z}(H)} \sqrt{c^{-2}(z) - H^2}\, dz, \tag{4}$$

$\theta=\partial G/\partial I$, $p=\partial G/\partial z$, and the turning depths $\hat{z}$ and $\check{z}$ satisfy $c(\hat{z})=c(\check{z})=-1/H$. The $\pm$ signs in Eq. (3) correspond to $\pm p$. In terms of (I,θ), $H=H(I)$ and the ray-travel time equations are

$$\frac{dI}{dr} = -\frac{\partial H}{\partial \theta} = 0, \quad \frac{d\theta}{dr} = \frac{\partial H}{\partial I} = \omega(I),$$

$$\frac{dT}{dr} = I\omega(I) - H(I) + \frac{d}{dr}(G - I\theta). \tag{5}$$

Here $H(I)$ is defined by Eq. (4). Note that the same symbol H is used to denote the Hamiltonian whether the independent variables are (p,z) or (I,θ); this is done because the transformation from (p,z) to (I,θ) leaves unchanged the numerical value of H and its physical interpretation as minus the r-component of the ray slowness vector, $H=-p_r$. The frequency $\omega(I)$ is the spatial frequency of a ray, $\omega(I)=2\pi/R_\ell(H(I))$, where $R_\ell(H)=2\pi dI/dH$ is the horizontal ray cycle (double loop) distance. Equation (5) can be integrated by inspection. Solutions are

$$I(r) = I_0, \quad \theta(r) = \theta_0 + \omega(I)r,$$

$$T(r) = (I\omega(I) - H(I))r + (G - I\theta)\big|_{r=0}^{r}. \tag{6}$$

Note that in a range-independent environment the action I is constant following each ray. Thus, in such an environment I can be thought of as a ray label. Also, note that θ increases monotonically along each ray, increasing by 2π over each ray cycle.

A directionally narrow transient beam spreads both spatially and temporally as it propagates. Furthermore, in an environment in which scattering takes place, there are both deterministic and scattering-induced contributions to both spatial and temporal spreads. The beam itself can be thought of as being composed of a continuum of rays with a narrow band of actions $|I-I_0| \lesssim \Delta I/2$. (But note that, as described below, in the presence of scattering the I-domain grows with increasing range.) We assume that the source array excites energy within a narrow band of launch angles $\Delta\phi_s$. Using $cp=\sin\phi$ and the definitions of $H(p,z)$ and $H(I)$ [see Eqs. (2) and (4), respectively], $\Delta\phi_s$ can be converted to an equivalent spread in ray action ΔI. Beam construction is discussed in the Sec. II B. We shall be concerned with the spread of the beam as a function of range r.

We have so far focused on a ray-based description of propagation. It is insightful to consider, in addition, a mode-based description. The mode-based theory considered here is based on asymptotic results, i.e., WKB theory or its uniform asymptotic counterpart. This is done both to keep our discussion simple and to allow a simple correspondence between ray-and mode-based estimates of time spreads to be established. Surprisingly few results are required to derive mode-based estimates of beam spreads. First is the modal quantization condition, which for the two turning point problem is well known (see, e.g., Ref. 40),

$$\sigma I(p_r) = m + \frac{1}{2}, \quad m = 0,1,2,\ldots, \tag{7}$$

where $I(p_r)=I(-H)$ is defined in Eq. (4). The phase slowness $p_r=1/\hat{c}=1/\check{c}$ of a mode is generally different than its group slowness,

$$S_g(p_r) = T_\ell(p_r)/R_\ell(p_r). \tag{8}$$

Here $T_\ell(p_r)=2\pi I(p_r)+p_r R_\ell(p_r)$ and $R_\ell(p_r)=-2\pi dI/dp_r$ are the travel time and range, respectively, of a ray double loop whose turning depths coincide with the modal turning depths. Note that asymptotically (this is not an exact result) the dependence of S_g on σ and m enters only through its dependence on p_r, i.e., $S_g=S_g(p_r(m,\sigma))$. Equations (7) and (8) define parametrically—via p_r—a set of curves $S_g(f;m)$ which, when plotted together, constitute a dispersion diagram (see, e.g., Refs. 29 and 30). Modal group time delays are simply (see, e.g., Ref. 40)

$$T(m,\sigma) = S_g(p_r(I(m,\sigma)))r. \tag{9}$$

The validity of these mode-based results is restricted to range-independent environments. Scattering effects, resulting in mode coupling, will be discussed below.

B. Spatial beam spreads

In this subsection, we consider the lateral spread of a beam as a function of range r. The measure of lateral spread that we consider is the spread in r at a fixed value of θ. For example, constraining the beam to have $\theta=(17/2)\pi$ con-

strains all of the rays that comprise the beam to make four complete cycles plus an additional quarter cycle. For small ΔI fixing θ is an excellent approximation (but not identical) to fixing the starting and ending depths of all of the rays that make up the beam. If our constant θ beam width estimates are viewed as an approximation to constant z beam width estimates, then the approximation leads to small errors in the treatment of partial ray cycles. We emphasize, however, that our constant θ measure is as sensible a measure of beam width as the constant z measure. Also, it should be clear that we are *not* neglecting partial ray cycles.

Consider first the deterministic spread in range Δr_d of a beam. It follows from the second equation of Eq. (6) with θ fixed that $0=\omega(I)\Delta r+r\omega'(I)\Delta I$. Thus the deterministic spread of the beam whose width in I is ΔI, centered on I_0, is

$$\Delta r_d = -r\frac{\omega'(I_0)}{\omega(I_0)}\Delta I. \tag{10}$$

This is the deterministic beam spread as a function of range. The sign of Δr_d is controlled by the sign of $\omega'(I_0)$; the most common situation in deep water ocean acoustics is $\omega'(I)<0$, for which R_ℓ increases with increasing I or $\Delta r_d>0$. Equation (10) is consistent with the estimate of the deterministic beam spread derived in Ref. 15 using a different argument. (The expressions differ by a factor of 2 owing to the difference in the way ΔI is defined.)

We consider now the scattering-induced spread in range Δr_s of a narrow beam. We assume that the scattering of acoustic energy is caused by small scale sound-speed structure $\delta c(z,r)$ that is superimposed on the background $c(z)$. Rather than reformulating the ray results in the perturbed environment, we shall employ a simple phase-screen model of ray scattering; at selected ranges the small-scale inhomogeneity causes a ray to be kicked from one action surface to another. A similar scattering model will be employed when time spreads are considered. As described below, a sequence of such random scattering events leads to the diffusive spreading of the I-domain of the beam. This diffusive spreading has been put on a firm mathematical foundation in Refs. 11 and 12 using results from the theory of stochastic differential equations. Here we use a simple argument to derive an estimate of Δr_s.

The simplest deep-ocean scattering model is the apex approximation in which scattering is assumed to take place only at the upper turning point of the ray. After n such scattering events the scattering-induced spread in range is approximately

$$\Delta r_n = \frac{dR_\ell}{dI}(I_0)\sum_{i=1}^{n}(I_i - I_0), \tag{11}$$

where partial ray cycles are not accounted for, I_0 is the unperturbed ray action, and I_i is the ray action after the ith scattering event. Let δI_j denote the jump in action at the jth scattering event; then

$$I_i - I_0 = \sum_{j=1}^{i}\delta I_j. \tag{12}$$

To a good approximation δI_j can be assumed to satisfy $\langle \delta I_j\rangle=0$, $\langle \delta I_i \delta I_j\rangle=(\delta I)^2\delta_{ij}$ where the angular brackets denote ensemble average. Then it follows that $\langle (I_n-I_0)^2\rangle=(\delta I)^2 n \approx (\delta I)^2 r/R_\ell$, and that, for large n,

$$\langle (\Delta r)^2\rangle = \left(\frac{dR_\ell}{dI}(I_0)\right)^2(\delta I)^2\frac{1}{3}\left(\frac{r}{R_\ell}\right)^3. \tag{13}$$

But $\omega(I)=2\pi/R_\ell(I)$ and $(\delta I)^2/R_\ell$ can be replaced by a general action diffusivity D, where $\langle (I(r)-I_0)^2\rangle=Dr$, so

$$\Delta r_s = \frac{|\omega'(I_0)|}{\omega(I_0)}\sqrt{\frac{D}{3}}r^{3/2} \tag{14}$$

is the rms scattering-induced beam spread as a function of r. In spite of the strong assumptions that were made in the derivation of Eq. (14), ray-based numerical simulations reveal that this expression is a good approximation except for near-axial rays. [See Ref. 12 for a discussion of near-axial ray scattering and a correction to the simple action diffusion model on which Eq. (14) is based.]

In the limit of small r both Δr_d [Eq. (10)] and Δr_s [Eq. (14)] approach zero. But in this limit Δr must approach Δr_0, the beam width in close proximity to the source region. If a horizontal array is used to generate the beam, then Δr_0 is simply an appropriate measure of the length of the source array. If a vertical source array is used to generate the beam, then Δr_0 can be approximated as $\Delta z_0/|(dz/dr)_0|$, where Δz_0 is the array length and $(dz/dr)_0$ is the tangent of the angle made by the ray at the center of the beam at $r=0$. (If the initial beam angle is zero, $\Delta z_0/|(dz/dr)_0|$ can be evaluated at a short distance from the source array.) Because Δr_0, Δr_d, and Δr_s are independent it is natural to assume that these contributions to the total spread in range of a narrow beam combine approximately in quadrature (see also Refs. 14 and 30), so the total spread

$$\Delta r = \sqrt{(\Delta r_0)^2+(\Delta r_d)^2+(\Delta r_s)^2}. \tag{15}$$

While we do not have a rigorous justification for Eq. (15), we note that the variance of the sum of independent zero-mean Gaussian random variables is the sum of the individual variances.

Related to the spatial beam spreads considered here are the diffractive and scattering-induced measures of the "width of a ray" considered in Ref. 14. An important difference, however, is that the measures of the width of a ray considered earlier were constrained to satisfy a two-point boundary value problem, i.e., both endpoints were fixed. Like our Δr_d [Eq. (10)] and Δr_s [Eq. (14)] both contributions to the effective width of a ray were shown to be controlled by $\omega'(I_0)$ [or $\alpha(I_0)$ as described below].

C. Temporal beam spreads

We turn our attention now to ray-theoretical estimates of the time spreads of narrow transient beams. Unlike the spatial beam spreads described in Sec. II B, the temporal beam

spreads described here are not constrained by the condition that the final θ is fixed for all rays that comprise the beam; the time spread described here is the total spread in time of the energy contained in the beam, without regard to the final depth (or θ) of the energy at the range of interest. There are two reasons for defining spatial and temporal beam spreads differently. First, the definitions of both spatial and temporal beam spreads that we have adopted conform to the "obvious" way to define these quantities, as illustrated in the numerical results and figures described below. Second, the constrained (θ fixed) time spread estimate has been considered elsewhere (see Refs. 8 and 9) using arguments similar to those used here. The basic result is that at long range the scattering-induced contribution to the constrained time spread (the broadening of a branch of the time front) is second order in δI and is proportional to $\omega'(I)$. Importantly, however, that result is independent of the angular aperture of the source, and, as such, should not be thought of as a beam-related result. Finally, we note that throughout this section we shall neglect the travel time contributions in Eqs. (5) and (6) involving the $G-I\theta$ term. The neglected terms give small end-point corrections to ray travel times; in addition to being small, $G-I\theta$ oscillates about zero, with two zero-crossings per ray cycle, exhibiting no secular (in r) growth. Neglect of the end-point correction terms is consistent with the approximate treatment of partial ray cycles in Sec. II B.

Consider first the deterministic spread in time ΔT_d of a narrow beam of rays in the background $c(z)$ environment. We shall refer to ΔT_d as the dispersive spread of the beam. It follows from Eq. (6) that at a fixed range the dispersive time spread of the beam whose central ray has label I_0 is

$$\Delta T_d \approx \frac{\partial T}{\partial I}(I_0)\Delta I = I_0\omega'(I_0)r\Delta I. \tag{16}$$

Interestingly, the end-point terms in Eq. (6) give no first-order contribution to ΔT_d. (But the end-point terms are neglected below, so the comment made above about neglecting those terms is relevant.)

Next, we consider the scattering-induced contribution to the time spread. Consistent with the type of scattering model that was used above (but now allowing scattering to take place at any position along a ray), the total travel time of a scattered ray is

$$\begin{aligned} T_s &= \sum_i [I_i\omega(I_i) - H(I_i)]\Delta r_i \\ &\approx \sum_i [I_0\omega(I_0) - H(I_0) + I_0\omega'(I_0)(I_i - I_0)]\Delta r_i \\ &= [I_0\omega(I_0) - H(I_0)]r + I_0\omega'(I_0)\sum_i (I_i - I_0)\Delta r_i. \end{aligned} \tag{17}$$

The constraint $r=\Sigma_i r_i$ has been used, and, as noted earlier, the endpoint contributions to the travel time involving $G-I\theta$ have been neglected. The second term in the above expression, which we denote ΔT_s, is the scattering-induced contribution to the time spread of the beam. In the limit of small Δr_i, this term can be written as

$$\Delta T_s = I_0\omega'(I_0)\int_0^r (I(\xi) - I_0)d\xi. \tag{18}$$

Consistent with the arguments leading to Eq. (14), Eq. (18) reduces, in a rms sense, to

$$\Delta T_s = I_0|\omega'(I_0)|\sqrt{\frac{D}{3}}r^{3/2} \tag{19}$$

where, as above, $\langle (I(r)-I_0)^2\rangle = Dr$.

Interestingly, Δr_d [Eq. (10)], Δr_s [Eq. (14)], ΔT_d [Eq. (16)], and ΔT_s [Eq. (19)] are all proportional to $\omega'(I)$, a quantity that depends on the background sound-speed profile and is generally different for different rays. The *ray stability parameter*[5,7,9,14,31]

$$\alpha(I) = I\omega'(I)/\omega(I), \tag{20}$$

so that $I\omega'(I)=\alpha(I)\omega(I)$ and $\omega'(I)/\omega(I)=\alpha(I)/I$.

We turn our attention now to mode-based estimates of time spreads. Note that no mode-based estimates of the spatial spreads of beams were presented above. This choice was made in part to avoid the conceptual difficulty that the ray equivalent of a mode is a superposition of up- and down-going rays whose turning depths coincide with the mode turning depths. In spite of this correct conceptual picture, a narrow beam that is approximately centered on a single ray can be described as a superposition of modes. If that is done, the asymptotic quantization condition below leads to the identification of each mode with a ray cycle distance, which in turn leads to a modal analysis that looks much like the ray analysis presented above (see, e.g., Ref. 41). In contrast, the modal description of the time spreads of narrow beams involves an argument that differs from the ray-based argument. It will be shown that the seemingly different mode-based arguments presented below lead to the same results as the ray-based arguments presented above.

Consider first the deterministic dispersive spreading in time of a narrow beam which occupies a band ΔI of I-values. The beam is composed of modes covering a range of m-values over a band of σ, but to compute the dispersive spreading of the beam we need only to consider ΔI, and not the explicit m- and σ-dependences of the modes that make up the beam. It follows from Eq. (9) that the dispersive time spread of the beam is

$$\Delta T_d = \frac{dS_g}{dp_r}(I_0)\frac{dp_r}{dI}(I_0)r\Delta I, \tag{21}$$

where I_0 is action at the beam center. The quantity

$$\beta = -\frac{dS_g}{dp_r} \tag{22}$$

is referred to as the *waveguide invariant*.[16–21] [Note, however, that some authors (e.g., Ref. 16) define β as the reciprocal of this quantity.] In general $\beta=\beta(m;\sigma)$; consistent with our asymptotic analysis we shall sometimes write $\beta=\beta(I)$ with $I=I(m,\sigma)$ defined by Eq. (7). Also note that $dp_r/dI = -dH/dI = -\omega(I) = -2\pi/R_\ell$. Thus Eq. (21) can be written as

$$\Delta T_d = \beta(I_0)\omega(I_0) r \Delta I. \tag{23}$$

In Ref. 31 it is shown that asymptotically $\alpha(I)=\beta(I(m,\sigma))$; it is straightforward to derive this relationship using the results presented above. Thus, $\beta(I)\omega(I)=I\omega'(I)$, so Eq. (23) is seen to be identical to the ray-based estimate (16) of ΔT_d.

We consider now the mode-based estimate of the scattering-induced contribution to the total time spread. The delay time of the modal energy corresponding to a particular action history $\{I_1, I_2, \ldots\}$ is

$$T_s = \sum_i S_g(I_i)\Delta r_i, \tag{24}$$

where the total range $r=\Sigma_i \Delta r_i$. [Note that mode number history could be used in place of action history; we have chosen the latter primarily because that choice makes the connection to ray-based results more direct. Also, note that even for very simple initial ($r=0$) conditions, e.g., all energy in one I-value, the number of action histories that make up the total wavefield is generally very large.] It follows from Eq. (24) that

$$\begin{aligned} T_s &\approx \sum_i \left(S_g(I_0) + \frac{dS_g}{dp_r}(I_0)\frac{dp_r}{dI}(I_0)(I_i - I_0) \right) \Delta r_i \\ &= S_g(I_0) r + \beta(I_0)\omega(I_0) \sum_i (I_i - I_0)\Delta r_i. \end{aligned} \tag{25}$$

Taking the limit of small Δr_i, it is seen that the second term on the rhs, the scattering-induced time spread,

$$\Delta T_s = \beta(I_0)\omega(I_0) \int_0^r (I(\xi) - I_0) d\xi, \tag{26}$$

which, after noting that $\beta(I)\omega(I)=I\omega'(I)$, is seen to be identical to the ray-based estimate of ΔT_s [Eq. (18)], from which Eq. (19) follows.

The equivalence between the ray- and mode-based estimates of ΔT_d and ΔT_s should come as no surprise inasmuch as both the ray- and mode-based estimates describe the spreads in time of the same acoustic energy. Our demonstration of the equivalence of ray- and mode-based estimates of ΔT_d and ΔT_s was facilitated by our asymptotic treatment of the modal results and our use of action-angle variables to describe the ray results.

A third contribution to the total time spread of an acoustic beam is the reciprocal bandwidth,

$$\Delta T_{\mathrm{bw}} = (\Delta f)^{-1}. \tag{27}$$

This quantity is the minimum time spread of the beam and the time spread at $r=0$. Under most circumstances ΔT_{bw} is negligible compared to ΔT_d and ΔT_s. Independence of the three contributions to the total time spread ΔT suggests that the contributions should combine in quadrature, as was argued above for Δr, so

$$\Delta T = \sqrt{(\Delta T_{\mathrm{bw}})^2 + (\Delta T_d)^2 + (\Delta T_s)^2}. \tag{28}$$

Our simulations suggest that Eq. (28) is a very good approximation, but, as was the case for Eq. (15), we do not have a rigorous argument to support its validity. We note, however, that the assumption that ΔT_{bw} and ΔT_s are in quadrature is widely (see, e.g., Ref. 42) made, and that it was shown in Ref. 29 that ΔT_{bw} and ΔT_d are in quadrature for a narrowband Gaussian source spectrum. Also, we note that the arguments leading to Eqs. (23), (26), and (28) are very similar to those used in Ref. 30 to derive modal group time spread estimates.

III. NUMERICAL SIMULATIONS OF BEAM SPREADS

In this section, numerical simulations of the spatial and temporal spreads of narrow beams are presented and discussed in light of the theoretical estimates of beam spreads that were presented in the previous section. We begin with a discussion of some details relating to the numerical generation of beams and a description of the environments considered. Numerical results are presented in the final subsection.

A. Numerical generation of beams

The numerical simulations of underwater acoustic beams that are presented below were generated by solving the Thomson–Chapman form[43] of the parabolic wave equation (PE) using the split-step Fourier algorithm. Transient beams were constructed by Fourier synthesis. The source spectrum was assumed to have the shape of a Hanning window, whose total width was 128 Hz; note that the "effective" bandwidth Δf is somewhat less than this value.

To generate narrow directional beams, a Gaussian starting field (in both z and k_z) was employed (see, e.g., Ref. 44 for a discussion of this topic). If Δz is defined as the separation between the peak of the z distribution and the distance to the e^{-1} amplitude decay point, and similarly for Δk_z, then

$$\Delta z \Delta k_z = \sigma \Delta z \Delta p = 2\pi f \Delta z \Delta p = 2. \tag{29}$$

At 250 Hz the choice $\Delta z=80$ m corresponds to $\Delta p = 0.025$ s km^{-1}, or, approximately, $\Delta\phi=1.4°$. This value of Δz was used in our simulations. Note that the effective array length and beam width are approximately twice these values. Initial PE phases were chosen to be independent of depth, corresponding to horizontally directed (centered on $p_z=0$, as shown in Fig. 1) initial beams.

Finally, we address what appears to be a mismatch between our use of Thomson–Chapman PE simulations and theoretical results that are based on solutions to the Helmholtz equation. In fact, there need not be any mismatch. The reason is that all of the theoretical results that we seek to test are expressed using action-angle variables. The same expressions are applicable in the context of many parabolic approximations, including the Thomson–Chapman approximation, to the (one-way) Helmholtz equation; the only modification required is that the integrand in Eqs. (3) and (4) be replaced by the appropriate form of $p(H)$. For the Thomson–Chapman approximation $p(H)=\sqrt{c_0^{-2}-(H-c_0^{-1}+c^{-1}(z))^2}$. In the numerical results presented and described below, both the wavefields and the relevant action-angle-based quantities that appear in theoretical expressions, e.g., $\omega(I)$ and $\alpha(I)$, are computed in a way that is consistent with the Thomson–Chapman parabolic approximation.

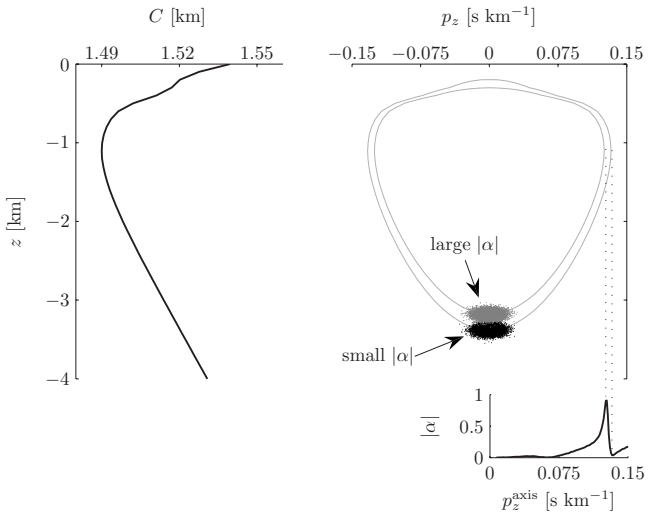


FIG. 1. Left panel: background sound-speed profile used in the simulations shown in the subsequent figures. Upper right panel: two surfaces of constant action in phase space corresponding to rays with large $|\alpha|$ and small $|\alpha|$. Note that the entire phase space is foliated by surfaces of constant action, so that each point in phase space is associated with a unique value of action. The dots shown in the lower portion of the figure have Gaussian distributions which approximate the distribution of energy in depth and vertical wavenumber for two PE starting fields extensively used in this study. Lower panel: $|\alpha|$ vs p_z^{axis}. The latter variable, like the action variable, can be thought of as a ray label in the background environment.

B. The environments considered

The choice of environments used in our simulations was motivated by the simple theoretical results presented above. Of particular interest is the prediction that the contributions to time spreads from deterministic dispersion and scattering are both proportional to $\alpha(I)$—or $\beta(m;\sigma)$—which is a function of the background sound-speed structure. For this reason we have chosen to use a background sound-speed profile with a nontrivial $\alpha(I)$ structure (Fig. 1). This environment consists of a perturbed canonical[45] profile,

$$c(z) = c_0 + \epsilon c_0(e^{\eta} - \eta - 1) + \delta c_0 e^{-(z-z_1)^2/2b^2}, \tag{30}$$

where $n=2(z-z_0)/B$, $c_0=1.49$ km s^{-1}, $z_0=-1.1$ km, $z_1=-0.35$ km, $B=1.0$ km, $b=0.1$ km, $\epsilon=5.7\times10^{-3}$, and $\delta=8\times10^{-3}$. Our use of narrow beams allows us to isolate bundles of rays (or groups of modes) with different values of α, so we need not consider different background sound-speed structures to test the predicted dependence of ΔR and ΔT on $\alpha(I)$.

Although we have chosen a background sound-speed profile with a nontrivial $\alpha(I)$ structure, we have intentionally avoided zero-crossings of $\alpha(I)$. Near such points, the simple first order expansions used above to obtain expressions (10), (11), (16), (17), (21), and (25) must be extended to include second-order terms. This topic will be discussed briefly in Sec. IV.

Equations (14), (19), and (26), which describe scattering-induced contributions to beam spreads, suggest that beam spreads should not show much sensitivity to details of $\delta c(z,r)$. This is because after several scattering events, the probability density function (PDF) of an ensemble of terms $\int_0^r (I(\xi)-I_0)\mathrm{d}\xi$ [or the discrete counterpart of this expression given in Eq. (11)] should not be sensitive to the details of the perturbation. Virovlyansky *et al.*[12] derived the relevant probability distribution functions for both off-axial (the problem on which we have focused) and near-axial (i.e., close to zero) I-values; in the former case the PDF is Gaussian. Those results are valid for any perturbation that leads to locally diffusive spreading of energy in I.

To test our expectation that both spatial and temporal beam spreads show little sensitivity to the details of $\delta c(z,r)$, we have performed simulations using two very different choices of δc. The first perturbation field, δc_{IW}, is a simulated internal-wave-induced sound-speed perturbation which correctly accounts for the inhomogeneity, anisotropy, and near power-law spectrum of mid-latitude internal waves in the deep-ocean.[46] Our δc_{IW} was computed using Eq. (19) of Ref. 47 with y and t set to zero, i.e., a frozen vertical slice of an internal-wave field was used. The range-averaged buoyancy frequency profile measured during the AET experiment[48,49] was used. The dimensionless parameter μ was set to 17.3 and the dimensionless strength E was varied; internal-wave strengths are specified as the fraction E/E_{GM} where $E_{\text{GM}}=6.3\times10^{-5}$ is the nominal Garrett–Munk strength parameter.[46] Horizontal wavenumber and vertical mode number cutoffs of 2π km^{-1} and 30, respectively, were used in our simulations.

In contrast to the oceanographic realism—in a statistical sense—of δc_{IW}, the second perturbation field that we have used, δc_{HIS}, has statistical properties that are not realistic oceanographically. This perturbation is consistent with a highly idealized model of turbulence that is homogeneous, isotropic, and has a Gaussian wavenumber spectrum. The Fourier transform of the Gaussian wavenumber spectrum is the autocorrelation function, which is also a Gaussian. The standard deviation of the Gaussian autocorrelation function, which we have set equal to 250 m, defines a unique length scale associated with δc_{HIS}. In contrast, the power-law δc_{IW} spectrum does not have a unique length scale. To satisfy $\delta c_{\text{HIS}}=0$ at the sea surface and bottom, a discrete set of vertical wavenumbers (each corresponding to a mode) is used. Then δc_{HIS}, like δc_{IW}, can be constructed by Fourier synthesis (over horizontal wavenumber) of a sum over a discrete set of vertical modes with random phases and amplitudes consistent with the specified energy spectrum, as described in Ref. 47; construction of δc_{HIS} is identical to construction of δc_{IW} except that, when constructing δc_{HIS}, no depth-stretching is applied and the Garrett–Munk internal-wave spectrum is replaced by a Gaussian spectrum.

C. Numerical results

Spatial spreads of beams are illustrated below using cw (fixed-frequency) wavefields with $f_0=250$ Hz. The spatial spreads of transient wavefields with this center frequency are the same as the cw beam spreads shown provided the transient wavefield is not rich in low frequency energy. (Interference effects are, of course, frequency dependent but these effects do not modify the bounds of the spatial domain of the insonified region. More importantly, frequency-dependent effects that we have not accounted for in our simple theoretical

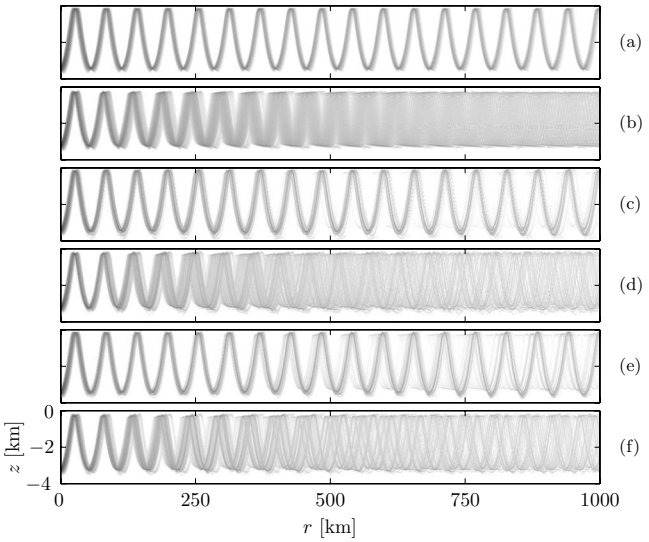


FIG. 2. Acoustic field intensity in the range-depth plane in cw ($f=250$ Hz) wavefields, showing the spatial spreading of narrow beams. The angular aperture of the beam at $r=0$ is the same in all cases. The same background sound-speed structure is used in all cases. Small $|\alpha|$ beams (corresponding to a source array midpoint at approximately 3350 m depth) are shown in panels (a), (c), and (e). Large $|\alpha|$ beams (corresponding to a source array midpoint at approximately 3150 m depth) are shown in panels (b), (d), and (f). In panels (a) and (b) no small-scale sound-speed perturbation is superimposed on the background; $\delta c=0$. In panels (c) and (d) an internal-wave-induced sound-speed perturbation is superimposed on the background; $\delta c=\delta c_{\mathrm{IW}}$. In panels (e) and (f), an idealized homogeneous isotropic single-scale sound-speed perturbation is superimposed on the background; $\delta c=\delta c_{\mathrm{HIS}}$.

beam spread estimates are important at frequencies sufficiently low that the influence of the ocean boundaries is felt, i.e., when a ray turning depth is within a few wavelengths of the ocean surface or bottom.) Temporal spreads of transient beams are illustrated below using broadband wavefields with $f_0=250$ Hz and $\Delta f\approx 100$ Hz (recall that a Hanning spectral window was used whose total width was 128 Hz). In the time-depth plane at $r=0$ the transient source appears (this is not shown in any figure) as a narrow Gaussian distribution in depth, centered at the source depth, with an approximately Gaussian distribution in time whose standard deviation is approximately $10\ \mathrm{ms}=(100\ \mathrm{Hz})^{-1}$. All of the wavefields shown were generated using the same vertical source array with a Gaussian distribution of energy in depth whose half-width is approximately 80 m; this corresponds [from Eq. (29) with $\Delta p\approx 0.025\ \mathrm{s\ km}^{-1}$] to a beam half-width of roughly 1.4°. In a fixed background sound-speed profile the corresponding initial spread in action ΔI depends weakly on the depth at which the beam is centered. Most of the numerical simulations presented use beams that are centered at $r=0$ on either the local maximum or the local minimum in $|\alpha(I)|$ (recall Fig. 1). These are referred to below as the large $|\alpha|$ and the small $|\alpha|$ beams, respectively. To an excellent approximation ΔI is the same for both of these beams.

We consider first the spatial spreading of cw beams. Figure 2 shows the spatial spreading of beams in the range-depth plane. The two upper panels show beam spreading in the background range-independent environment, i.e., under conditions in which $\delta c=0$ so $\Delta r_s=0$. Under such conditions Eqs. (10) and (15) predict that, except close to the source, Δr increases linearly in r at a rate proportional to $|\omega'(I_0)|$ (or $|\alpha(I_0)|$). Figure 2 is consistent, qualitatively at least, with this prediction inasmuch as the beam with large $|\alpha|$ is seen to spread much more rapidly that the beam with small $|\alpha|$. [Because ΔI and $\omega(I_0)$ are, to an excellent approximation, the same in both cases, the different rates of beam spreading is due almost entirely to the difference in $\omega'(I_0)$.] The middle two panels of Fig. 2 are the same as the upper two panels except that an internal-wave-induced sound-speed perturbation was superimposed on the background to compute the wavefields shown. Under these conditions Δr is described by Eq. (15); because both $|\Delta r_d|$ and Δr_s are proportional to $|\alpha(I_0)|$, Δr is predicted to be nearly (neglecting the small term Δr_0) proportional to $|\alpha(I_0)|$. Again, this is consistent with the wavefields shown in Fig. 2. The lower two panels are the same as the middle two panels except that a homogeneous isotropic sound-speed perturbation was used in place of the internal-wave-induced perturbation. The same comments that were made about the middle two panels also apply to the lower two panels. In a manner that is largely independent of details of the perturbation, beams with large $|\alpha|$ are seen to spread more rapidly than beams with small $|\alpha|$. Weak sensitivity to the perturbation enters via the diffusivity D. Note, however, that in spite of the very different statistical properties of the two types of perturbation considered, both lead to similar estimates of D and hence also similar behavior with regard to beam spreading.

Unfortunately, for the large $|\alpha|$ beams, beam spreading in Fig. 2 is seen to be sufficiently rapid that the self-intersection of beams in the range-depth plane after only a few hundred kilometers leads to difficulties in obtaining quantitative estimates of beam widths. This difficulty could be largely overcome by producing narrower beams, i.e., by decreasing ΔI. This could be accomplished in our simulations by either increasing f_0 or increasing the length of the source array (or both). Increasing the length of a source array beyond the length that we have assumed would be extremely difficult to achieve in an experimental setting. Increasing the source center frequency could easily be achieved, but because attenuation increases rapidly with increasing frequency, the maximum experimentally accessible range would then be less than the maximum range plotted in Fig. 2. In short, we could produce numerically simulated wavefields to quantitatively test our theoretical prediction of the spatial spreading of narrow beams at long range, but the simulations would be somewhat unphysical. In contrast, the temporal spreads of beams presented below, with $f_0=250$ Hz and $\Delta f\approx 100$ Hz, do not have this drawback. For this reason we shall focus on the temporal spread of transient directionally narrow beams throughout the remainder of this section.

The temporal spreading of transient beams is illustrated in Fig. 3; wavefields in the (z,T) plane are shown at selected values of r. Four examples of distributions of acoustic energy of this type are shown. In this and all subsequent plots, the range at which wavefields are plotted—and the range at which time spreads are estimated—is taken to be an integer multiple of the unperturbed complete cycle distance of the ray at the center of the $r=0$ distribution of energy in phase space (Fig. 1). This choice was made because energy tends to

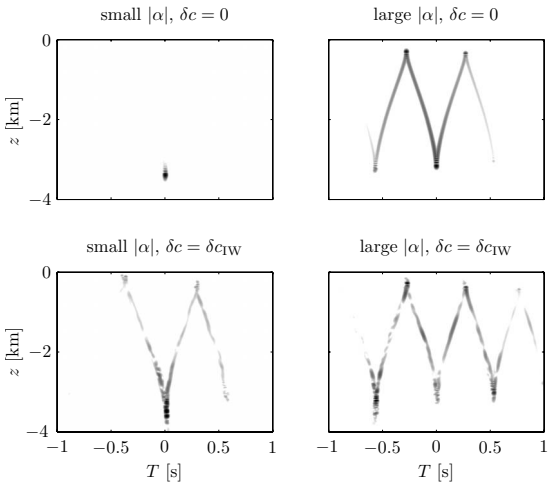


FIG. 3. Distribution of acoustic energy in time and depth for transient narrow beams with small $|\alpha|$ (left) and large $|\alpha|$ (right) after 19 complete unperturbed cycles of the ray at the center of the beam (cf. Fig. 1). With this constraint the final range is approximately 1070 km in the left panels and 1015 km in the right panels. Time is plotted relative to the travel time of the unperturbed central ray. In the upper panels the perturbation field $\delta c=0$ while in the lower panels $\delta c=\delta c_{\mathrm{IW}}$ with $E/E_{\mathrm{GM}}=1$. The dynamic range in all four panels is 35 dB.

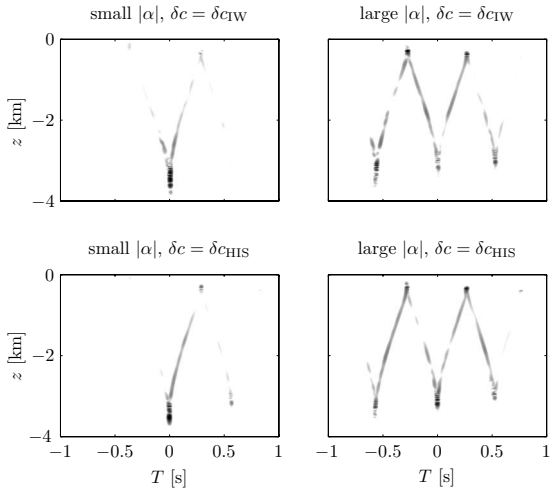


FIG. 5. Distribution of acoustic energy in time and depth for narrow transient beams illustrating strong sensitivity to $|\alpha|$ and weak sensitivity to details of δc. Left panels: small $|\alpha|$. Right panels: large $|\alpha|$. Upper panels: $\delta c=\delta c_{\mathrm{IW}}$. Lower panels: $\delta c=\delta c_{\mathrm{HIS}}$.

focus at the upper and lower turning depth. This choice reduces partial cycle deterministic propagation effects that we have treated approximately; such partial cycle effects are most evident at short range where time spreads are small.

Consider first the $\delta c=0$ wavefields shown in the top panels of Fig. 3. The predicted time spreads are described by Eq. (28) with $\Delta T_s=0$. Consistent with the theoretical predictions, deterministic time spreads are seen to be large (small) when $|\alpha|$—or $|\beta|$—is large (small). The same trend is seen in the bottom panels of Fig. 3, corresponding to wavefields in an environment including an internal-wave-induced perturbation $\delta c=\delta c_{\mathrm{IW}}$. This trend is not surprising because both the deterministic dispersive contribution (16) or (21) and the scattering-induced contribution (19) or (26) (again note the equivalence of the ray- and mode-based estimates) to the total time spread are seen to be controlled by α or β.

Three of the remaining figures, including Fig. 4, show numerical estimates of time spreads. These were computed according to

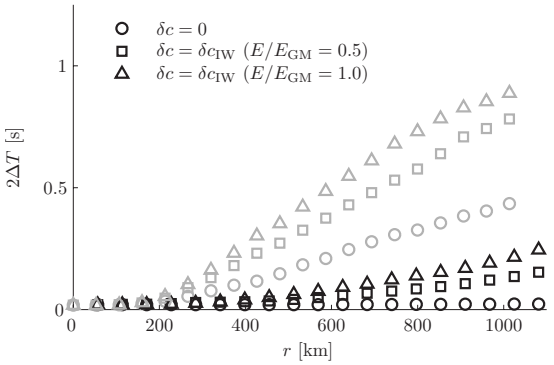


FIG. 4. Six sets of points showing the evolution of time spreads as a function of range for beams with small $|\alpha|$ (black symbols) and large $|\alpha|$ (gray symbols), and three different values of the strength of δc_{IW}.

$$\Delta T = \sqrt{\frac{\int t^2 P dt}{\int P dt}}, \tag{31}$$

where $P(t;r)$ is the wavefield intensity integrated over depth, and $t=0$ is taken to be the arrival time of the unperturbed ray at the center of the beam. Note that the quantity plotted in Figs. 4, 6, and 7 is $2\Delta T$.

Figure 4 shows two sets of three ΔT vs r curves, one set for small $|\alpha|$ and one set for large $|\alpha|$ (recall Fig. 1). Within each set, the three curves correspond to three values of the strength of the internal-wave-induced perturbation field, $E/E_{\mathrm{GM}}=0,0.5,1$. Consider first the small $|\alpha|$ curves. When $\delta c=0$, ΔT is seen to remain very small for long r, consistent with Eqs. (16) and (23). When $\delta c\neq 0$, ΔT at fixed r is seen to increase with increasing perturbation strength, consistent with Eqs. (18), (19), and (26), but with small $|\alpha|$ ΔT remains relatively small. Consider now the large $|\alpha|$ curves. Note that, even when $\delta c=0$, ΔT is relatively large owing to the α-dependence of the dispersive contribution (16) and (23) to ΔT. Again, when $\delta c\neq 0$, ΔT at fixed r is seen to increase with increasing perturbation strength, consistent with Eqs. (18), (19), and (26).

We turn our attention now to the structure of the sound-speed perturbation δc. Figure 5 shows the distribution of acoustic energy in depth and time for beams with large and small $|\alpha|$ after 19 complete beam cycles (cycles of the unperturbed ray at the center of the beam) in environments with two very different sound-speed perturbation fields δc_{IW} and δc_{HIS} as discussed above. In spite of the significant differences between δc_{IW} and δc_{HIS}, Fig. 5 reveals only remarkably small differences in the corresponding wavefields—and one expects that most of these differences would be eliminated by computing an ensemble-averaged wavefield. In contrast, Fig. 5 shows a strong dependence on the background $c(z)$ via α.

Figures 3–5 show the time spreads only for beams centered on the local maximum or minimum of $|\alpha|$, as illustrated in Fig. 1. This restriction is relaxed in Fig. 6 where the initial

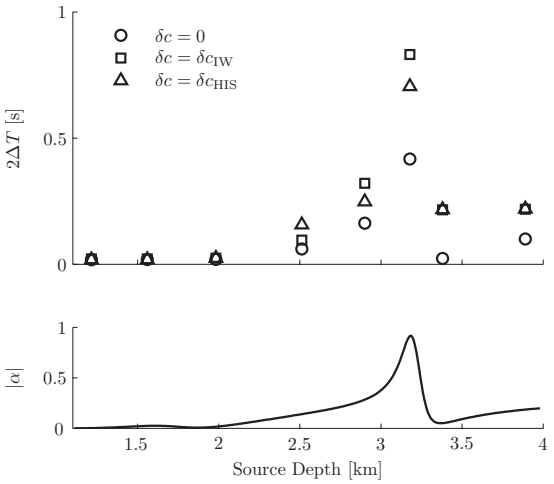


FIG. 6. Upper panel: time spreads of transient narrow beams after 19 complete unperturbed cycles of the ray at the center of the beam as a function of beam initial depth. The final ranges of the points plotted, from left to right, are approximately 785, 805, 820, 860, 915, 1015, 1070, and 1100 km. Lower panel: the solid curve shows the corresponding variation of $|\alpha|$ (cf. Fig. 1). Note the qualitative agreement between the time spread estimates and $|\alpha|$ variations.

$(r=0)$ beam depth is varied, providing a stronger test of the predicted α-dependence in Eq. (16) or (19) [or the β-dependence in Eq. (23) or (26)]. Three sets of computed time spreads are plotted—for $\delta c=0$, $\delta c=\delta c_{\mathrm{IW}}$ and $\delta c=\delta c_{\mathrm{HIS}}$—along with the corresponding α-dependence. (The latter are α-values for rays with zero launch angles at the depths of the centers of the initial beams.) Consistent with the results presented earlier, both theoretical and numerical, Fig. 6 shows that beam time spreads, both with $\delta c=0$ and $\delta c\neq 0$, are largely controlled by α (or β), and that beam time spreads show remarkably little sensitivity to the structure of δc.

The comparison between simulations and theoretical predictions shown in Figs. 2–6 is qualitative, emphasizing the importance of $|\alpha|$. A quantitative comparison is shown in Fig. 7. In that figure the range evolution of simulation-based estimates of ΔT is compared to the prediction based on Eqs. (28), (27), and (16) or (23) and (19) or (26) for large and small $|\alpha(I_0)|$, and for both $\delta c=0$ and $\delta c=\delta c_{\mathrm{IW}}$. (The same simulation-based estimates of ΔT are shown in Fig. 4, but the

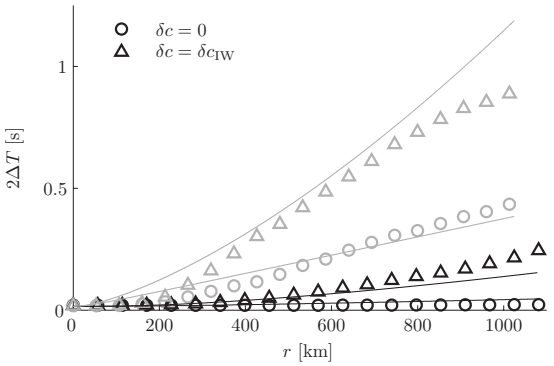


FIG. 7. Theoretical predictions (solid lines) and simulated-wavefield-based estimates (circles and triangles) of the temporal spreads of directionally narrow transient beams as a function of range. Beams with small and large $|\alpha|$ are shown in black and gray, respectively.

points corresponding to the weak δc_{IW} field are omitted in Fig. 7.) When evaluating Eq. (28) some care must be exercised to normalize all of the terms in this expression in a self-consistent fashion. Consistent with Eq. (31) and assumptions made earlier, ΔT should be interpreted as the standard deviation of an approximately Gaussian distribution of acoustic intensity as a function of time. This choice fixes $\Delta f=100$ Hz, as noted earlier, and $\Delta I=0.0025$ s (for both beams considered). The diffusivity for each beam was estimated from the relationship $\langle (I(r)-I_0)^2\rangle = D_T$ using an ensemble of scattered rays. The estimates obtained were $D=1.6\times 10^{-7}\ \mathrm{s^2\,km^{-1}}$ for the large $|\alpha|$ beam and $D=2.0\times 10^{-7}\ \mathrm{s^2\,km^{-1}}$ for the small $|\alpha|$ beam. These values are close to the values reported in Refs. 12, 14, and 30. A final remark concerning evaluation of ΔT_d [Eqs. (16) or (23)] and ΔT_s [Eqs. (19) or (26)] is that these quantities depend on the choice of the reference action I_0. There is some ambiguity associated with this choice because the Taylor series used to derive those results are valid (approximately) for any I_0 within the beam's initial ΔI-window. We have evaluated $I_0\omega'(I_0)=\omega(I_0)\alpha(I_0)$ in Eqs. (16) and (19) using $\omega(I_0)=0.11\ \mathrm{km^{-1}}$ for both beams, and $|\alpha(I_0)|=0.075$ and 0.70 for the small- and large-$|\alpha|$ beams, respectively. Both $|\alpha|$ values are representative of the respective beams. Note also that $\omega(I)$ is nearly constant within each beam, so the choice of $\omega(I_0)$ is subjected to much less uncertainty than the choice of $\alpha(I_0)$.

Agreement between theory and simulations in Fig. 7 is seen to be generally good. One important caveat is that in the presence of scattering $(\delta c\neq 0)$, the simulations of $\Delta T(r)$ at large r are seen to deviate from the theoretical predictions in an apparently systematic way. The reason for this behavior is that scattered rays that have an initial α-value close to a local maximum of $|\alpha(I)|$ will, on average, be scattered onto I-surfaces with lower $|\alpha|$-values, and, conversely, rays that have an initial α-value close to a local minimum of $|\alpha(I)|$ will, on average, be scattered onto I-surfaces with higher $|\alpha|$-values. In other words, in the presence of scattering the $|\alpha|$ in Eqs. (19) and (26) should be replaced by a locally averaged value where the amount of averaging increases with increasing range. Also, at ranges of approximately 500 km or less, simulation-based estimates of ΔT are systematically less than the theoretical predictions, especially for large $|\alpha|$. This is likely due, in part, to a significant clustering of energy near the caustic adjacent to the beam's lower turning depth at the ranges at which points are plotted. In addition to these trends, note that expressions (16) and (21) grow like r, while expressions (19) and (26) grow like $r^{3/2}$. (Recall also that the scattering of near-axial rays, or small m modes, requires special care—see Ref. 12.)

IV. SUMMARY AND DISCUSSION

In this paper it was shown, using PE simulations of directionally narrow acoustic beams in deep-ocean environments, that, consistent with both ray- and mode-based theoretical predictions, both the spatial and temporal spreads of such beams are largely controlled by the background sound-speed structure through the ray stability parameter $\alpha(I)$—or

its mode equivalent $\beta(m;\sigma)$. This trend was shown to hold both with and without a small-scale perturbation (due, for example, to internal waves) superimposed on the background; details of the sound-speed perturbation were shown to have only a minor influence on controlling beam spreads.

The above conclusion has important consequences. Directionally broad sound fields generated by a compact source can be decomposed into many directionally narrow beams. Because the dynamics of each narrow beam is governed by α (or β), so too is the total wavefield, which is simply the phase-coherent superposition of the constituent narrow beams. Our demonstration that time spreads of broadband signals are largely controlled by α implies that phase fluctuations of narrow-band signals are also largely controlled by α. Also, geometric intensities are inversely proportional to spatial beam spreads; consistent with the results presented here, geometric intensities, in both unperturbed ($\delta c=0$) and perturbed ($\delta c\neq 0$) environments, have been shown[7,31] to be largely controlled by α. It follows from these observations that wavefield structure and stability are largely controlled by α. [A complication is that in directionally broad wavefields, rays (modes) with many α (β) values contribute to the wavefield, which may obscure the role of α (β). This problem may be severe in fixed-frequency wavefields where arrivals—whether they are interpreted as rays or modes—cannot be separated in time.] A consequence of these simple observations is that in environments consisting of a range-independent (or slowly-varying in range) background on which a weak highly structured perturbation is superimposed, the background sound-speed structure, via α (or β), plays a critical role in mediating the transfer of environmental variability to wavefield variability.

We turn our attention now to the four lines of research mentioned in Sec. I that are related to the present study. We shall begin with Tappert's[32] prediction of "exploding beams." Prior to discussing the connection between our work and this topic, it is necessary to explain the rationale for the exploding beam prediction. The beams that are the object of the present study (and Tappert's[32] earlier study) correspond at $r=0$ to an approximately elliptical blob of acoustic energy in phase space (recall Fig. 1). We assume that $\delta c(z,r)$ is small but nonzero and that ray trajectories are predominantly chaotic. The evolution in range of the phase space blob is analogous to the evolution in time of a patch of dye in a turbulent $2d$ incompressible fluid flow. To more fully appreciate this statement, note that it follows from Eq. (1) that the flow in phase space is incompressible, $\partial(dz/dr)/\partial z+\partial(dp/dr)/\partial p=0$ (analogous to the incompressibility condition, $\partial u/\partial x+\partial v/\partial y=0$, for a $2d$ incompressible fluid). With this fluid mechanical analogy in mind, the following comments about evolution of the beam should be clear. As the beam propagates, the corresponding phase space blob undergoes counterclockwise motion in phase space, staying close to the I-surface on which it was initially centered. Meanwhile, provided the frequency is sufficiently high that diffractive smearing effects are initially negligible, evolution of the blob is constrained as follows: (1) the perimeter of the blob grows exponentially (because ray motion is predominantly chaotic), (2) the area enclosed remains constant (a consequence of the incompressibility of the flow in phase space), and (3) the I-domain that encloses essentially all of the blob grows approximately diffusively [an approximate result that we have used to derive Eqs. (14), (19), and (26)]. As the blob evolves, it becomes increasingly convoluted, acquiring structure on increasingly small scales. At sufficiently long-range diffractive effects will smear out the finescale structure of the blob. Tappert's[32] exploding beam prediction is linked to the initial exponential growth of the perimeter of the blob.

Tappert[32] did not perform any numerical simulations to test his prediction. Subsequently, however, Morozov and Colosi[33] did perform such simulations. They did not see any evidence of explosive (exponential) spreading of beams, but they did show that their computed wavefields had a positive Kolmogorov–Sinai entropy, which suggests that the wavefields evolve chaotically. Consistent with the work reported in Ref. 33, we have not in the present study observed exponential spreading of beams (although the motivation for our work was not to search for such behavior). We regard the reasoning behind Tappert's[32] prediction as correct, so the fact that neither we nor Morozov and Colosi[33] observed this behavior requires some explanation. Of critical importance is the fact that the wavefields presented here and in Ref. 33 are fields in z that evolve as a function of r (the presence or absence of temporal structure is not important in this context). At each r the observed z-structure is the projection of some structure in phase space (p,z) onto z. As the perimeter of the aforementioned phase space blob grows exponentially in r, the projection of this structure onto z will not, in general, grow exponentially, except possibly as an initial (short r) transient; typically, power-law growth in the z-spread of the blob is expected. In order to observe the exponential growth of the perimeter of the phase space blob, phase space distributions (e.g., Wigner or Husimi distributions) must be computed at many ranges. Also, to observe several e-folding intervals, the wave frequency must be high enough to appropriately postpone the large r regime in which diffractive effects dominate. We suspect that in typical deep-ocean environments, this dictates that the wave frequency be much higher than 250 Hz. These considerations suggest that only with great care can the predicted explosive behavior be observed. With the above discussion as background, we see no contradiction between our work (or that of Morozov and Colosi[33]) and Tappert's[32] prediction of exploding beams. A secondary point relating to explosive beam growth is that the work presented in Refs. 32 and 33 does not account for the important role played by α; consistent—qualitatively, at least—with the results presented here, numerical simulations presented in Ref. 7 show that average Lyapunov exponents (inverse e-folding ranges) are approximately proportional to $|\alpha(I)|$. We note also that a later paper[50] by the authors of Ref. 33 included numerical simulations of beams, but the focus of the latter work was mode coupling, rather than beam dynamics.

There is a close connection between the results presented here and the earlier work[34–37] on weakly divergent beams, which satisfy the condition $\alpha(I_0)=0$. Clearly, those beams are a special limiting case of the beams considered

here. [In Refs. 34–37 the condition $\alpha(I)=0$ is expressed as $dR_\ell(p_r)/dp_r=0$.] The focus of the earlier work on weakly divergent beams was on the spatial spreading of beams—including beams embedded in directionally broad distributions of acoustic energy such as fields produced by a point source—and associated implications for beam intensities. We have not discussed in this paper the $\alpha(I_0)\to 0$ limit, but it is straightforward to derive the second-order corrections to the deterministic spatial and temporal beam spread estimates, Eqs. (10) and (16), that are required to describe this limit. The corrected expressions are

$$\Delta r_d = \frac{r}{\omega(I_0)}\left[-\omega'(I_0)\Delta I + \left(\frac{(\omega'(I_0))^2}{\omega(I_0)} - \frac{\omega''(I_0)}{2}\right)(\Delta I)^2\right] \tag{32}$$

and

$$\Delta T_d = r\left[I_0\omega'(I_0)\Delta I + \frac{1}{2}(\omega'(I_0) + I_0\omega''(I_0))(\Delta I)^2\right], \tag{33}$$

where $\omega''(I_0)$ is the second derivative of $\omega(I)$ evaluated at I_0. Note that, like Eqs. (10) and (16), these expressions treat partial ray loops approximately. As noted earlier, the neglected partial loop corrections are small and they oscillate about zero, exhibiting no secular (in r) growth. Note that in the limit $\omega'(I_0)=0$ [so $\alpha(I_0)=0$] Δr_d and ΔT_d (1) are nonzero, (2) grow linearly in r, like Eqs. (10), (16), and (3) are second order in the small parameter ΔI. Also, recent work[51] reveals that rays that satisfy $\omega'(I)=0$ often correspond to robust barriers in phase space; this leads, for example, to the expectation that the region between two neighboring zero-crossings of $\omega'(I)=0$ might serve as an effective trap of acoustic energy. Perhaps the most interesting insight of the present work relative to the earlier work on weakly divergent beams is that those beams are weakly divergent temporally as well as spatially.

Finally, we reiterate that the results presented here are consistent with a rapidly growing body of work[1–31] that shows that numerous wavefield properties and/or diagnostics are controlled by $\alpha(I)$—or its modal counterpart $\beta(m;\sigma)$. [Note, however, that some of those references used a notation different than that used here, so that the α (or β) dependence is sometimes not immediately obvious.]

ACKNOWLEDGMENTS

We thank Ilya Udovydchenkov, Irina Rypina, and John Colosi for the benefit of many useful discussions relating to this work. This research was supported by the National Science Foundation under Grant No. CMG0825547 and Code No. 321OA of the Office of Naval Research.

[1] A. L. Virovlyansky, "Travel times of acoustic pulses in the ocean," Sov. Phys. Acoust. **31**, 399–401 (1985).
[2] A. L. Virovlyansky, "Temporal structure of a pulse signal in an underwater sound channel," Sov. Phys. Acoust. **31**, 480–481 (1985).
[3] A. L. Virovlyansky, "On general properties of ray arrival sequences in oceanic acoustic waveguides," J. Acoust. Soc. Am. **97**, 3180–3183 (1995).
[4] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, "Theory and application of ray chaos to underwater acoustics," Phys. Rev. E **64**, 036221 (2001).
[5] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long-range deep-ocean sound propagation," J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[6] F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242 (2003).
[7] F. J. Beron-Vera and M. G. Brown, "Ray stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **144**, 123–130 (2003).
[8] A. L. Virovlyansky, "Ray travel times at long range in acoustic waveguides," J. Acoust. Soc. Am. **113**, 2523–2532 (2003).
[9] F. J. Beron-Vera and M. G. Brown, "Travel time stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **115**, 1068–1077 (2004).
[10] A. L. Virovlyansky and G. M. Zaslavsky, "Ray and wave chaos in problems of sound propagation in the ocean," Acoust. Phys. **53**, 282–297 (2007).
[11] A. L. Virovlyansky, "Statistical description of ray chaos in an underwater acoustic waveguide," Acoust. Phys. **51**, 71–80 (2005).
[12] A. L. Virovlyansky, A. Yu. Kazarova, and L. Ya. Lyubavin, "Statistical description of chaotic rays in a deep water acoustic waveguide," J. Acoust. Soc. Am. **121**, 2542–2552 (2007).
[13] A. L. Virovlyansky, *Ray Theory of Long-Range Sound Propagation in the Ocean* (Institute of Applied Physics, Nizhny Novgorod, 2006), in Russian.
[14] I. I. Rypina and M. G. Brown, "On the width of a ray," J. Acoust. Soc. Am. **122**, 1440–1448 (2007).
[15] A. L. Virovlyansky, A. Yu. Kazarova, L. Ya. Lyubavin, and D. F. Nefedova, "Ray description of the field of a distributed source in a waveguide," Acoust. Phys. **54**, 654–663 (2008).
[16] S. D. Chuprov, "Interference structure of an acoustic field in a layered waveguide," in *Acoustics of the Ocean*, edited by L. M. Brekhovskikh, and I. B. Andreevoi (Nauka, Moscow, 1982).
[17] G. A. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**, 33–35 (1993).
[18] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, New York, 1991).
[19] G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[20] G. L. D'Spain, G. L. Rovner, P. Gerstoft, W. A. Kuperman, and W. S. Hodgkiss, "Determination of the waveguide invariant in general environments," U.S. Navy J. Underwater Acoustics **51**, 123–142 (2002).
[21] P. Gerstoft, G. L. D'Spain, W. A. Kuperman, G. L. Rovner, and W. S. Hodgkiss, "Calculating the waveguide invariant β by ray-theoretic approaches," Marine Physical Laboratory Report No. TM-468, University of California San Diego, La Jolla, CA, 2002.
[22] S. Kim, W. A. Kuperman, W. S. Hodgkiss, M. C. Song, G. F. Edelmann, and T. Akal, "Robust time reversal focusing in the ocean," J. Acoust. Soc. Am. **114**, 145–157 (2003).
[23] 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 (1999).
[24] 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).
[25] V. M. Kuz'kin, "The effect of variability of ocean stratification on a sound field interference structure," Acoust. Phys. **41**, 300–301 (1995).
[26] V. M. Kuz'kin, A. V. Ogurtsov, and V. G. Petnikov, "The effect of hydrodynamic variability on frequency shifts of the interference pattern of a sound field in shallow water sea," Acoust. Phys. **44**, 77–82 (1998).
[27] V. M. Kuz'kin, "Frequency shifts of the interference pattern of a sound field in shallow water," Acoust. Phys. **45**, 224–229 (1998).
[28] W. A. Kuperman, S. Kim, G. F. Edelman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Group and phase speed analysis for predicting and mitigating the effects of fluctuations," in *Impact of Litoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Lerici, 2002), pp. 279–286.
[29] M. G. Brown, I. Viechnicki, and F. D. Tappert, "On the measurement of modal group time delays in deep-ocean," J. Acoust. Soc. Am. **100**, 2093–2102 (1995).
[30] I. A. Udovydchenkov and M. G. Brown, "Modal group time spreads in weakly range-dependent deep ocean environments," J. Acoust. Soc. Am. **123**, 41–50 (2008).
[31] M. G. Brown, F. J. Beron-Vera, I. Rypina, and I. A. Udovydchenkov, "Rays, modes, wavefield structure and wavefield stability," J. Acoust. Soc. Am. **117**, 1607–1610 (2005).

[32]F. D. Tappert, "Theory of explosive beam spreading due to ray chaos," J. Acoust. Soc. Am. **114**, 2775–2781 (2003).
[33]A. K. Morozov and J. A. Colosi, "Entropy and scintillation analysis of acoustical beam propagation through ocean internal waves," J. Acoust. Soc. Am. **117**, 1611–1623 (2005).
[34]L. M. Brekhovskikh, V. V. Goncharov, and V. M. Kurtepov, "Weakly divergent bundles of sound rays in the Arctic," Atmos. Oceanic Phys. **31**, 441–446 (1995).
[35]Y. V. Petukhov, "A sound beam with minimal wavefront divergence in a stratified waveguide," Acoust. Phys. **40**, 97–105 (1994).
[36]V. V. Goncharov and V. M. Kurtepov, "Formation and propagation of weakly diverging bundles of rays in a horizontally inhomogeneous ocean," Acoust. Phys. **40**, 685–692 (1994).
[37]Y. V. Petukhov, "Slowly-diverging acoustic beams in smoothly inhomogeneous ocean waveguides," Acoust. Phys. **43**, 196–201 (1997).
[38]E. Svensson, "Gaussian beam summation in shallow waveguides," Wave Motion **45**, 445–456 (2008).
[39]M. P. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**, 1349–1359 (1987).
[40]W. Munk and C. Wunsch, "Ocean acoustic tomography: Rays and modes," Rev. Geophys. Space Phys. **21**, 1–37 (1983).
[41]A. L. Virovlyansky, A. Yu. Kazarova, and L. Ya. Lyubavin, "Ray-based description of normal mode amplitudes in a range-dependent waveguide," Wave Motion **42**, 317–334 (2005).
[42]S. M. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University, Cambridge, 1979).
[43]D. J. Thomson and N. R. Chapman, "A wide-angle split-step algorithm for the parabolic equation," J. Acoust. Soc. Am. **74**, 1848–1854 (1983).
[44]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 1994).
[45]W. H. Munk, "Sound Channel in an Exponentially Stratified Ocean with Application to SOFAR," J. Acoust. Soc. Am. **55**, 220–226 (1974).
[46]W. H. Munk, "Internal Waves and Small Scale Processes," in *Evolution of Physical Oceanography*, edited by B. Warren and C. Wunsch (MIT, Cambridge, 1981), pp. 264–291.
[47]J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[48]P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, J. A. Colosi, B. M. Howe, J. A. Mercer, A. B. Baggeroer, and K. Metzger, "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the Eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[49]J. A. Colosi, S. M. Flatté, B. D. Cornuelle, M. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "Comparison of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the Eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[50]A. K. Morozov and J. A. Colosi, "Stochastic differential equation analysis for sound scattering by random internal waves in the ocean," Acoust. Phys. **53**, 335–347 (2007).
[51]I. I. Rypina, M. G. Brown, F. J. Beron-Vera, H. Koçak, M. J. Olascoaga, and I. A. Udovydchenkov, "Robust transport barriers resulting from strong KAM stability," Phys. Rev. Lett. **98**, 104102 (2007).

Determination of power-law attenuation coefficient and dispersion spectra in multi-wall carbon nanotube composites using Kramers–Kronig relations

Joel Mobley[a)]
Department of Physics and Astronomy and Jamie Whitten National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

Richard A. Mack
Jamie Whitten National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

Joseph R. Gladden
Department of Physics and Astronomy, University of Mississippi, 108 Lewis Hall, University, Mississippi 38677

P. Raju Mantena
Department of Mechanical Engineering, University of Mississippi, 201D Carrier Hall, University, Mississippi 38677

(Received 14 October 2008; revised 2 January 2009; accepted 2 April 2009)

Using a broadband through-transmission technique, the attenuation coefficient and phase velocity spectra have been measured for a set of multi-wall carbon nanotube (MWCNT)-nylon composites (from pure nylon to 20% MWCNT by weight) in the ultrasonic frequency band from 4 to 14 MHz. The samples were found to be effectively homogeneous on spatial scales from the low end of ultrasonic wavelengths investigated and up (>0.2 mm). Using Kramers–Kronig relations, the attenuation and dispersion data were found to be consistent with a power-law attenuation model with a range of exponents from $y=1.12$ to $y=1.19$ over the measurement bandwidth. The attenuation coefficients of the respective samples are found to decrease with increasing MWCNT content and a similar trend holds also for the dispersion. In contrast, the mean phase velocities for the samples rise with increasing MWCNT content indicating an increase in the mechanical moduli.
 [DOI: 10.1121/1.3125323]

PACS number(s): 43.35.Cg, 43.20.Jr, 43.35.Yb, 43.35.Zc [RLW] Pages: 92–97

I. INTRODUCTION

Since the discovery of carbon nanotubes (CNTs) by Iijima in 1991,[1] significant efforts have been made to incorporate these nanoparticles with conventional materials in order to improve the mechanical strength and stiffness or other physical properties (e.g., electrical and thermal conductance) of the resulting composite.[2–4] CNTs include both single- and multi-walled (MWCNT) structures, with the former having typical outside diameter (OD) of 1–2 nm while the latter an OD of 8–12 nm. Their lengths range from the typical 10 μm to as much as 100 μm with very high aspect ratios (length-to-diameter) of order 1000:1. CNTs have about 50 times the tensile strength of stainless steel (100 GPa vs 2 GPa) and five times the thermal conductivity of copper. Incorporating nano-scale particles into a matrix to construct a macro-scale composite can potentially offer improved performance over composites with larger inclusions (e.g., conventional carbon fibers) for several reasons, including the increased effective surface area of contact between the nanoparticles and the matrix, and higher crystallinity. CNT composites aim to capitalize on both the extraordinary mechanical properties of the individual CNTs and the potential advantages of nanoscale reinforcing particles. Research on CNT composites is diverse incorporating a variety of matrix materials and CNT types and sizes. Realizing the promise of enhanced mechanical properties relies on the ability to disperse the CNTs uniformly in the host material and on achieving good interfacial bonding for transferring loads across the matrix-fiber interface.[2–4] The dynamic properties of composite materials in the ultrasonic region of the spectrum can provide information about the mechanical moduli, fiber/matrix coupling, and structural integrity of composite materials.[5–8] The attenuation spectra of ultrasonic stress waves are sensitive to the coupling of the matrix and reinforcing inclusion as well as the homogeneity over length scales relevant for structural applications. The phase velocity spectra can be used to determine the dynamic mechanical moduli of a material, while the velocity dispersion is linked to the attenuation (as discussed below) and thus sensitive to a similar list of properties. Nonlinear mechanical properties of composites can also be used to find signs of microstructural degradation.[9]

Broadband ultrasonic spectroscopy is a technique utilizing time-localized signals to determine the phase velocity and attenuation coefficient spectra over a range of frequen-

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

cies simultaneously.[10] The attenuation coefficient and phase velocity spectra are components of the complex wavenumber and are further interlinked through the Kramers–Kronig (KK) relations. Fundamentally rooted in causality, KK relations provide linkages between the physical properties that govern the response of matter and materials to external stimuli. Due to their general foundations[11] KK relations have proven to be adaptable and applicable to a wide array of tasks which include measuring fundamental material parameters, establishing the consistency of laboratory data, and building causally-consistent physical models. KK relations between components of the complex wavenumber using the method of subtractions have been established for both homogeneous and inhomogeneous materials.[12] One complication in adapting KK relations for the analysis of data is the knowledge gap that exists between the infinite bandwidth required by the KK integrals and the measured data which are inherently bandlimited. The impact of this gap on KK calculations depends on many factors, both general and system-dependent. However, finite bandwidth approximations to the KK relations have proven to be applicable to measured ultrasonic attenuation and velocity spectra for suspensions exhibiting resonant features,[13–16] and have also proven accurate for the analysis of systems exhibiting monotonic behavior where the attenuation varies as a frequency power-law over limited experimental bandwidths.[17,18] For the power-law attenuation, the KK relations predict that the velocity dispersion also varies as a power-law (or logarithmically in the case that the attenuation is linear in frequency), although with a scaling factor that is a function of the power-law exponent.[17,19] Due to convergence problems with a KK formulation in use at the time, these results were first derived with an alternate technique, known as the time-causal method.[19] Using the method of subtractions,[12,20] the convergence problems were circumvented permitting valid KK calculations to be performed. In this work, we present data for the attenuation coefficient and phase velocity spectra of longitudinal mode elastic waves in the ultrasonic frequency regime for a series of nylon matrix materials containing various concentrations of MWCNTs. The attenuation coefficient spectra for the samples examined in this work are found to follow a power-law dependence on frequency and the dispersion data exhibits the variation predicted by the KK relations.

II. THEORY

In a variety of media (including some liquids, soft mammalian tissues, and solid polymers) over a finite bandwidth, the attenuation coefficient of ultrasonic waves appears to be adequately modeled by a power-law dependence on frequency

$$\alpha(f) = \alpha_0 |f|^y, \tag{1}$$

where α_0 and y are real constants, with $1 \leq y \leq 2$. The frequency response of a medium of thickness h can be characterized by its transfer function

TABLE I. The densities of the samples examined in this work.

Sample	Density (g/cm^3)
20% MWCNT	1.25
10% MWCNT	1.20
5% MWCNT	1.17
Nylon (0% MWCNT)	1.14

$$H(f,h) = \exp[iK(f)h], \tag{2}$$

where

$$K(f) = \frac{2\pi f}{c_p(f)} + i\alpha(f) \tag{3}$$

is the conventional complex wavenumber, and $c_p(f)$ is the phase velocity. Given that the power-law attenuation persists throughout the spectrum, the KK relations predict that the phase velocities at two frequencies f and f_0 have the following relationship:[17,19]

$$\frac{1}{c_p(f)} = \frac{1}{c_p(f_0)} + \frac{\alpha_0}{2\pi}\tan\left(y\frac{\pi}{2}\right)(|f|^{y-1} - |f_0|^{y-1}) \quad \text{for} \quad 1 < y \leq 2, \tag{4a}$$

which takes the form

$$\frac{1}{c_p(f)} = \frac{1}{c_p(f_0)} - \frac{\alpha_0}{\pi^2}\ln\left|\frac{f}{f_0}\right| \quad \text{in the limit } y \to 1. \tag{4b}$$

These causally-consistent functional forms for the attenuation and phase velocity spectra have been shown to accurately describe the behavior of real materials over band limited windows in the low-megahertz region of the acoustic spectrum.[17–19]

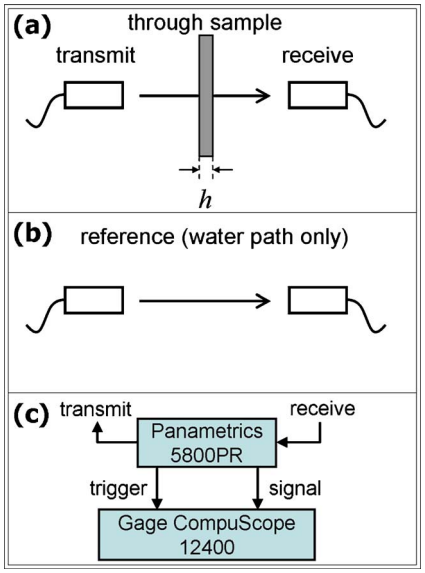


FIG. 1. (Color online) (a) The through sample and (b) reference (water path only) signal acquisition steps of the substitution method. (c) Schematic diagram of the instrumentation in the measurement system.

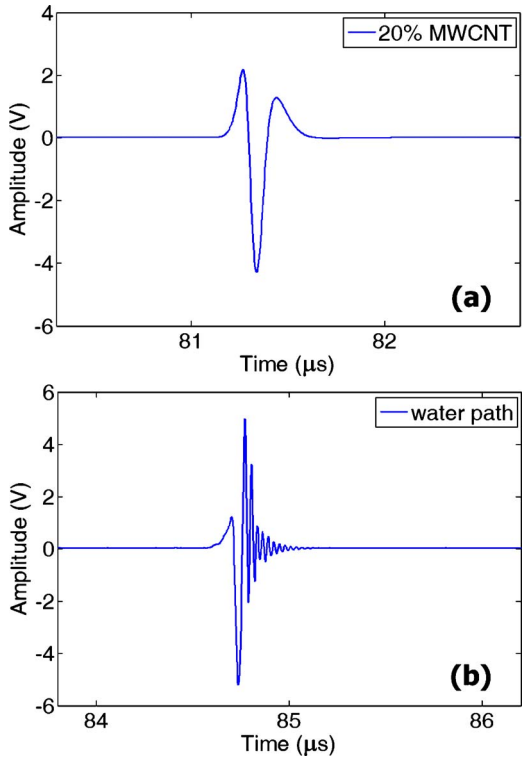


FIG. 2. (Color online) Representative waveforms captured during (a) through-sample acquisition (20% MWCNT sample) and (b) reference (water path only) acquisition.

III. MATERIALS

The samples were provided as extruded plates by Ensinger Inc. with weight fractions of MWCNTs of 0% (pure nylon), 5%, 10%, and 20%. The polymer matrix for the samples is nylon 6,6. The MWCNTs were produced by Hyperion Catalysis International, Inc. (Cambridge, MA). The densities of the samples are shown in Table I.

IV. DATA ACQUISITION AND ANALYSIS

The velocity and attenuation data were determined using the broadband ultrasonic spectroscopy technique implemented in a through-transmission set-up, as shown in Fig. 1. The ultrasound was generated and received by a pair of PVDF transducers (Olympus NDT/Panametrics) immersed in a water bath and separated by 12.5 cm. The transmitter was excited by a broadband pulser/receiver unit (Olympus NDT/Panametrics 5800). The received signals were captured by a digital oscilloscope (GaGe Applied Compuscope 12400) where they were digitized to 12 bits at a rate of 400 Msamples/s. For each sample, through-transmitted ultrasonic signals were acquired [Fig. 1(a)] from five sites at normal incidence. A representative through-sample waveform is shown in Fig. 2(a). In addition to the through-sample acquisitions, waterpath only waveforms were captured to serve as the reference data in the analysis, as shown in Fig. 1(b). A captured signal from a waterpath only acquisition is shown in Fig. 2(b). Sample thicknesses were measured ultrasonically at each acquisition site on the sample and are derived from time-of-flight measurements from pulse-echo signals off the near sample face for each transducer and a waterpath

TABLE II. The fit values for the attenuation constant, the exponent of the power law attenuation coefficients, and the offset a_0 for the four samples.

Sample	α_0 (Np mm^{-1} MHz^{-y})	y	a_0 (Np/mm)
20% MWCNT	0.0217	1.169	0.0108
10% MWCNT	0.0272	1.123	−0.0048
5% MWCNT	0.0264	1.154	–0.0046
Nylon (0% MWCNT)	0.0267	1.187	−0.0038

only through-transmission acquisition.[21] Waveform data were acquired by accumulating 5000 raw signals and the accumulated signal was recorded to disk for further processing. The discrete Fourier transforms of the through-transmitted and waterpath signals were taken, and the amplitude and phase spectra from each were then used to compute the attenuation coefficient and phase velocity for each site on a given sample. The attenuation coefficient was determined using the following relation:

$$\alpha(f) = \alpha_w(f) + \frac{\ln \frac{A_{\text{ref}}(f)}{A_{\text{thru}}(f)} + \ln T}{h}, \tag{5}$$

where $\alpha_w(f)$ is the attenuation coefficient for water, $A_{\text{thru}}(f)$ and $A_{\text{ref}}(f)$ are the Fourier amplitude spectra of the through-sample and reference waveforms, respectively, h is the sample thickness, and $T=4Z_wZ_s/(Z_w+Z_s)^2$ is the single-pass

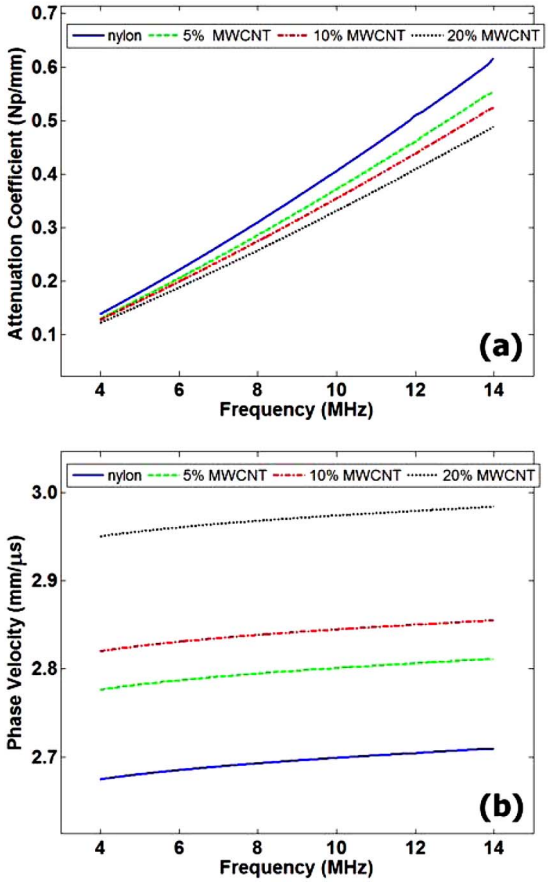


FIG. 3. (Color online) The measured (a) attenuation coefficient and (b) phase velocity spectra for the four samples studied in this work.

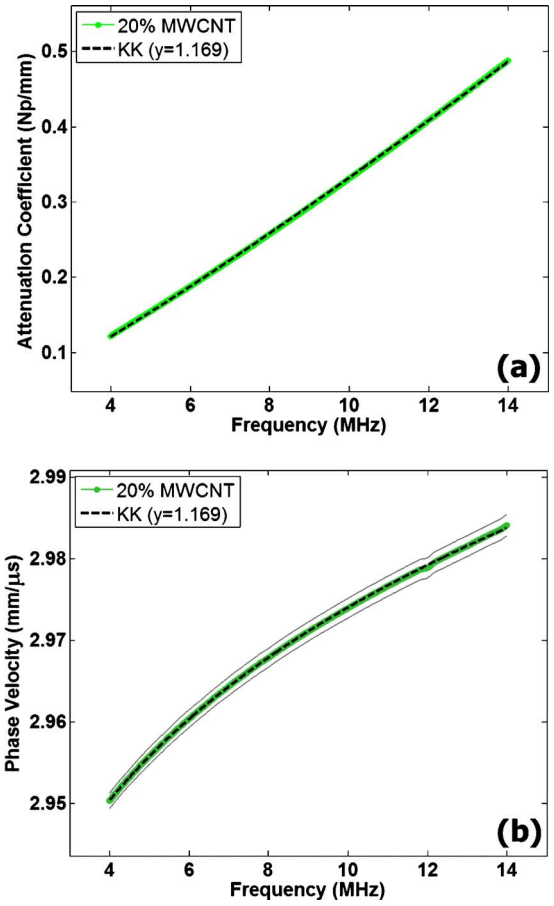


FIG. 4. (Color online) The experimental attenuation coefficient and phase velocity spectra for the 20% MWCNT sample with their associated KK predicted curves. The solid lines in the bottom panel are the standard deviation limits of the measured quantities. (The standard deviation limits for the attenuation data are smaller than the width of the plotted curve.)

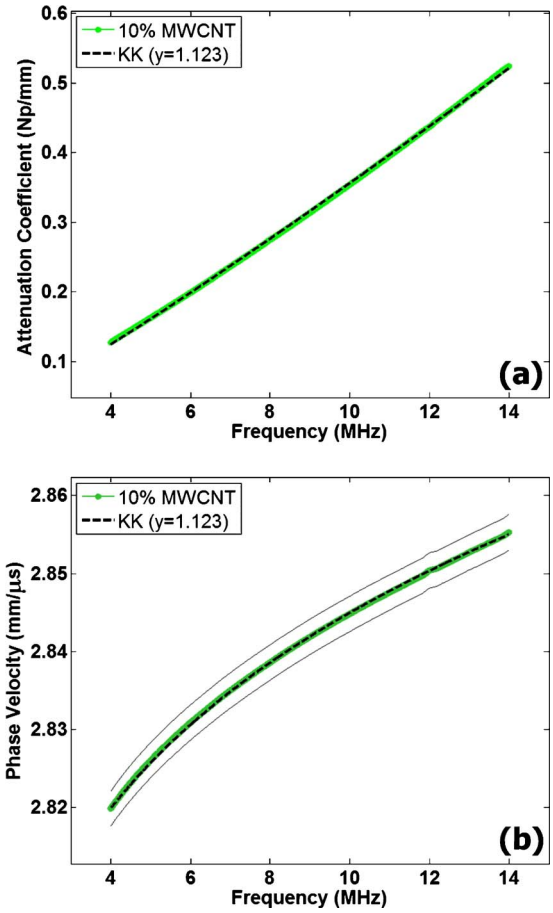


FIG. 5. (Color online) The experimental attenuation coefficient and phase velocity spectra for the 10% MWCNT sample with their associated KK predicted curves. The solid lines in the bottom panel are the standard deviation limits of the measured quantities. (The standard deviation limits for the attenuation data are smaller than the width of the plotted curve.)

amplitude transmission coefficient including both the entry and exit water/sample interfaces, where $Z_w = \rho_w c_w$ and $Z_w = \rho_s c_p(f)$ are the characteristic acoustic impedances of the water and the sample, respectively (the sample densities, $\{\rho_s\}$, are given in Table I). The phase velocity relation is

$$c(f) = \frac{c_w}{1 - c_w \dfrac{\Delta\phi(f)}{2\pi f h}}, \tag{6}$$

where $\Delta\phi(f) = \phi_{\text{thru}}(f) - \phi_{\text{ref}}(f)$ is the difference in the unwrapped phase spectra from the two signals compensated for sheet offsets. The speed of sound in water, c_w, was determined by the water temperature using the formula from Ref. 22.

After the attenuation coefficient and phase velocity spectra were measured at five sites on a given sample, the spectra were averaged across the sites to yield a single attenuation and velocity spectrum for each sample. To determine the parameters α_0 and y in each case, the attenuation data were fitted to a model function of the form $\alpha(f) = a_0 + \alpha_0 f^y$, where a_0 is the offset of the attenuation data at zero frequency (beyond the low frequency limits of the measurement spectrum). This was done over a range of exponents from $y = 1.001$ to $y = 1.349$, and the exponent y and associated coefficient α_0 that gave the best fit to the dispersion data using Eq. (4a) were used. The values for the parameters α_0 and y as determined by this procedure are given in Table II. (See Table III for phase velocity values at $f_{\text{low}} = 4$ MHz.)

V. RESULTS AND DISCUSSION

The results for the attenuation coefficient and phase velocity for all four samples are shown together in the panels of Fig. 3. The comparisons of the measured attenuation and dispersion data with the KK power-law model predictions are shown for each respective sample in Figs. 4–7. The top panel in each of Figs. 4–7 shows the attenuation coefficient and its power-law fit. The bottom panel in each figure compares the measured phase velocity spectra with the KK prediction in the form of Eq. (4a) using the parameters from the attenuation fit.

The attenuation coefficient spectrum for the pure nylon sample exhibited the highest values throughout the measurement bandwidth, and across the samples the attenuation was found to decrease with increasing MWCNT content (in the same way as with the dispersion as discussed below), as shown in Fig. 3(a). The attenuation coefficient of the nylon-only sample also exhibited the steepest rise with frequency, and the largest power-law exponent of about 1.19. The at-

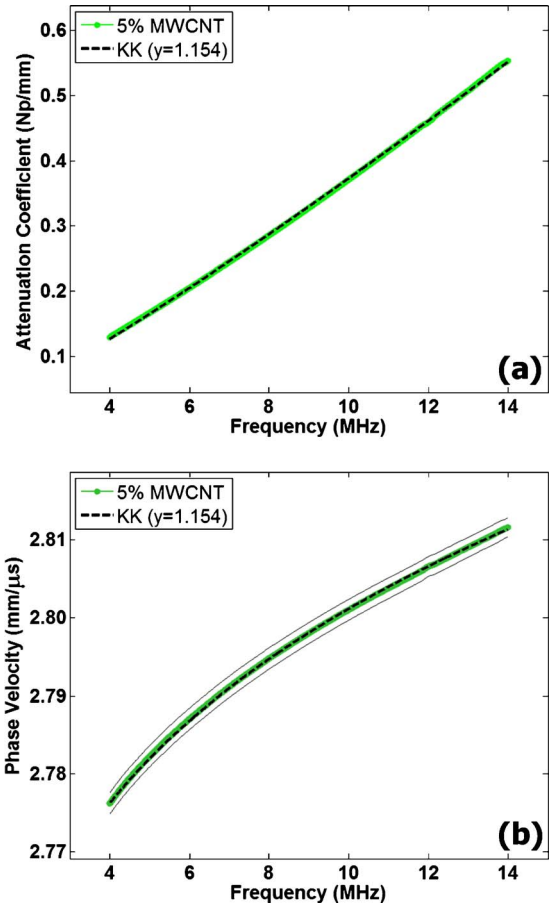


FIG. 6. (Color online) The experimental attenuation coefficient and phase velocity spectra for the 5% MWCNT sample with their associated KK predicted curves. The solid lines in the bottom panel are the standard deviation limits of the measured quantities. (The standard deviation limits for the attenuation data are smaller than the width of the plotted curve.)

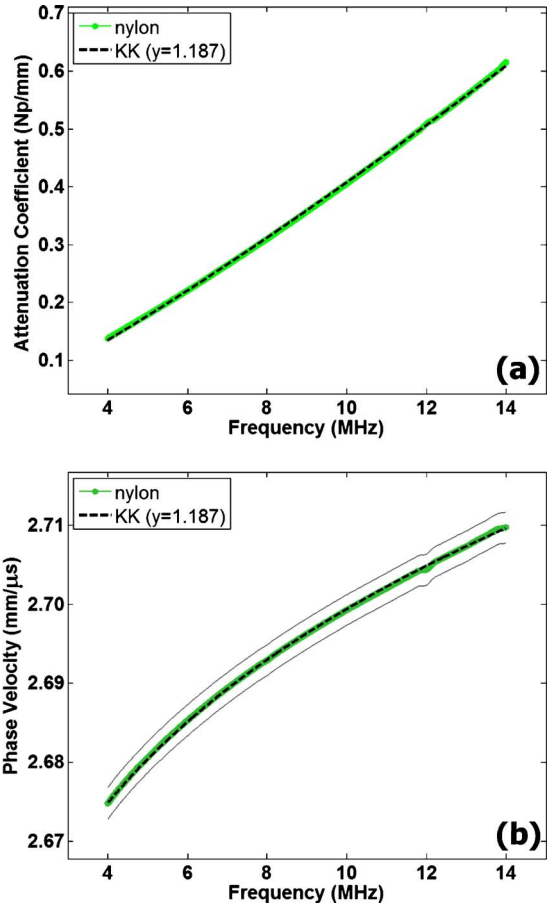


FIG. 7. (Color online) The experimental attenuation coefficient and phase velocity spectra for the nylon (0% MWCNT) sample with their associated KK predicted curves. The solid lines in the bottom panel are the standard deviation limits of the measured quantities. (The standard deviation limits for the attenuation data are smaller than the width of the plotted curve.)

tenuation coefficient of the 20% MWCNT sample was the smallest among the samples across the spectrum and the slowest rise with frequency. The power-law exponent of the 5% MWCNT sample and 10% MWCNT sample fell monotonically from the nylon value, but the 20% MWCNT sample had a greater exponent than both the 5% and 10% MWCNTs while still remaining lower than the nylon-only. The reasons for the monotonic decrease in attenuation with increasing MWCNT concentration are not clear, although this could due to differences in the bulk attenuation of the two media (volume related effect) or some aspect of the coupling between the two phases (interphase boundary effects and scaling with the interphase surface area). Simple series- and parallel-type volumetric law of mixtures models were unable to account for the differences in attenuation among the samples. It is likely that a combination of factors contributes to the observed trend, but a more definitive judgment on this manner is beyond the scope of the present work.

The phase velocity and dispersion results are summarized in Table III. The mean phase velocities (averaged across the 4–14 MHz bandwidth) for the four samples exhibit an increase as MWCNT content rises, starting from a value of 2.695 mm/μs for the pure nylon (0% MWCNT) sample up to 2.970 mm/μs for the 20% MWCNT sample. This is indicative of an increase in the Young's modulus of about 28% from pure nylon to the composite with the highest fraction of MWCNT, assuming that nylon's Poisson's ratio of about 0.39 continues to hold.[23] The dispersion, defined as $\Delta(1/c)=[1/c(f_{\text{low}})]-[1/c(f_{\text{high}})]$ [as is natural based on the form predicted in Eq. (4)], varies from sample to sample with the dispersion decreasing with increasing MWCNT content. This is the same trend as seen with the attenuation data as expected from the causally-consistent forms of Eqs. (3) and (4). (An interesting observation is that the change in phase velocity, $\Delta c=c(f_{\text{high}})-c(f_{\text{low}})$, as dispersion is often defined, did not change significantly from sample to sample.) The pure nylon exhibited the greatest dispersion, with a variation in $\Delta(1/c)$ about 1.26 times higher than that of the 20% MWCNT sample. These dispersion data are also consistent with the attenuation coefficient determinations as linked by the KK relations, as can be seen in the Figs. 4(b), 5(b), 6(b), and 7(b). In each case the KK prediction is almost indistinguishable from the measured phase velocity. It is clear that the KK predicted forms are quantitatively consistent with these data. It is not clear, however, how far beyond the measurement bandwidth one could expect these forms to continue to hold.

VI. CONCLUSION

The phase velocity and attenuation coefficient spectra of composite materials with varying amounts of MWCNT con-

TABLE III. Phase velocity measurements from the four samples. The first three columns are the values at 4, 9, and 14 MHz, respectively. The fourth column is the phase velocity averaged over the entire bandwidth, and the last column is the dispersion.

Sample	$c_p(f_{\text{low}})$ (mm/μs)	$c_p(f_{\text{median}})$ (mm/μs)	$c_p(f_{\text{high}})$ (mm/μs)	$\bar{c}_p$ (mm/μs)	$1/c_p(f_{\text{low}})-1/c_p(f_{\text{high}})$ (mm/μs)$^{-1}$
20% MWCNT	2.950	2.971	2.984	2.970	3.83×10^{-3}
10% MWCNT	2.820	2.842	2.855	2.840	4.40×10^{-3}
5% MWCNT	2.776	2.798	2.812	2.797	4.53×10^{-3}
Nylon (0% MWCNT)	2.675	2.696	2.710	2.695	4.82×10^{-3}

tent (from 0 to 20% by weight) were measured using a broadband technique. The samples were found to be effectively homogeneous on spatial scales from the low end of ultrasonic wavelengths investigated and up (>0.2 mm). Over the measurement bandwidth, these spectra were found to be a consistent KK model that utilizes a frequency power-law form for the attenuation coefficient. The mean phase velocity increased monotonically with the rising MWCNT content, indicating an increase in the mechanical moduli with MWCNT concentration. The attenuation coefficient and the dispersion both showed the opposite trend, decreasing with increasing MWCNT content, consistent with the predictions of the KK model.

ACKNOWLEDGMENT

Support for this research by the Office of Naval Research, Solid Mechanics Program, ONR Grant No N00014-07-1-1010 (Dr. Yapa D. S. Rajapakse, Program Manager), is acknowledged.

[1]S. Iijima, "Helical microtubules of graphitic carbon," Nature (London) **354**, 56–58 (1991).
[2]P. J. F. Harris, "Carbon nanotube composites," Int. Mater. Rev. **49**, 31–43 (2004).
[3]J. N. Coleman, U. Khan, W. J. Blau, and Y. K. I. Gun'ko, "Small but strong: A review of the mechanical properties of carbon nanotube-polymer composites," Carbon **44**, 1624–1652 (2006).
[4]J. Njuguna, K. Pielichowski, and J. R. Alcock, "Epoxy-based fibre reinforced nanocomposites," Adv. Eng. Mater. **9**, 835–847 (2007).
[5]S. I. Rokhlin, W. Huang, and Y. C. Chu, "Ultrasonic scattering and velocity methods for characterization of fiber-matrix interphases," Ultrasonics **33**, 351–364 (1995).
[6]Y. C. Chu and S. I. Rokhlin, "Determination of macromechanical and micromechanical and interfacial elastic properties of composites from ultrasonic data," J. Acoust. Soc. Am. **92**, 920–931 (1992).
[7]K. Balasubramaniam, S. Alluri, P. Nidumolu, P. R. Mantena, J. G. Vaughan, and M. Kowsika, "Ultrasonic and vibration methods for the characterization of pultruded composites," Composites Eng. **5**, 1433–1451 (1995).
[8]J. Wu, C. Layman, S. Murthy, and R.-B. Yang, "Determine mechanical properties of particulate composite using ultrasound spectroscopy," Ultrasonics **44**, e793–e800 (2006).
[9]K. Y. Jhang, "Applications of nonlinear ultrasonics to the NDE of material degradation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 540–548 (2000).
[10]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).
[11]J. S. Toll, "Causality and the dispersion relations: Logical foundations," Phys. Rev. **104**, 1760–1770 (1956).
[12]R. L. Weaver and Y.-H. Pao, "Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media," J. Math. Phys. **22**, 1909–1918 (1981).
[13]J. Mobley, "Finite-bandwidth Kramers-Kronig relations for acoustic group velocity and attenuation derivative applied to encapsulated microbubble suspensions," J. Acoust. Soc. Am. **121**, 1916–1923 (2007).
[14]J. Mobley, K. R. Waters, and J. G. Miller, "Causal determination of acoustic group velocity and frequency derivative of attenuation with finite-bandwidth Kramers-Kronig relations," Phys. Rev. E **72**, 016604 (2005).
[15]J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **108**, 2091–2106 (2000); "Erratum: 'Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles [J. Acoust. Soc. Am. 108, 2091–2106 (2000)]'," J. Acoust. Soc. Am. **112**, 760–761 (2002).
[16]J. Mobley and R. E. Heithaus, "Ultrasonic properties of a suspension of microspheres supporting negative group velocities," Phys. Rev. Lett. **99**, 124301 (2007).
[17]K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers–Kronig relations for media with ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[18]K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "Differential forms of the Kramers–Kronig dispersion relations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 68–76 (2003).
[19]T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[20]H. M. Nussenzveig, *Causality and Dispersion Relations* (Academic, New York, 1972).
[21]D. K. Hsu and M. S. Hughes, "Simultaneous ultrasonic velocity and sample thickness measurement and application in composites," J. Acoust. Soc. Am. **92**, 669–675 (1992).
[22]W. Marczak, "Water as a standard in the measurements of speed of sound in liquids," J. Acoust. Soc. Am. **102**, 2776–2779 (1997).
[23]G. S. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987), p. 550.

Sensitivity of acoustic microscopy for detecting three-dimensional nanometer gaps embedded in a silicon structure

Hironori Tohmyoh[a] and M. A. Salam Akanda
Department of Nanomechanics, Tohoku University, Aoba 6-6-01, Aramaki, Aoba-ku, Sendai 980-8579, Japan

(Received 21 January 2009; revised 8 May 2009; accepted 8 May 2009)

The sensitivity of acoustic microscopy for detecting three-dimensional defects in a Si structure is reported. Circular, nanometer gaps with diameters ranging from 5 to 1000 μm were embedded in Si disks by a direct bonding technique, and these were visualized using acoustic microscopy. The limits of detection for the gap thickness and diameter were observed simultaneously in samples with gaps of 4 and 140 nm. The behavior of the sensitivity in detecting the gaps can be explained by a simple analytical model. It is shown experimentally and theoretically that the gap thickness and diameter are not independent variables as regards detection. The sensitivity of acoustic microscopy is governed by the three-dimensional features of the embedded defects.
 [DOI: 10.1121/1.3147493]

PACS number(s): 43.35.Sx, 43.35.Zc, 43.58.Vb [TDM] Pages: 98–102

I. INTRODUCTION

Acoustic microscopy/imaging is one of a number of image-based nondestructive techniques,[1–4] and has been an indispensable tool for evaluating the reliability of microelectromechanical systems (MEMSs) or nanoelectromechanical systems (NEMSs). What size of defect can be observed by acoustic microscopy? The answer to this question will clarify the applicability of this technique for the inspection of future MEMS/NEMS. This simple question can be interpreted in two ways. One interpretation is with regard to the lateral resolution, which indicates the separation between two sound sources in the object plane that can be clearly distinguished from each other.[2,5–7] The resolution determines the minimum dimension of detectable defects in the plane perpendicular to the sound propagation. The other interpretation is in terms of the minimum thickness of defect that can be detected, which is the dimension of the defect in the direction of sound propagation. To avoid confusion with the depth of the defect beneath the surface of the structure, from this point onwards, the defect thickness is referred to as the separation. The lateral resolution is related to the wavelength of the transmitted waves and it can be predicted theoretically from the frequency characteristics of the transmitted waves and the geometrical configuration of the acoustic transducers. On the other hand, it is very difficult to quantify the minimum detectable separation both experimentally and theoretically because the minimum detectable separation may be on the nanometer scale. Newton's ring experiments enable us to observe the interaction between acoustic waves and thin gaps.[8] However, the gaps that give rise to Newton's ring are not surrounded by other structures, and moreover, the gap separation varies in the object plane. Therefore, it is difficult to discuss the detection limit for defects embedded in a sample using Newton's ring experiments. Ogura[6] reported that 5 nm wide gaps introduced in a Si dioxide layer grown on a substrate and embedded in a sample by a Si-on-insulator technique could be detected by acoustic microscopy. However, the sensitivity for detecting the gaps as a function of the separation was not determined and the detection limit for observing defects embedded in the sample was not clearly expressed.

Until now, the sensitivities for the minimum detectable defect separation and the lateral resolution have usually been discussed separately. However, these should simultaneously be considered because the separation of the defects changes as a result of the acoustic pressure, and this change depends on the size of the defects and the measurement conditions. For this reason, a technique for embedding precise nanometer gaps in samples was developed,[9] and the gaps were visualized by acoustic microscopy employing a 100 MHz focused ultrasonic transducer. Moreover, the intensity of the acoustic pressure applied to the gap during imaging was increased by inserting a polymer acoustic matching layer between the water and the Si sample.[10] However, only patterns with gap separations of 9.7 nm and widths up to 300 μm were analyzed and the detection limit for the gap separation was not found. In this paper, the detection limits for both the gap separation and the gap diameter are simultaneously observed in samples with circular gaps with separations of 4 and 140 nm and diameters in the range 5–1000 μm.

II. CONCEPT OF SENSITIVITY FOR DETECTING 3D NANOMETER GAPS

Consider a circular gap of diameter D and separation H embedded in a structure. The top surface of the gap is located at a depth h from the surface of the structure. For a rough estimate of the elastic movement at the top of the circular gap under acoustic pressure, a simple disk model with the

[a]Author to whom correspondence should be addressed. Electronic mail: tohmyoh@ism.mech.tohoku.ac.jp

 0001-4966/2009/126(1)/98/5/$25.00

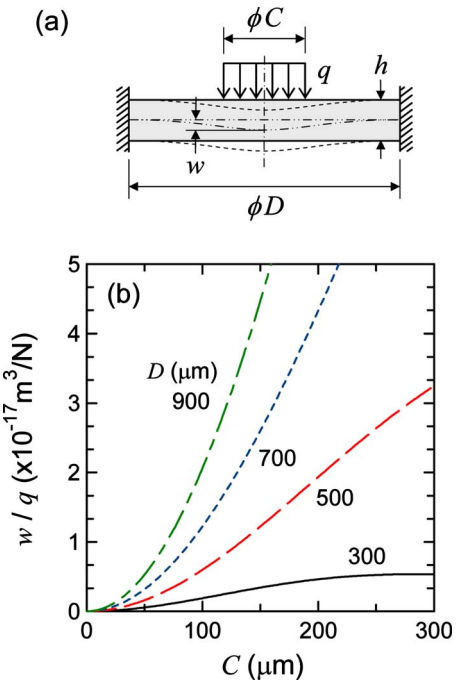


FIG. 1. (Color online) Concept of sensitivity for detecting 3D gaps. (a) Model for deformation of a circular gap embedded in a sample. The displacement at the top surface of the gap can be determined from the circular disk model. Here the edge of the disk is rigidly fixed, and a uniformly distributed load is applied symmetrically with respect to the center of the disk. (b) Relationships between w/q and C for various values of D.

dimensions of the top part covering the gap is used. The disk is considered to deform under the loading conditions where the load, q, is uniformly distributed over a central part of the disk with diameter C, as shown in Fig. 1(a). The cylindrical boundary surface of the model is considered to be rigidly fixed. In this case, the displacement, w, of the disk at its center is given by[11]

$$w = \frac{qC^4}{256A}\left[\left(\frac{D}{C}\right)^2 - \ln\left(\frac{D}{C}\right) - \frac{3}{4}\right], \tag{1}$$

where E is Young's modulus, ν Poisson's ratio, and $A = Eh^3/[12(1-\nu^2)]$.

The structural material is considered to be Si [100] because it is frequently used for MEMS/NEMS. The values of E and ν for Si [100] are 130.8 GPa (Ref. 12) and 0.28,[13] respectively, and the gaps are located at a depth $h = 0.5$ mm. Figure 1(b) shows w/q for $D = 300$, 500, 700, and 900 μm as functions of C. The value of w/q increases as both C and D increase. For example, at $C = 100$ μm, the value of w/q for $D = 900$ μm is about 2.1×10^{-17} m^3/N but it is only about 0.2×10^{-17} m^3/N for $D = 300$ μm. In the case of $q = 1$ GPa, the values of w for $D = 900$ and 300 μm are 21 and 1.9 nm, respectively. Here, the detection sensitivity is considered to decrease as the gap closes. Therefore, a gap with $D = 900$ μm and $H = 10$ nm may be difficult to detect by an ultrasonic transducer with $q = 1$ GPa and $C = 100$ μm, since $w > H$ and the gap will close. On the other hand, a gap with $D = 300$ μm and $H = 10$ nm may be detectable by the same ultrasonic transducer provided that the transducer has sufficient lateral resolution. In the above discussion the three-dimensional (3D) nature of the defects, i.e., D and H for circular gaps, is considered in determining the detection limits of acoustic microscopy, in addition to the characteristics of the incident acoustic waves, i.e., q and C. Note that the sensitivity for detecting gaps is also dependent on h.

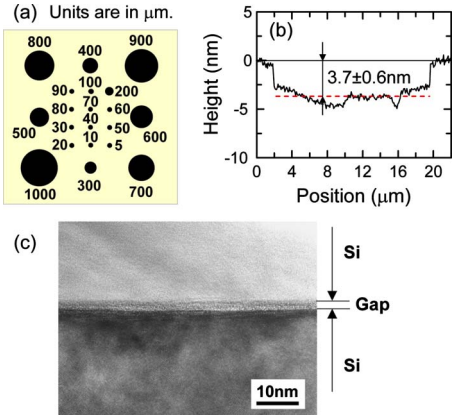


FIG. 2. (Color online) Details of the 3D nanometer gaps embedded in the Si sample. (a) Pattern of gaps fabricated on a Si [100] chip. (b) An example of the height profile of a circular gap for $H = 4$ nm and $D = 20$ μm obtained by AFM. (c) An example of a TEM image of a gap with $H = 4$ nm and $D = 500$ μm.

III. EXPERIMENT

Three chips measuring $20 \times 20 \times 0.5$ mm^3 were cut-out from Si [100] wafers, and circular patterns with D ranging from 5 to 1000 μm, and $H = 4$ nm were introduced onto two of the chips using photolithography and fast-atom-beam etching, as shown in Fig. 2(a). One of the two patterned chips and the non-patterned chip were bonded together by direct bonding using a wafer bonder so that the gaps were embedded. The other Si chip was used for measuring the height of the gaps by atomic force microscopy (AFM). More details on the fabrication procedure for preparing these samples are described elsewhere.[9] Figure 2(b) shows an example of the depth profile of a 20 μm-diameter gap obtained by AFM, which shows the depth to be 3.7 ± 0.6 nm. Another bonded sample with gaps of 140 ± 1 nm separation was prepared using the same procedure.

Pure water at 295 K was used as the coupling liquid. Two types of broadband, focused ultrasonic transducers were used to record acoustic images from the back surface of the samples, which was in contact with the water. One transducer had a nominal frequency of 100 MHz and a focal length of 12.7 mm. The other had a nominal frequency of 30 MHz and a focal length of 19.1 mm. The diameter of the piezoelectric element of both transducers was 6.35 mm. After acquiring the acoustic images, the sample with $H = 4$ nm was cut using a micro-sampling technique, and its cross-section viewed using a transmission electron microscope (TEM). Figure 2(c) shows an example of the cross-sectional view around a 500 μm-diameter gap obtained by TEM. The contrast in the image between the top and bottom Si chips is due to the difference in crystal orientation. The gap was filled with carbon to get a clearer picture. The TEM analysis showed that the gap separation was unaffected by the bonding operation.

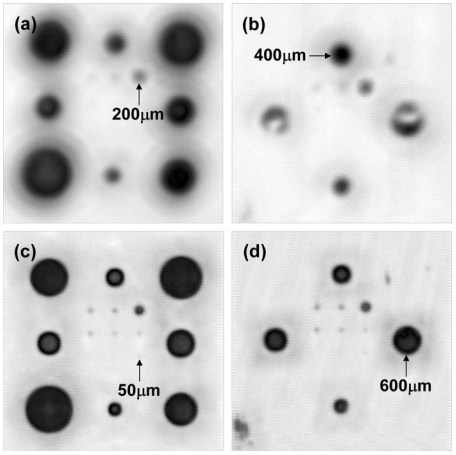


FIG. 3. Acoustic images of 3D nanometer gaps embedded in Si samples (4.5×4.5 mm^2). The images in (a) and (b) were obtained with the 30 MHz ultrasonic transducer and those in (c) and (d) were obtained with the 100 MHz ultrasonic transducer. The value of H for the images in (a) and (c) was 140 nm, and that for the images in (b) and (d) was 4 nm.

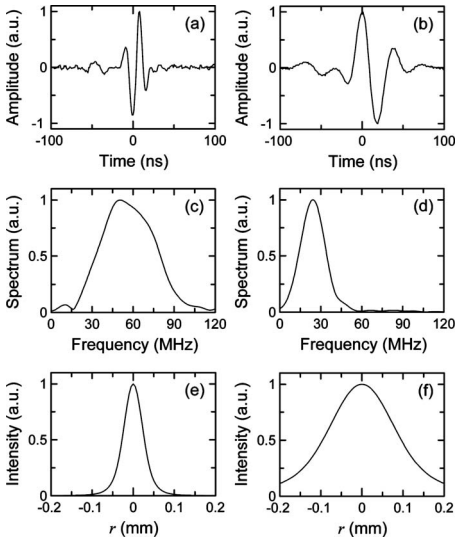


FIG. 4. Echo waveforms from the back surface of the bonded Si sample obtained with (a) the 30 MHz ultrasonic transducer and (b) the 100 MHz ultrasonic transducer. The corresponding amplitude spectra of (a) and (b) are shown in (c) and (d), respectively. The distributions of beam intensity at the focal point for (e) the 30 MHz ultrasonic transducer and (f) the 100 MHz ultrasonic transducer.

IV. RESULTS AND DISCUSSION

A. Acoustic images

Acoustic images of the samples with $H=140$ and 4 nm obtained using the 30 MHz ultrasonic transducer are shown in Figs. 3(a) and 3(b), respectively. Images obtained using the 100 MHz ultrasonic transducer are shown in Figs. 3(c) and 3(d), respectively. In both images obtained using the 30 MHz ultrasonic transducer, gaps smaller than $D=100$ μm are hardly visible. On the other hand, gaps smaller than $D=40$ μm are difficult to identify in both images obtained with the 100 MHz ultrasonic transducer. This is due to limitations in the lateral resolution for each of the ultrasonic transducers used.

The lateral resolution, d^{PE}, is defined as the separation between two sound sources, which can clearly be distinguished from each other, and can be determined from the beam intensity at the point at which the ultrasonic transducer is focused; see Appendix.[5–7] Figures 4(a) and 4(b) show the echo waveforms from the back surface of the sample at a position where there is no gap, obtained using the 30 and 100 MHz ultrasonic transducers, and Figs. 4(c) and 4(d) show the corresponding amplitude spectra of Figs. 4(a) and 4(b), respectively. The distributions of the beam intensity at the point at which the transducers are focused are determined as shown in Figs. 4(e) and 4(f), respectively. The estimated values of d^{PE} at the back surface of the samples for the 30 and 100 MHz ultrasonic transducers are 165 and 48 μm, respectively, and these are reasonable compared with the experimental values.

The gaps with $D=500$ and 600 μm in the sample with $H=4$ nm are seen as faded circular marks in the acoustic image obtained with the 30 MHz ultrasonic transducer, but these are clearly visible in the image obtained using the 100 MHz ultrasonic transducer; see Fig. 3. Although the echo transmittance via thin gaps changes depending on the gap thickness,[14] this cannot explain the deference in sensitivity for gaps with different diameters. The blurred 500 and 600 μm-diameter images for the 30 MHz ultrasonic transducer must be due to closure of the gaps as a result of the ultrasonic transmission. From Eq. (1), the value of w/q for $D=500$ μm when $C=d^{PE}=165$ μm is 1.43×10^{-17} m^3/N. If the authors assume the acoustic pressure field formed by the 30 MHz ultrasonic transducer is uniformly distributed, then q becomes 0.28 GPa for $w=H=4$ nm.

In the acoustic image of the sample with $H=4$ nm obtained with the 100 MHz ultrasonic transducer, gaps larger than $D=700$ μm disappear [Fig. 3(d)]. The receiver gains for acoustic imaging were 45 dB for the 30 MHz ultrasonic transducer and 63 dB for 100 MHz ultrasonic transducer, i.e., the gain of the 100 MHz ultrasonic transducer was 18 dB higher than that of the 30 MHz ultrasonic transducer. For simplicity, the acoustic pressure on the gap from the 100 MHz ultrasonic transducer was considered to be -9 dB for pulse echo system (half of -18 dB) compared with the 30 MHz transducer, meaning that the acoustic power applied by the 100 MHz ultrasonic transducer is estimated to be about 35% of that of the 30 MHz ultrasonic transducer. The value of q for the 100 MHz ultrasonic transducer was estimated to be 1.16 GPa for $C=d^{PE}=48$ μm, and the value of D for which the gaps close was estimated to be about 760 μm. Although the estimated detection limit showed reasonable agreement with the experimental one, none of the edges of gaps with $D\geq700$ μm are visible in Fig. 3(d) even though the displacement at the edges should be zero. From this experimental fact, it was suspected that the surfaces of the 4 nm gaps with $D\geq700$ μm were bonded together in the initial bonding process.

B. Sensitivity for detecting gaps

Now consider the sensitivity for detecting gaps by using the relative intensity of the echo from the back surface ($=A_g/A_0$) as a measure, where A_0 is the amplitude of the echo from the back of the sample obtained at a position

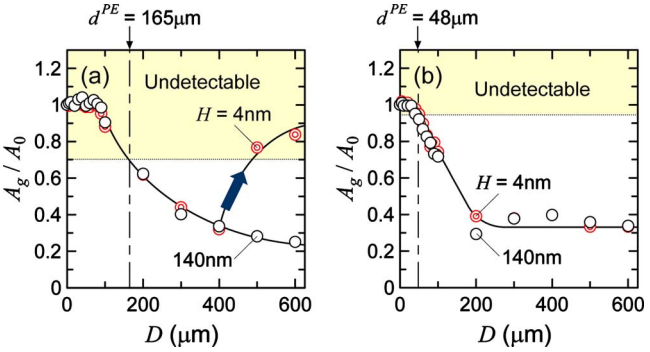


FIG. 5. (Color online) Relationships between A_g/A_0 and D obtained for (a) the 30 MHz ultrasonic transducer and (b) the 100 MHz ultrasonic transducer.

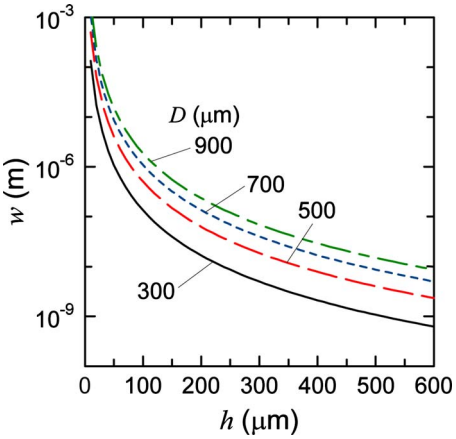


FIG. 6. (Color online) Relationships between w and h for $D=300-900$ μm for the case where $q=0.28$ GPa and $C=165$ μm.

without a gap, and A_g is that obtained at the center of a gap. Lower values of A_g/A_0 indicate a higher sensitivity for the embedded gaps. Figure 5(a) shows the relationships between A_g/A_0 and D in the case of the 30 MHz ultrasonic transducer; those for the 100 MHz ultrasonic transducer are shown in Fig. 5(b).

No gap separation dependency for the values of A_g/A_0 was observed in the case of the 100 MHz ultrasonic transducer in the range of D examined. On the other hand, in the case of the 30 MHz ultrasonic transducer, although A_g/A_0 slightly decreases with increasing D when $H=140$ nm and $D>400$ μm, it increases with D when $H=4$ nm. This is because the top and bottom surfaces of the gaps make contact. In the case of the 30 MHz ultrasonic transducer, in the range of D shown in Fig. 5(a), the value of A_g/A_0 for $H=140$ nm increases as D decreases. A similar tendency was observed for the 100 MHz ultrasonic transducer, which shows an increase in A_g/A_0 as D decreases for $D<200$ μm [Fig. 5(b)]. The decrease in sensitivity for smaller D is due to the dispersion of the acoustic energy applied to the gap, and this is closely related with the distribution of the acoustic pressure in the object plane of the ultrasonic transducers used.

The condition for visible circular gaps is given by $D_L \leq D \leq D_U$, where D_L and D_U are the lower and upper values of the diameter, respectively. The value of D_L is the lateral resolution of the ultrasonic transducer, and D_U is governed by H, h, and the intensity of the applied acoustic waves. For example, in the case of $H=4$ nm and $h=0.5$ mm, the visible range of D is given by 165 μm$\leq D \leq$400 μm for the 30 MHz ultrasonic transducer.

Based on the experimental results and analytical considerations, in the case of smaller defects, e.g., voids, the lateral resolution governs the detectability of defects. The minimum detectable separation $w(=H)$ for smaller defects may be of the order of less than 1 nm. On the other hand, in the case of larger defects, e.g., delamination, the separation of the defects may govern whether the defect can be detected or not. Note that the detectable separation for larger defects increases with the increasing size of the defects.

In this paper, the gaps were located at $h=0.5$ mm. From Eq. (1), it can be seen that the minimum detectable separation $(w=H)$ of the defects depends very much on h. Figure 6 shows the relationships between h and w for different gap diameters ($D=300-900$ μm) obtained with the 30 MHz ultrasonic transducer ($q=0.28$ GPa, $C=165$ μm). The value of w increases with decreasing h. Note that the detectable gap separation increases for the defects located at shallower positions.

To verify the present model, the top surface of the sample with gaps of $H=140$ nm was lapped and polished down to $h=0.2$ mm. Figure 7 shows an acoustic image of this sample obtained after thinning using the 30 MHz ultrasonic transducer. For this transducer, the value of D for $w=H=140$ nm at $h=0.2$ mm was calculated to be 710 μm from Eq. (1). As shown in Fig. 7, the central parts of gaps with $D\geq800$ μm, which were dark in Fig. 3(a) for $h=0.5$ mm, become light, indicating that the surfaces of the gaps had come into contact in the central regions. The experimental results are in good agreement with predictions obtained from Fig. 6, and the facts support the validity of the present analytical model for calculating the sensitivity of acoustic microscopy for detecting embedded defects.

V. CONCLUSIONS

The detection limits for both the depth and width of gaps detected using acoustic microscopy were studied using Si

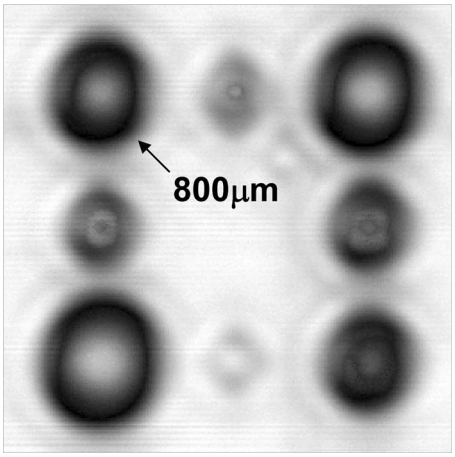


FIG. 7. Acoustic image of 3D nanometer gaps with $H=140$ nm obtained with the 30 MHz ultrasonic transducer after lapping and polishing the surface of the Si sample (4.5×4.5 mm^2). The acoustic image of the same sample before polishing is shown in Fig. 3(a).

samples with embedded circular gaps with depths of 4 and 140 nm and diameters in the range from 5 to 1000 μm. Acoustic imaging of the samples was performed using 30 and 100 MHz focused ultrasonic transducers. For both transducers, the sensitivity for detecting the gaps decreases with decreasing gap diameter due to the dispersion of the acoustic energy applied to the gaps. On the other hand, the sensitivity also decreases with increasing gap diameter for the 30 MHz ultrasonic transducer due to the surfaces of the gap coming into contact. The phenomenon was interpreted by a simple analytical model, in which the acoustic pressure was uniformly distributed in the central region of the disk with a fixed circular boundary.

ACKNOWLEDGMENTS

The author would like to acknowledge Professor M. Saka for valuable discussions throughout this work. The author also thanks Mr. H. Hirayama for his help in the experiments and Dr. T. Fukushima for his help in preparing the samples. This work was partly supported by Grant-in-Aid for Young Scientists (B) Grant No. 19760057 and by Grant-in-Aid for JSPS Fellows Grant No. 19 07379. A part of this work was performed at the Micro/Nano-Machining Research and Education Center of Tohoku University.

APPENDIX: CALCULATION OF BEAM INTENSITY

The beam intensity at the point at which the ultrasonic transducer is focused can be expressed in terms of the amplitude spectrum, ϕ, and the geometrical configuration of the ultrasonic transducer and is given as[5–7]

$$\xi_e(r) = \sum_{f_L}^{f_U} \{\phi[2J_1(B)/B]^2\}, \tag{A1}$$

where r is the radial distance from the central axis of the ultrasonic transducer, J_1 is a Bessel function of the first kind and first order, $B=k_W ra/z_0$, $k_W(=2\pi f/c_W)$ is the ultrasonic wave number in water, f is frequency, c_W is the longitudinal wave velocity in water (=1500 m/s), f_L is the lower frequency limit, and f_U is the upper frequency limit. In this study, the frequencies for which ϕ was 10% of the maximum amplitude were used for f_L and f_U. $2a$ and z_0 are the diameter of the piezoelectric element and the focal length of used ultrasonic transducer, respectively. The lateral resolution, d^{PE}, must satisfy the following equation:[7] $\xi_e(0) - 2\xi_e(d^{PE}/2) + \xi_e(d^{PE}) = 0$.

[1] Z. Yu and S. Boseck, "Scanning acoustic microscopy and its applications to material characterization," Rev. Mod. Phys. **67**, 863–891 (1995).
[2] R. S. Gilmore, "Industrial ultrasonic imaging and microscopy," J. Phys. D **29**, 1389–1417 (1996).
[3] R. Puers and A. Cozma, "Bonding wafers with sodium silicate solution," J. Micromech. Microeng. **7**, 114–117 (1997).
[4] L. Wang, "The contrast mechanism of bond defects with the scanning acoustic microscopy," J. Acoust. Soc. Am. **104**, 2750–2755 (1998).
[5] L. Bechou, Y. Ousten, B. Tregon, F. Marc, Y. Danto, R. Even, and P. Kertesz, "Ultrasonic images interpretation improvement for microassembling technologies characterization," Microelectron. Reliab. **37**, 1787–1790 (1997).
[6] Y. Ogura, "High-sensitivity detection of voids at direct wafer bonding interface by ultrasonic imaging method," Oyo Butsuri **66**, 467–471 (1997).
[7] S. Canumalla, "Resolution of broadband transducers in acoustic microscopy of encapsulated ICs: Transducer selection," IEEE Trans. Compon. Packag. Technol. **22**, 582–592 (1999).
[8] D. K. Hsu and V. Dayal, "Ultrasonic Newton's rings," Appl. Phys. Lett. **60**, 1169–1171 (1992).
[9] H. Tohmyoh, M. Saka, and H. Hirayama, "Potential of acoustic imaging in the detection of nanometer gaps," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 2481–2483 (2006).
[10] H. Tohmyoh, "Polymer acoustic matching layer for broadband ultrasonic applications," J. Acoust. Soc. Am. **120**, 31–34 (2006).
[11] S. Timoshenko, "Circular plate concentrically loaded," *Strength of Materials, Part II: Advanced Theory and Problems* (Van Nostrand, New York, 1930), pp. 500–502.
[12] J. J. Wortman and R. A. Evans, "Young's modulus, shear modulus, and Poisson's ratio in silicon and germanium," J. Appl. Phys. **36**, 153–156 (1965).
[13] H. J. McSkimin, W. L. Bond, E. Buehler, and G. K. Teal, "Measurement of the elastic constants of silicon single crystals and their thermal coefficients," Phys. Rev. **83**, 1080 (1951).
[14] J. Krautkrämer and H. Krautkrämer, "Plane sound waves at boundaries," *Ultrasonic Testing of Materials*, 4th ed. (Springer-Verlag, Berlin, Germany, 1990), pp. 18–23.

Lamb wave characterization of corrosion-thinning in aircraft stringers: Experiment and three-dimensional simulation

Jill Bingham and Mark Hinders[a)]
Department of Applied Science, The College of William & Mary in Virginia, NDE Lab @ 116 Jamestown Road, Williamsburg, Virginia 23187-8795

(Received 22 January 2009; revised 14 April 2009; accepted 17 April 2009)

The development of automatic guided wave interpretation for detecting corrosion in aluminum aircraft structural stringers is described. The dynamic wavelet fingerprint technique (DWFT) is used to render the guided wave mode information in two-dimensional binary images. Automatic algorithms then extract DWFT features that correspond to the distorted arrival times of the guided wave modes of interest, which give insight into changes of the structure in the propagation path. To better understand how the guided wave modes propagate through real structures, parallel-processing elastic wave simulations using the finite integration technique (EFIT) has been performed. Three-dimensional (3D) simulations are used to examine models too complex for analytical solutions. They produce informative visualizations of the guided wave modes in the structures and mimic the output from sensors placed in the simulation space. Using the previously developed mode extraction algorithms, the 3D EFIT results are compared directly to their experimental counterparts. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3132505]

PACS number(s): 43.35.Zc, 43.40.Le, 43.60.Hj [TDM] Pages: 103–113

I. INTRODUCTION

Structural health monitoring (SHM) combines the use of onboard sensors with artificial intelligence algorithms to automatically identify and monitor structural health issues. Because they propagate relatively long distances within structures, guided waves allow inspection of large areas with fewer sensors, making them attractive for a variety of applications. Ultrasonic guided wave SHM methods push us to fully exploit the interaction of complex multimode signals with flaws of interest in key structural components.

This paper describes the experimental development of automatic guided wave interpretation for detecting the thinning effects of corrosion on aluminum aircraft structural stringers. The signals received are too complex for interpretation without knowledge of the guided wave physics. We employ a signal processing technique called the dynamic wavelet fingerprint technique (DFWT) in order to render the guided wave mode information in two-dimensional binary images. The use of wavelets allows us to keep track of both time and scale features from the original signals. With simple image processing, we have developed automatic extraction algorithms for features that correspond to the arrival times of the guided wave modes of interest for each of the applications. Due to the dispersive nature of the guided wave modes, the mode arrival times give details of the structure in the propagation path. For further understanding of how the guided wave modes propagate through the real structures, we have developed parallel processing, three-dimensional (3D) elastic wave simulations using the finite integration technique (EFIT). This full field, numeric simulation technique easily examines models too complex for analytical solutions. We have developed the algorithm to handle built up 3D structures as well as layers with different material properties and surface detail. The simulations produce informative visualizations of the guided wave modes in the structures but also directly mimic output from sensors placed in the simulation space for direct comparison to experiment. Using our previously developed mode extraction algorithms, we were then able to compare our 3D EFIT data to their experimental counterparts with consistency.

The basic formulation and discussion of the guided waves can be found in several now classic texts,[1–5] as well as extensive literature reviews of work done in the field of guided elastic waves for application in SHM.[6,7] The work presented here builds on our recent line of research[8–24] with the formalism and basic techniques dating back to 1885.[25–29] In extended structures where the wavelength is on the same order as the thickness, the two boundaries cause multiple reflections and mode conversions, so the "plate" develops new wave packets that propagate throughout the thickness. There are an infinite number of modes generated for higher frequency-thickness products in even the simplest of plates, although each of the modes except the two lowest has a cutoff frequency where the phase velocity approaches infinity while the group velocity approaches zero. The number of propagating modes present in the structure is determined by the frequency-thickness product and the way in which the waves are generated. The dispersive properties of the generated modes allow us to gain information about the structure through which they propagate. With a known nominal thickness, we can choose an excitation frequency that generates a highly dispersive mode and then exploit the dispersive nature of the guided waves to design SHM methods to monitor structures for deviations from this, i.e., flaws.

Our philosophy is to excite complicated signals in order to keep all of the time-series information, then post-process

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

the waveforms to extract the most useful information about the modes of interest using the DWFT. The DWFT implemented here relies on filtering the data with a discrete stationary wavelet (SWT) filter to remove a few layers of detail then passing the filtered signal through the fingerprinting algorithm.

Wavelets are very useful for analyzing time-series data because the wavelet transform allows us to keep track of both time and frequency, or scale features. Whereas Fourier transforms break down a signal into a series of sines and cosines in order to identify the frequency content of the entire signal, wavelet transforms keep track of local frequency features in the time domain. Ultrasonic signal analysis with wavelet transforms was first studied by Abbate in 1994 who found that if the mother wavelet was well defined there was good peak detection even with large amounts of added white noise.[30] Massicotte *et al.* then found that even noisy EMAT sensor signals were resolvable using the multi-scale method of the wavelet transform.[31] One of the strengths compared to the fast Fourier transform was that the extraction algorithm did not need to include the inverse transform, the arrival time could be taken directly from the time frequency domain of the wavelet transform. In 2002 Perov *et al.*[32] considered the basic principles of the formulation of the wavelet transform for the purpose of an ultrasonic flaw detector and concluded that any of the known systems of orthogonal wavelets are suitable for this purpose as long as the number of levels does not drop below 4–5. In 2003 Lou and Hu found that the wavelet transform was useful in suppressing non-stationary wideband noise from speech.[33] In a comparison study between the Wigner–Ville distribution and the wavelet transform, performed by Zou and Chen, the wavelet transform outperformed the Wigner–Ville in terms of sensitivity to the change in stiffness of a cracked rotor.[34] In 2002 Hou and Hinders developed a multi-mode arrival time extraction tool that rendered the time-series data in two dimensional (2D) time-scale binary images.[35] Since then this technique has been applied to multi-mode extraction of Lamb wave signals for tomographic reconstruction,[17,36] time domain reflectometry signals for wiring flaw detection,[37,38] and a periodontal probing device.[39]

In general most of the information in a signal is contained in the approximations of the first few levels of the SWT. The details of these low levels often have mostly high frequency noise information. If we set the details of these first few levels to zero, when we reconstruct the signal with the inverse SWT we have effectively de-noised our signal to keep just information of the Lamb wave modes of interest. In our work, we start with the filtered ultrasonic signal and take a continuous wavelet transform. The continuous wavelet transform gives a surface of wavelet coefficients, and this surface is then normalized between [0–1]. Then we perform a thick contour slice operation where the user defines the number of slices to use; the more slices, the thinner the contour slice. The contour slices are given the value of 0 or 1 in alternating fashion. They are then projected down to a 2D image where the result often looks remarkably like the ridges of a human fingerprint. The problem has thus been transformed from one-dimensional signal identification problem to a 2D image recognition scenario. The power of the DWFT is that it reduces the time-series data into a binary matrix that is easily stored and transferred. There is also a robustness to the algorithm, since different mother wavelets emphasize different features in the signals. For the most part in the research, we have manually chosen the mother wavelet based on experience and using wavelets roughly shaped like the excitation pulses. The last piece of the DWFT is the image recognition of the binary features that correspond to the modes of interest. We have found that different modes are represented in unique features in our applications. We have found that using a simple ridge counting algorithm on the 2D images is often a helpful way to identify some of the features of interest. Once a feature has been identified in the time scale space, we have determined its arrival in the time domain as well and we can draw conclusions based on our knowledge of the guided wave theory, supplemented by high-resolution 3D simulations.

Along with finite element techniques, numeric modeling using the EFIT has proven very useful for modeling guided wave behavior. Fellinger *et al.*,[40] developed the basic equations of EFIT along with a way to discretize the material parameters for ensuring continuity of stress and displacement across the staggered grid in 1995. Schubert *et al.*[41] then adapted the EFIT equations into cylindrical coordinates (CEFIT) to investigate axisymmetric wave propagation in pipes with a 2D grid. In 2001 Schubert and Koehler[42] presented results looking at elastic wave propagation in porous concrete but due to computational limitations could only model $5\times5\times10$ cm^3 spaces with periodic boundary conditions. Then in 2004 Schubert[43] gave an overview of the flexibility of EFIT with discretization in Cartesian, cylindrical, and spherical coordinates and showed a wide range of modeling applications. Rudd *et al.* extended this to a full 3D massively parallel implementation to model phased array focusing after pipe bends[44] ultrasonographic periodontal probing[45] and non-linear acoustic security screening.[46]

Our motivation is to identify problem areas in structures before failure occurs by developing techniques using ultrasonic guided waves to provide quantitative information about the structure. Most of the development is carried out experimentally in the lab, using the DWFT to present the complex data in a form that is easier to interpret. Our specific goal is to better understand Lamb wave propagation in airframe stringers and their interaction with corrosion and thickness loss flaws. First, using an incremental thickness loss experiment and then an accelerated corrosion test, we employ the DWFT to automatically extract mode arrivals. We then perform high-resolution 3D simulations of Lamb wave propagation in the stringers and compare directly to the experimental results.

The stringer samples provided by Alcoa are made of a high-strength aluminum alloy, Al 2024 T3511, in which copper is the major alloying element along with a small magnesium content. The approximately 4% copper and 1.5% magnesium provide increased strength and work-hardening characteristics. However, due to the higher copper content, this alloy is less resistant to corrosion. Minute copper particles on the surface and grain boundaries of the alloy create

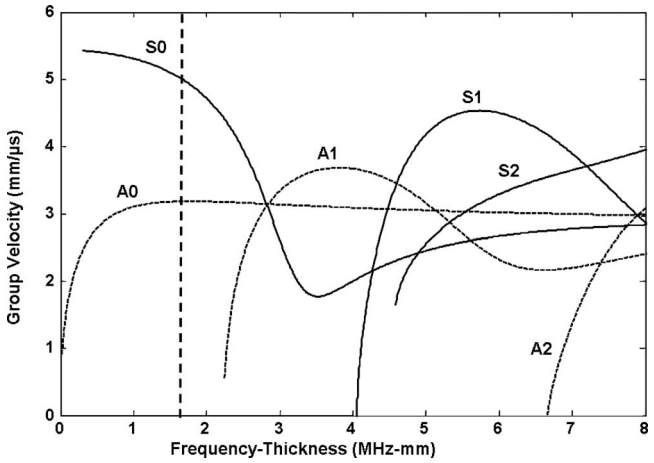


FIG. 1. Aluminum dispersion curve. For a frequency-thickness product of 1.6 mm MHz (dotted line), we expect the arrival of two guided wave modes, *S*0 and *A*0.

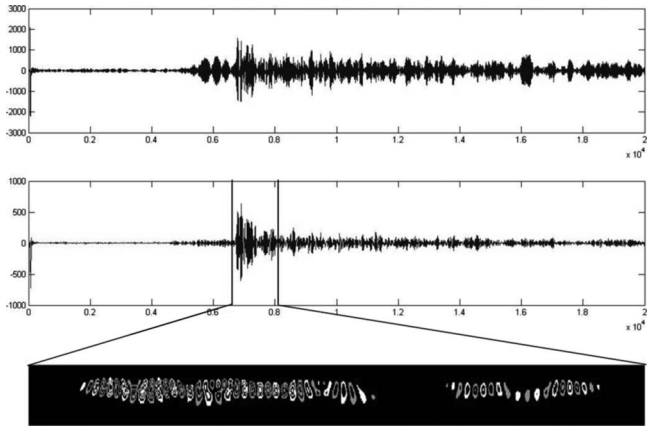


FIG. 2. (Top) Raw waveform collected from clean, un-milled sample. (Middle) Filtered waveform using a DSW filter. (Bottom) Extracted wavelet fingerprint for the *A*0 mode, corresponding to the gated portion of the filtered waveform. The horizontal axis in each is time. For the waveforms the vertical axis is amplitude, and for the wavelet fingerprint the vertical axis is scale.

small galvanic cells in the material. The copper is more noble than the aluminum so it has potential to reduce the adjacent aluminum, causing corrosion. As this intergranular corrosion continues, the copper ions replate themselves on the alloy increasing the corrosion rate. This is why the temper, T3511, and work-hardening are important; the heat treatment can affect the amount, size, and distribution of the intermetallic precipitates.[47] In extruded structures, the grains become elongated down the length of the structure so the intergranular corrosion proceeds in an exfoliating manner. The grain boundaries expand, flaking off mostly noncorroded layers in a fashion that we can model as a thinning of the plate-like structure. The sample aluminum stringers used are 1 m in length and have a "T" cross-section with an original flange thickness of 1.6 mm. We used piezoelectric shear wave contact transducers on the flange in a pitch-catch arrangement to inspect the samples, with the Lamb waves propagating from one end of the stringer to the other primarily in the flange. The transmitting transducer, aligned in the shear vertical (SV) configuration where the polarization of the PZT crystal is parallel to the length of the stringer, excites Lamb wave modes that are then recorded by the receiving transducers at the far end of the stringer. Since the Lamb wave modes are dispersive, the presence of a flaw shifts the arrival times and amplitudes of the Lamb wave modes received. For thinning flaws, we expect that the faster *S*0 mode will speed up while the slower *A*0 mode will gradually slow down with increased thinning (Fig. 1). One of the advantages of guided waves for SHM is that they are sensitive to corrosion on either side of the structure, which here means that the transducer can be placed on whichever side of the flange is accessible and then corrosion can be detected on both that side and on the inaccessible face of the structural member.

II. INCREMENTAL THICKNESS MILLING TEST

Our first step consisted of simulating the effect of corrosion by incrementally decreasing the thickness of the stringer flange with a milling machine so that we know precisely how much material we are removing at each step. The milling increments were taken from the middle 40 cm of one side of the flange, from the outside edge 21.5 mm toward the web. The transducers were placed 10 cm from the ends of the stringers to separate the reflections from the ends out in time from the direct signal. Modeling clay on the ends of the stringer was used to damp out some of the standing waves and to keep the stringer from touching the laboratory bench. The total propagation length was 80 cm with the 40 cm flaw in the middle. We show the data processing steps in Fig. 2. This progression begins with a SWT filter, and then by windowing a portion to pass to the DWFT algorithm to find the fingerprints. The window was determined using the expected arrival time for the *A*0 mode. Figure 3 shows the binary fingerprint images produced from the incremental milling tests. The top image is for the un-milled clean sample. In each of the images, a circle-in-circle feature was automatically extracted via straightforward image processing which searched at each time delay for the corresponding vertical ridge-count pattern, with the feature's location indicated by the vertical lines. As material is removed from the flange, we would expect from the dispersion curves that the *A*0 mode would slow down. In these images a slowing of the mode would mean a later arrival time and a shift of the feature to the right. It can be seen in the images that the feature shifts right. Furthermore the movement of the double feature corresponds to the arrival time of the *A*0 mode. The extracted arrival times are labeled to the left. If we calculate the expected arrival times using the mode velocities from the dispersion curves (Fig. 1), taking into account that the waves are only traveling through the thinned region for half of the propagation length, we can compare to our experimentally extracted arrival times and we find that they match the expected arrival times. The arrival times change as the thickness changes in the manner expected from the dispersion curve predictions.

III. ACCELERATED CORROSION TEST

To keep in accordance with the incremental milling tests, we masked off the test section leaving the middle

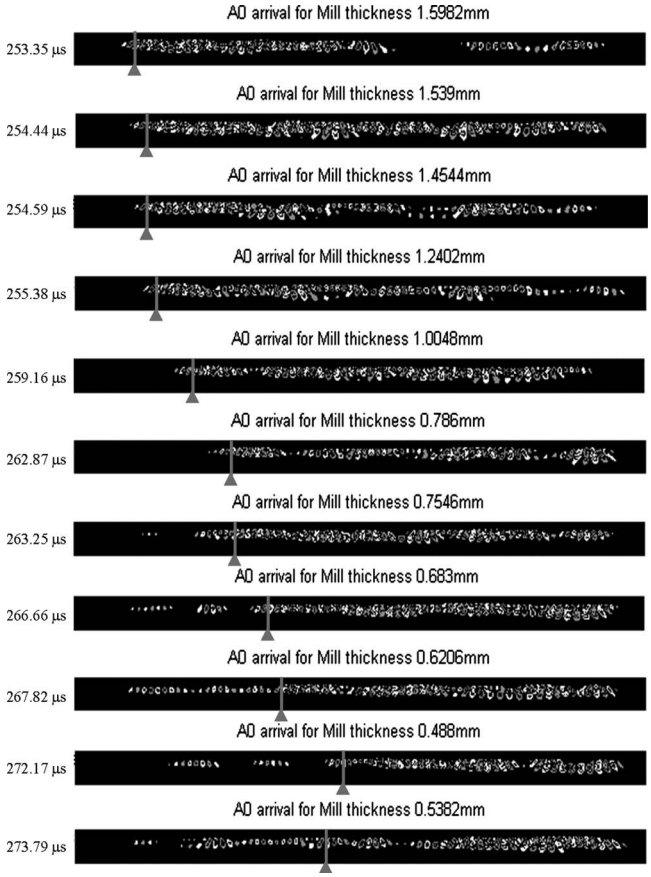


FIG. 3. Material was removed incrementally from top to bottom. Fingerprints show the automatically extracted *A*0 arrival, depicted by the lines. That arrival time is noted to the left of each.

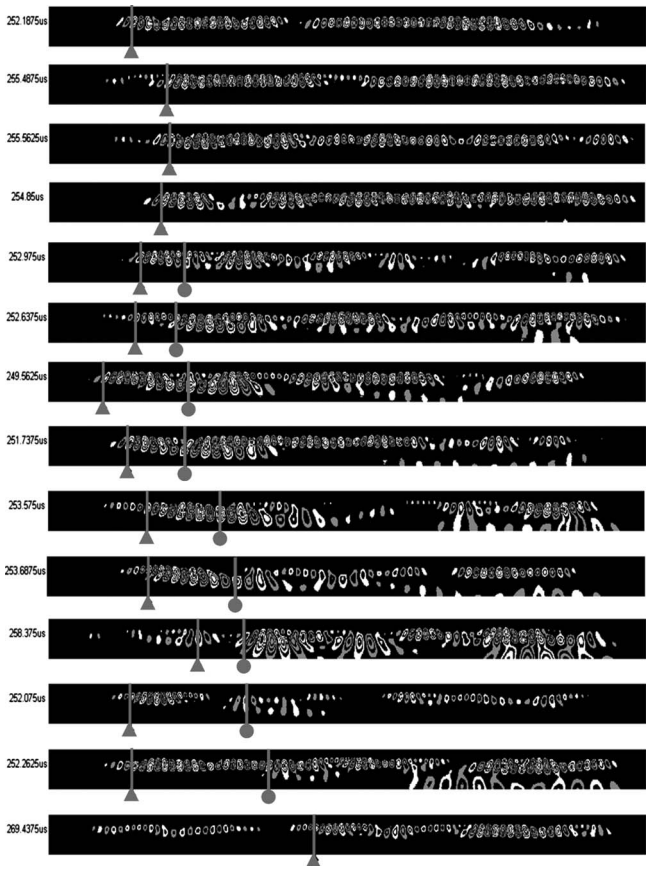


FIG. 4. 12 h increments of accelerated corrosion EXCO test from top to bottom represented in thumbprints obtained by the DWFT. The vertical lines indicate the automatic extraction of a double circular feature from the thumbprints. The second lines in several of the thumbprints indicate a second feature of interest that appears to split off from the first with increased corrosion.

40 cm of the flange to be corroded. The accelerated corrosion test follows an ASTM standard test method known as the EXCO test.[48] This test is especially designed to corrode high-strength 2XXX and 7XXX series aluminum alloys. The solution is made up of sodium chloride, potassium nitrate, nitric acid, and hydrogen peroxide for added acceleration. The EXCO solution produced an exfoliation form of corrosion which flakes off layers of the exposed surface. Our procedure was to record baseline waveforms, apply the EXCO solution to the test area, and then collect data every 12 h. Data collection consisted of recording a waveform with the EXCO solution still on the surface, and then putting the sample in a nitric acid bath to remove corrosion products collecting another waveform before taking multiple thickness measurements. After data collection for a particular timestep, we reapplied the EXCO solution to sit for another 12 h.

The recording and processing of the guided wave signals were the same as for the milling tests. The contact transducers were placed 10 cm from each of the ends of the 1 m T stringer. However, the raw waveforms from these tests were much more complex than from the incremental milling tests. This is because of the nature of the exfoliation, each of the flakes becomes a scatterer of the elastic wave energy. We recorded Lamb waveforms from both the fluid loaded stringers as well as after the rinse. Once again from the raw time-series signals, we could not directly extract useful mode information that tells the extent of the thickness thinning corrosion. Using a window from the SWT filtered signal, we can employ the DFWT developed with the incremental milling test to identify arrival information of the *A*0 mode. The resulting thumbprints are similar to those from the previous tests. Figure 4 shows the progression of the accelerated corrosion in the thumbprints. Indicated by the vertical lines, we see that the automatic extraction of the same double circular feature as was found in the milling tests shift around an arrival of 250 μs. It is also interesting that after the first few 12 h corrosion shifts, we see a split of the features that make up the *A*0 mode. The second lines in several of the thumbprints in Fig. 4 are placed on the first feature that goes from a left leaning inclination to the right, signifying the start of another possible mode. There is a split between the two feature extractions. The first lines stay about at the arrival time of the original thickness while the second lines slow down as expected for thickness loss due to corrosion.

In order to determine if there is actually a splitting of the *A*0 mode, we have to understand more fully how the wave is propagating in this built up T structure better. This structure is more complicated than just the theoretical plate model that is used to compute the dispersion curves used in this section for the approximate expected arrival times for the modes.[6,7] Here we not only have the reflections from the top and bot-

tom surfaces of the flange, but its width and the presence of the web also affect how the elastic energy propagates along the stringer.

IV. EFIT SIMULATION METHOD

The EFIT evolves from the basic wave equations for elastic solids.[40] We start with Hooke's law and Cauchy's equation of motion to give the fundamental equations. The differential form of the equation of motion

$$\rho \dot{v}_x = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z} + f_x, \tag{1}$$

$$\rho \dot{v}_y = \frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} + f_y, \tag{2}$$

$$\rho \dot{v}_z = \frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} + f_z, \tag{3}$$

and the first time derivative of Hooke's law in differential form

$$\dot{\sigma}_{ij} = \lambda \dot{\epsilon}_{kk} \delta_{ij} + 2\mu \dot{\epsilon}_{ij}, \quad i,j = x,y,z, \tag{4}$$

where we sum over the repeated index k and

$$\dot{\epsilon}_{xx} = \frac{\partial v_x}{\partial x}, \quad \dot{\epsilon}_{yy} = \frac{\partial v_y}{\partial x}, \quad \dot{\epsilon}_{zz} = \frac{\partial v_z}{\partial z},$$

$$\dot{\epsilon}_{xy} = \frac{1}{2}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right), \quad \dot{\epsilon}_{xz} = \frac{1}{2}\left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z}\right),$$

$$\dot{\epsilon}_{yz} = \frac{1}{2}\left(\frac{\partial v_z}{\partial y} + \frac{\partial v_y}{\partial z}\right)$$

give the components of the velocity vector, v_i and the stress tensor, σ_{ij}, for a particle in the elastic solid. Here, the material is defined by the Lamé constants λ and μ and the material density ρ. The source terms are represented by f_x, f_y, and f_z in the velocity terms.

With the displacement and stress of a volume element specified in an elastic solid, we then consider the solid as many such cells next to each other[40] and place the velocity components on the edges of each cell with the diagonal elements of the stress tensor on the corners and the off-diagonal elements on the faces of the cells. Using this component placement we can then discretize Eqs. (1)–(4) as follows:

$$\rho \dot{v}_x^{(n)}(t) = \frac{\sigma_{xx}^{(n+\hat{x})}(t) - \sigma_{xx}^{(n)}(t)}{\Delta x} + \frac{\sigma_{xy}^{(n)}(t) - \sigma_{xy}^{(n-\hat{y})}(t)}{\Delta y} + \frac{\sigma_{xz}^{(n)}(t) - \sigma_{xz}^{(n-\hat{z})}(t)}{\Delta z} + f_x(t),$$

$$\rho \dot{v}_y^{(n)}(t) = \frac{\sigma_{xy}^{(n)}(t) - \sigma_{xy}^{(n-\hat{x})}(t)}{\Delta x} + \frac{\sigma_{yy}^{(n+\hat{y})}(t) - \sigma_{yy}^{(n)}(t)}{\Delta y} + \frac{\sigma_{yz}^{(n)}(t) - \sigma_{yz}^{(n-\hat{z})}(t)}{\Delta z} + f_y(t),$$

$$\rho \dot{v}_z^{(n)}(t) = \frac{\sigma_{xz}^{(n)}(t) - \sigma_{xz}^{(n-\hat{x})}(t)}{\Delta x} + \frac{\sigma_{yz}^{(n)}(t) - \sigma_{yz}^{(n-\hat{y})}(t)}{\Delta y} + \frac{\sigma_{zz}^{(n+\hat{z})}(t) - \sigma_{zz}^{(n)}(t)}{\Delta z} + f_z(t),$$

$$\dot{\sigma}_{xx}^{(n)}(t) = (\lambda + 2\mu)\frac{v_x^{(n)}(t) - v_x^{(n-\hat{x})}(t)}{\Delta x} + \lambda\left(\frac{v_y^{(n)}(t) - v_y^{(n-\hat{y})}(t)}{\Delta y} + \frac{v_z^{(n)}(t) - v_z^{(n-\hat{z})}(t)}{\Delta z}\right),$$

$$\dot{\sigma}_{yy}^{(n)}(t) = (\lambda + 2\mu)\frac{v_y^{(n)}(t) - v_y^{(n-\hat{y})}(t)}{\Delta y} + \lambda\left(\frac{v_x^{(n)}(t) - v_x^{(n-\hat{x})}(t)}{\Delta x} + \frac{v_z^{(n)}(t) - v_z^{(n-\hat{z})}(t)}{\Delta z}\right), \tag{5}$$

$$\dot{\sigma}_{zz}^{(n)}(t) = (\lambda + 2\mu)\frac{v_z^{(n)}(t) - v_z^{(n-\hat{z})}(t)}{\Delta z} + \lambda\left(\frac{v_x^{(n)}(t) - v_x^{(n-\hat{x})}(t)}{\Delta x} + \frac{v_y^{(n)}(t) - v_y^{(n-\hat{y})}(t)}{\Delta y}\right),$$

$$\dot{\sigma}_{xy}^{(n)}(t) = \mu\left(\frac{v_x^{(n+\hat{y})} - v_x^{(n)}}{\Delta y} + \frac{v_y^{(n+\hat{x})} - v_y^{(n)}}{\Delta x}\right),$$

$$\dot{\sigma}_{xz}^{(n)}(t) = \mu\left(\frac{v_x^{(n+\hat{z})} - v_x^{(n)}}{\Delta z} + \frac{v_z^{(n+\hat{x})} - v_z^{(n)}}{\Delta x}\right),$$

$$\dot{\sigma}_{yz}^{(n)}(t) = \mu\left(\frac{v_y^{(n+\hat{z})} - v_y^{(n)}}{\Delta z} + \frac{v_z^{(n+\hat{y})} - v_z^{(n)}}{\Delta y}\right).$$

Given that $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit steps in the x, y, and z directions, respectively, and n denotes the current cell.

We have to apply stress free boundary conditions on our simulation boundaries, since our interest is in guided elastic waves in structures at megahertz frequencies where the solid-air interfaces can be considered traction-free surfaces. This means that on the x boundaries σ_{xx}, σ_{xy}, and σ_{xz} all vanish while on y and z boundaries σ_{yy}, σ_{xy}, and σ_{yz} and σ_{zz}, σ_{xz}, σ_{yz} are zero, respectively. If we require the velocity components to be placed on the physical surface of our model, we find that the shear stress terms are also on the surface so we can set them to zero.[41] To ensure that the longitudinal stress at the surface vanishes, we set $\sigma_{ii}^{(\mathrm{surf})} = -\sigma_{ii}^{(\mathrm{surf}+\hat{i})}$ for a lower boundary and $\sigma_{ii}^{(\mathrm{surf}+\hat{i})} = -\sigma_{ii}^{(\mathrm{surf})}$ for an upper boundary.[44] This produces the equations for the surface velocity components for a lower boundary

$$\rho \dot{v}_i^{(\mathrm{surf})} = \frac{2\sigma_{ii}^{(\mathrm{surf}+\hat{i})}}{\Delta s} + f_i \tag{6}$$

and for an upper boundary

$$\rho \dot{v}_i^{(\mathrm{surf})} = -\frac{2\sigma_{ii}^{(\mathrm{surf})}}{\Delta s} + f_i \tag{7}$$

for $i = x,y,z$.

The temporal discretization is based on a central difference operator, "leap-frogging" through time across the staggered grid,

$$v_i^{[k]} = v_i^{[k-1]} + \dot{v}_i^{[k-1/2]}\Delta t, \tag{8}$$

$$\sigma_{ij}^{[k+1/2]} = \sigma_{ij}^{[k-1/2]} + \dot{\sigma}_{ij}^{[k]}\Delta t, \tag{9}$$

where the superscript k denotes full and $k \pm 1/2$ denotes half-time steps of Δt with the total time given by $T=k\Delta t$. First, the velocity components are updated, and then we update the stress tensor components using the previously updated velocity components.

In order for the 3D-EFIT algorithm to be numerically stable, we must satisfy the Courant–Friedrichs–Levy criterion:

$$\Delta t \leq \frac{1}{c_l\sqrt{1/(\Delta x)^2 + 1/(\Delta y)^2 + 1/(\Delta z)^2}}, \tag{10}$$

where c_l is the fastest longitudinal wave speed in the elastic medium. For simplicity, we use equal lengths for the sides of the cells, Δs, so Eq. (10) becomes

$$\Delta t \leq \frac{1}{c_l\sqrt{3/(\Delta s)^2}}. \tag{11}$$

Here Δs is determined such that the shortest wavelengths present are adequately discretized using

$$\Delta s \leq \frac{1}{8}\lambda_{\min} = \frac{1}{8}\frac{c_{\min}}{f_{\max}} \approx \frac{1}{10}\frac{c_{s,\min}}{f_{\max}}. \tag{12}$$

This assigns approximately ten grid points per shear wavelength, adjusting the exact size to correspond to the desired space thickness.

In this work we are dealing with thin plate-like structures so our simulation space thickness is small compared to the length and width. In order to inspect materials for changes in thickness, we have to use signals that have wavelengths on the same order as the thickness of our sample. This means that when dealing with a sample that is about 1 mm thick we have to be in the 1 MHz frequency range, resulting on the order of ten grid points per millimeter. If we set out to model one of the aircraft stringers samples, we use the Δs for aluminum of 0.000 145 m.

We split our EFIT simulation space across multiple processors using a typical 2D domain decomposition.[44] This domain decomposition consists of taking the 3D Cartesian simulation space and slicing it in the xz plane, and then again in the yz plane. This results in having the entire thickness of the simulation intact on each processor, allowing the algorithm to loop through the thickness without having to account for an edge of the processor in this direction. In the x and y dimensions, however, we do need to pay attention to the processor boundaries. For each half-time step, the cells on the processor edges need information from their neighbors in order to update themselves according to Eq. (5). Each processor updates the velocity components, then sends the edge velocity components to its neighbors before updating the stress components, and then it sends the edge stress components before increasing the time step.

Looking at Eq. (5), we see that the only material parameters that are present are the density, ρ, and the Lamé constants, λ and μ. For solid elastic materials, the interface with air is well approximated by a solid-vacuum interface so we can set the density of cells not in our test material to zero (or a large negative number in practice for convenience). In this fashion we can also include multiple materials; however, we have to take care to ensure continuity of stress and velocity across cell surfaces. If we choose the material parameter cell to coincide with the σ_{ii} integration cell, we can average the material parameters ρ and μ for use in Eq. (5).[42]

$$\begin{aligned}
\rho^{(n,x)} &= \frac{\rho^{(n)} + \rho^{(n+\hat{i})}}{2}, \\
\rho^{(n,y)} &= \frac{\rho^{(n)} + \rho^{(n+\hat{j})}}{2}, \\
\rho^{(n,z)} &= \frac{\rho^{(n)} + \rho^{(n+\hat{k})}}{2}, \\
\mu^{(n,xy)} &= \frac{4}{\frac{1}{\mu^{(n)}} + \frac{1}{\mu^{(n+i)}} + \frac{1}{\mu^{(n+j)}} + \frac{1}{\mu^{(n+i+j)}}}, \\
\mu^{(n,xz)} &= \frac{4}{\frac{1}{\mu^{(n)}} + \frac{1}{\mu^{(n+i)}} + \frac{1}{\mu^{(n+k)}} + \frac{1}{\mu^{(n+i+k)}}}, \\
\mu^{(n,yz)} &= \frac{4}{\frac{1}{\mu^{(n)}} + \frac{1}{\mu^{(n+j)}} + \frac{1}{\mu^{(n+k)}} + \frac{1}{\mu^{(n+j+k)}}}.
\end{aligned} \tag{13}$$

For layered materials we still update the model by rectangular volumes, but if two regions next to each other have different densities we change the material parameters at the interface.

Another flexibility of our EFIT code is that we can either apply boundary conditions to an entire flaw region or keep track of the density and Lamé material parameters cell by cell and step through the space applying the needed boundary conditions. This allows us to insert surfaces into our models. One of the main difficulties of adding a surface to the model is keeping in mind the stability criterion so that the EFIT model still behaves properly with the rough surface. This is achieved either by smoothing the surface somewhat or by reducing the cell size in the model. This is the final piece in the development of our 3D EFIT package that we need to simulate guided elastic wave propagation in aircraft stringers.

V. AIRCRAFT STRINGER SIMULATIONS

First, we simulate guided waves propagating down a plate which has dimensions identical to the flange of the experimental T stringer. We use a five-cycle sine wave, with SV excitation of a 1 MHz, 0.5 in.^{2} contact transducer for the

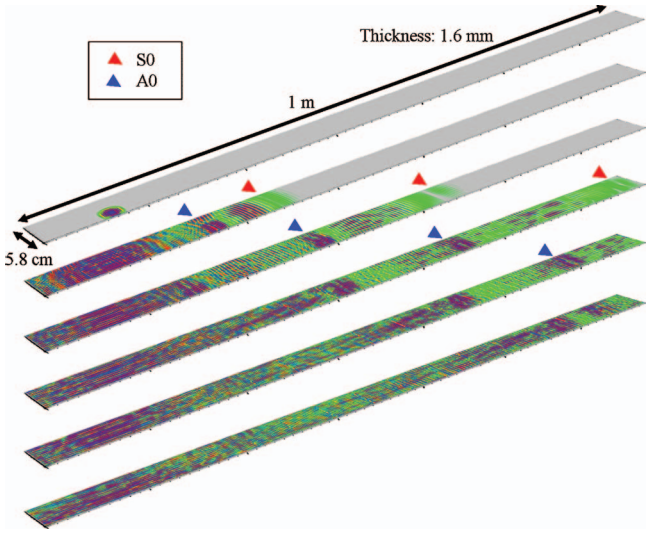


FIG. 5. Propagation of the Lamb waves as modeled by the 3D EFIT simulation. 1 MHz transducers were placed 10 cm from each end. For each of the frames the *S*0 and *A*0 modes are indicated by a red and blue triangle, respectively. Reflected modes from the ends of the sample are also seen present. Near the transmitter the finite width of the plate causes lateral reverberation which complicates the signal, but as the Lamb wave modes propagate down the length of the structure, these finite-width effects diminish and the Lamb wave modes propagate largely as expected from 2D plain-strain analyses.

$f_x(t)$ source term in Eq. (5), placed 10 cm from $x=0$ and 1 cm from $y=0$ on the surface of the simulated plate.

In Fig. 5, we can see that the *S*0 and *A*0 modes separate out as they propagate down the length of the flange. We even see the reflections off the end of the test piece that is nearer the transmitting transducer. When we include the web of the T stringer, we do still see the propagation of the expected guided waves in Fig. 6. One of the interesting features from this simulation is that the web of the stringer seems to damp out the *S*0 mode so that by the time it propagates the length of the stringer it is barely visible. This is consistent with the findings from our experimental tests where we were expecting to see a change in the *S*0 arrival times as well as the *A*0, but we had difficulty extracting the arrival information for the *S*0 mode.

We next apply the same DWFT algorithm to these new simulation results and are able to extract similar features from the simulation data as we could for the experimental data, see Fig. 7. In the EFIT simulations, it seems that the extracted arrival time is consistently 9 μs after the expected arrival from the dispersion curves, as well $\sim$6 μs after the experimental times. Other than this offset the structure of the thumbprints are quite similar between the simulations and the experimental data. This consistency gives us yet more confidence that we are modeling the guided waves correctly with our 3D EFIT.

We next simulated two steps in the milling experiment. In the simulation space, we masked off a 40×2.15 cm^2 rectangular area of the flange and left its thickness at 1, and 0.488 mm for two different runs. These two steps were chosen from the 11 steps in the milling test because they represent a midpoint and end of material loss from the experiment. Comparing the extracted arrival times for the *A*0 mode to those found experimentally in the controlled-milling tests, we can see in Fig. 8 that the same trend is present. As more material is removed from the flange, the *A*0 mode slows down more and more. We also see a consistent 9 μs delay in the extracted arrival time which is presumably due to not including transducer effects in the simulations. In the EFIT simulations, we force the motion of the surface grid points at the location of the transducer rather than including the transducer into the simulation space explicitly.

Although the EFIT simulation accurately models guided wave propagation for the 3D structure of the aircraft stringer, our goal is to better understand the propagation along a corroded surface. In order to simulate this, we need a thickness map of the corroded surface. To find this surface we performed an ultrasonic C-scan of the stringer, in pulse-echo mode with the test piece submerged in a water tank and using a focused 20 MHz transducer that raster scans across the corroded area. The corroded surface is mapped by gating the signal around the reflection of interest and recording the time-delay of this echo at each point in the scan.

Figure 9(a) shows the C-Scan map of the corrosion surface normalized in order to fit the cell size of our EFIT simulation. Recalling the stability criterion (12), we make sure that there are enough grid points per "ripple" in the surface. There are various ways to approach this: make the stepsize smaller, which increases the size of the simulation space adding computation time, memory usage, and disk storage space, or smooth out the surface somewhat to make sure that the ridges and bumps have enough lateral size to them. Figure 9(b) shows the final surface that we used in the simulation; we used a convolution averaging filter with a ten cell radius in order to obtain this smoothed surface.

This matrix that contains the surface was then mapped into the simulation space. This model was run on 100 processors, 25 in the x-direction and 4 in the y-direction. Since we know the step size and model dimensions for this simulation, we simply created individual files for the processors that each have a portion of the surface on them. The processors that do not contain a portion of the flaw are updated as usual. The ones that have the corrosion surface on them update around and under the flaw region as usual, and then they step cell by cell through the simulation space for the surface. The matrix for the surface gave a cell number for the top of the surface under which there is a density, above which the density is set to zero. Another array was then created holding a value for the boundaries of the cell. This number is 0 if it is in the interior of the solid, and has a 1 or 2 if it is on a boundary in a certain direction. The x-direction is determined by the hundreds place, the y by the tens, and z by the ones position. According to this boundary array, we apply the needed equations each time step for each cell. Figure 10 shows a snapshot of the guided wave propagation through the corroded region of the aircraft stringer. We see here that the *A*0 mode is much more distorted due to the corroded surface than is the *S*0 mode. This is because the through-thickness displacement profile of the *A*0 mode has the great-

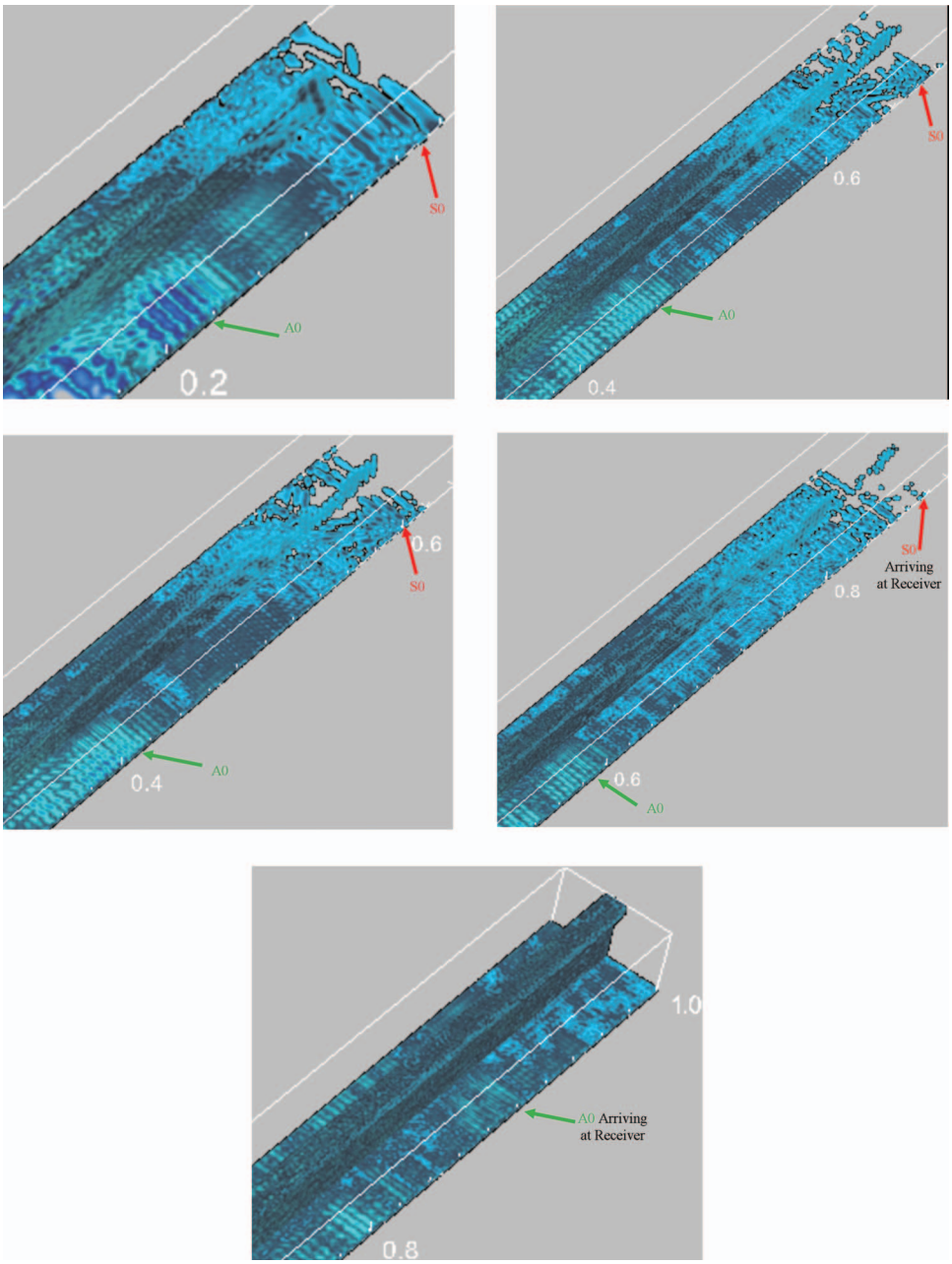


FIG. 6. Zoom of T stringer mode propagation for five time steps. The *A*0 and *S*0 Lamb wave modes separate out as they propagate due to their differing velocities, which helps to identify them even though they are no longer pure plate modes.

est magnitude at the surface of the plate, while the *S*0 is uniform through the center of the thickness.[4] This allows more of the *A*0 energy to be scattered in different directions by the uneven surface.

Considering the extracted arrival of the *A*0 mode from the corroded stringer simulation in the same fashion as before with the experimental data we again see consistency (Fig. 11). When we compare the thumbprints to those found in Fig. 4, we see the same features including an early doublet that is followed by the larger doublet feature corresponding to the *A*0 arrival for a thinned region. From the snapshot images, we can tell that the *A*0 wavefront is very much broken up by the corrosion surface, which could explain the splitting of the features that we discussed in Sec. III. In order to examine this in finer detail, we would need a much larger computer which we could track all of the displacement and stress components carefully instead of just recording the magnitude of the displacement vector. This would allow us to watch how the energy is scattered from the surface in great detail. Nevertheless, this simulation shows that the guided wave modes are still propagating under the corroded surface as expected.

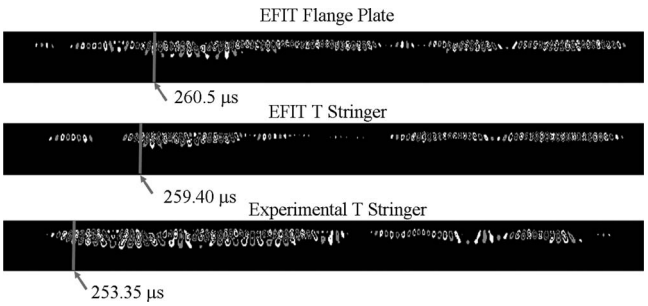


FIG. 7. Resulting thumbprints from experiment and simulation.

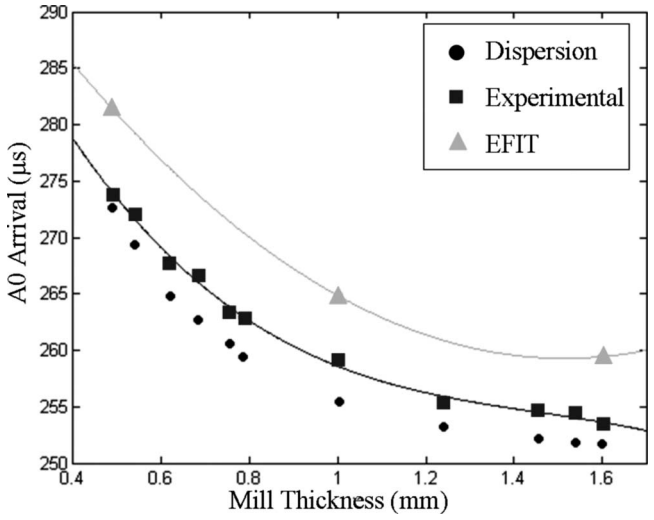


FIG. 8. Incremental milling test comparison of *A*0 arrival times for T-stiffener. The dots show the expected arrival time for each thickness step derived from the dispersion curves. The arrival times extracted via the DFWT thumbprints for experiment and simulated data are shown as well, all of which were automatically extracted using the same algorithm. Spline fits are shown to highlight trends.

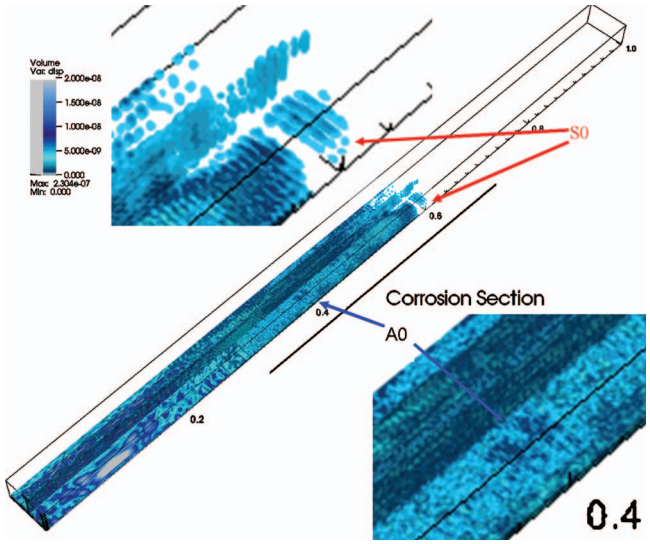


FIG. 10. Snapshot of EFIT propagation through corrosion surface. The *A*0 mode is much more distorted by the corroded surface than the *S*0.

VI. DISCUSSION AND FUTURE WORK

SHM is a family of emerging technologies intended to address both the international fleet of aging aircraft and optimum efficiency in new designs. The reliability of maintenance depends on whether or not we can accurately diagnose problem areas fast and cheaply. Repairing aircraft is a challenge because some problem areas are concealed inside the skin of the aircraft. It is wasteful to take apart sections of the airplane just to find that they do not need repair. New construction also has its own issues with maintenance and repair which arise when deciding which materials to use. Composites are strong and lightweight and as such give competition to aluminum airframes. Hence, aluminum companies must reduce the weight of their product in order to still be attractive to aircraft manufactures. Reducing the weight of the aluminum may translate to reducing the amount of aluminum that is used in the structure. We can place sensors to track structural integrity instead of over engineering the structure. Airframe stringers are one of the main structural components to which the outside skin is attached, and in key locations which are susceptible to corrosion an ultrasonic guided wave system may be ideal for monitoring large areas quickly without taking apart the structure.

Our tests of guided wave interaction with corrosion on aircraft structural stringers demonstrates robustness of the DWFT for identifying mode arrivals. Automatic extraction of the *A*0 mode with minimal prior knowledge of stringer state was shown to correspond to the thickness loss of a section of the stringer. Starting with the initial dimensions of the stringer, and knowledge of the guided wave mode behavior, we extracted guided wave arrivals automatically. For use in the field this could be implemented as an inspection technique where the extracted arrival time would tell the thickness of the stringer, deviation from the expected would be a flag for maintenance of that structure.

An interesting observation in the experimental work on the stringers was that it was difficult to extract the first arriving mode from the signals. Instead we used information from the later arriving *A*0 mode. In addition to confirming findings of the material-thickness loss causing a slowing of the *A*0 mode, the EFIT simulations shed light on the missing *S*0 waveform features. In the initial analysis, as with traditional guided wave analysis, the T stringer was approximated by a thin plate with the dimensions of the flange of the stringer. The EFIT simulation showed that this model, although correctly predicting the propagation of the *A*0 mode, is too simple to explain the *S*0 propagation. From the simulation visualization, we see that the *S*0 mode is attenuated by the web of the structure so much that by the time that mode

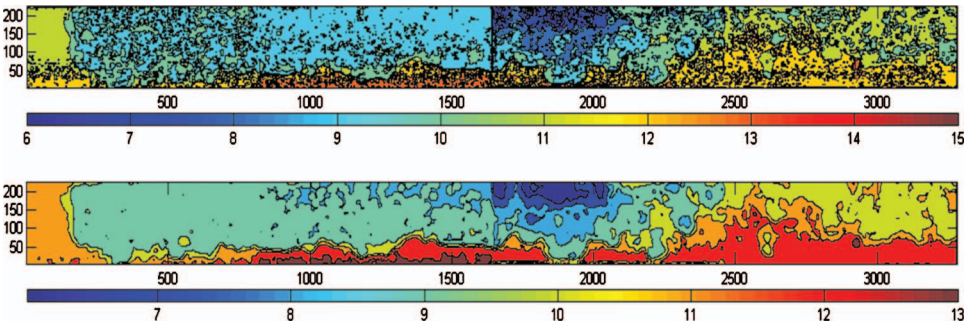


FIG. 9. C-Scan thickness map of the flange under the corrosion, with thickness or remaining material color coded. (Top) No smoothing. (Bottom) With a image convolution filter. Direction of Lamb wave propagation is left to right, corresponding to the simulations shown in the previous figures.

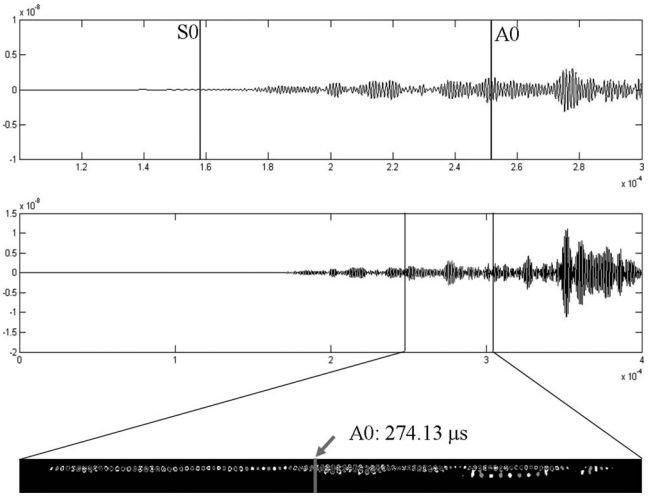


FIG. 11. Raw waveform, filtered signal, and resulting DWFT thumbprint for EFIT

propagates the stringer length it is not resolvable from the noise, which was the exact situation that we were seeing from the experimental data. Furthermore, the EFIT simulation accurately modeled the effect of the corroded surface on the guided waves. Here we saw the *A*0 mode breaking up from scattering by the rough surface because of its displacement magnitude being concentrated on that surface. With such EFIT simulations, further research could study in detail how much surface roughness is required to perturb the *A*0 mode propagating down the stringer; this could be used as a simple threshold alarm mechanism in a fieldable inspection device.

In this application, the idealized analytical approach was sophisticated enough to predict a portion of the results, but when considering the larger picture of the system we found that the 3D EFIT simulations provided needed analysis of the subtle guided wave propagation. Using EFIT in future work will continue to bring to light the subtleties of the propagation of guided waves in real structures. One of the best uses for this algorithm is in the planning of experimental tests where there is complex geometry that cannot be solved for analytically. One of the fastest ways of gaining initial understanding of the results from the EFIT simulations is by visualizing the data. Due to computer storage constraints, we collapsed the amount of output data down to just recording the magnitude of the displacement vector every 200 time steps. We are thus discarding 99% of the generated data. It would be of great interest to look at not only the displacement magnitudes but their direction and even some of the stress tensor field components. Visualizing the stresses as the guided waves propagate, or even as a test piece was under stress loading while being inspected would be a valuable tool to gain much more information about the structure under test.

Further simulations could also systematically investigate which frequency-thickness products excite modes that are sensitive to the thickness loss but not as affected by the corrosion surface roughness. Even with these few results from the 3D EFIT simulation, we have been able to further identify features and results from the experimental studies that were difficult to explain analytically. With extensive examination of the complete aircraft stringer model, we could extend our automatic corrosion detection algorithms to more of the aircraft structure.

ACKNOWLEDGMENTS

The authors would like to thank Hasso Weiland of the Alcoa Technical Center and Don Kennamer of Oceana Sensors for helpful discussions and in-kind support. Thanks also to Corey Miller and Cara Campbell for assistance with the experiments and to Chris Boarding for great help with the SciClone simulations.

[1]J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, New York, 1984).
[2]B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Kreiger, Malabar, FL, 1990).
[3]K. E. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[4]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, New York, 1999).
[5]I. A. Viktorov, *Rayleigh and Lamb Waves: Physical Theory and Applications* (Plenum, New York, 1967).
[6]J. L. Rose, "A baseline and vision of ultrasonic guided wave inspection potential," J. Pressure Vessel Technol. **124**, 273–282 (2002).
[7]A. Raghavan and C. E. S. Cesnik, "Review of guided-wave structural health monitoring," Shock Vib. Dig. **39**, 91–114 (2007).
[8]James C. P. McKeon, "Tomography applied to Lamb wave contact scanning," Ph.D. thesis, William and Mary, Williamsburg (1998).
[9]J. C. P. McKeon and M. K. Hinders, "Lamb wave scattering from a through hole," J. Sound Vib. **224**, 843–862 (1999).
[10]J. C. P. McKeon and M. K. Hinders, "Parallel projection and crosshole lamb wave contact scanning tomography," J. Acoust. Soc. Am. **106**, 2568–2577 (1999).
[11]J. C. P. McKeon and M. K. Hinders, "Lamb wave contact scanning tomography," Rev. Prog. Quant. Nondestr. Eval. **18**, 951–958 (1999).
[12]E. V. Malarenko, "Lamb wave diffraction tomography," Ph.D. thesis, William and Mary, Williamsburg (2000).
[13]E. V. Malyarenko and M. K. Hinders, "Fan beam and double crosshole lamb wave tomography for mapping flaws in aging aircraft structures," J. Acoust. Soc. Am. **108**, 1631–1639 (2000).
[14]E. V. Malyarenko and M. K. Hinders, "Ultrasonic Lamb wave diffraction tomography," Ultrasonics **39**, 269–281 (2001).
[15]K. R. Leonard, E. V. Malyarenko, and M. K. Hinders, "Ultrasonic Lamb wave tomography," Inverse Probl. **18**, 1795–1808 (2002).
[16]M. K. Hinders, K. R. Leonard, and E. V. Malyarenko, "Blind test of Lamb wave diffraction tomography," Rev. Prog. Quant. Nondestr. Eval. **21**, 278–283 (2002).
[17]K. R. Leonard and M. K. Hinders, "Guided wave helical ultrasonic tomography of pipes," J. Acoust. Soc. Am. **114**, 767–774 (2003).
[18]K. Leonard, "Ultrasonic guided wave tomography of pipes," Ph.D. thesis, College of William and Mary (2004).
[19]K. R. Leonard and M. K. Hinders, "Lamb wave helical ultrasonic tomography of pipes," Rev. Prog. Quant. Nondestr. Eval. **23**, 173–179 (2004).
[20]K. R. Leonard and M. K. Hinders, "Lamb wave tomography of pipe-like structures," Ultrasonics **43**, 574–583 (2005).
[21]K. R. Leonard and M. K. Hinders, "Multi-mode Lamb wave tomography with arrival time sorting," J. Acoust. Soc. Am. **117**, 2028–2038 (2005).
[22]K. R. Leonard and M. K. Hinders, "Lamb wave tomography of pipes and tanks using frequency compounding," Rev. Prog. Quant. Nondestr. Eval. **24**, 867–874 (2005).
[23]J. Hou, "Ultrasonic signal detection and recognition using dynamic wavelet fingerprints," Ph.D. thesis, College of William and Mary (2004).
[24]J. Hou, K. R. Leonard, and M. K. Hinders, "Multi-mode Lamb wave arrival time extraction of improved tomographic reconstruction," Rev. Prog. Quant. Nondestr. Eval. **24**, 736–743 (2005).
[25]Lord Rayleigh, "On waves propagated along the plane surface of an elastic solid," Proc. London Math. Soc. **s1**, 4–11 (1885).
[26]H. Lamb, "On waves in an elastic plate," Proc. R. Soc. London, Ser. A **93**, 114–128 (1917).
[27]R. D. Mindlin, "Influence of rotatory inertia and shear on flexural motions

of isotropic elastic plates," J. Appl. Mech. **18**, 31–38 (1951).
[28]D. C. Worlton, "Ultrasonic testing with Lamb waves," Nondestr. Test. (Chicago) **158**, 218–222 (1957).
[29]D. C. Worlton, "Experimental confirmation of Lamb waves at megacycle frequencies," J. Appl. Phys. **32**, 967–971 (1961).
[30]A. Abbate, J. Koay, J. Frankel, S. C. Schroeder, and P. Das, "Application of wavelet transform signal processor to ultrasound," Proc.-IEEE Ultrason. Symp. **2**, 1147–1152 (1994).
[31]D. Masscotte, J. Goyette, and T. K. Bose, "Wavelet-transform-based method of analysis for Lamb-wave ultrasonic NDE signals," IEEE Trans. Instrum. Meas. **49**, 524–529 (2000).
[32]D. V. Perov, A. B. Rinkevich, and Ya G. Smorodinskii, "Wavelet filtering of signals for ultrasonic flaw detector," Russian Journal of Nondestructive Testing **38**, 869–882 (2002).
[33]H. W. Lou and G. R. Hu, "An approach based on simplified KLT and wavelet transform for enhancing speech degraded by non-stationary wideband noise," J. Sound Vib. **268**, 717–729 (2003).
[34]J. Zou and J. Chen, "A comparative study on time-frequency feature of cracked rotor by Wigner–Ville distribution and wavelet transform," J. Sound Vib. **276**, 1–11 (2004).
[35]J. Hou and M. K. Hinders, "Dynamic wavelet fingerprint identification of ultrasound signals," Mater. Eval. **60**, 1089–1093 (2002).
[36]J. Hou, K. R. Leonard, and M. K. Hinders, "Automatic multi-mode Lamb wave arrival time extraction for improved tomographic reconstruction," Inverse Probl. **20**, 1873–1888 (2004).
[37]M. Hinders, J. Bingham, K. Rudd, R. Jones, and K. Leonard, "Wavelet thumbprint analysis of time domain reflectometry signals for wiring flaw detection," Rev. Prog. Quant. Nondestr. Eval. **25**, 641–648 (2006).
[38]M. Hinders, R. Jones, K. Leonard, and K. Rudd, "Wavelet thumbprint analysis of time domain reflectometry signals for wiring flaw detection," Engineering Intelligent Systems for Electrical Engineering and Communications **15**, 225–239 (2007).
[39]J. D. Hou, S. T. Rose, and M. K. Hinders, "Ultrasonic periodontal probing based on the dynamic wavelet fingerprint," EURASIP J. Appl. Signal Process. **2005**, 1137–1146 (2005).
[40]P. Fellinger, R. Marklein, K. J. Langenberg, and S. Klaholz, "Numerical modeling of elastic-wave propagation and scattering with EFIT—Elastodynamic finite integration technique," Wave Motion **21**, 47–66 (1995).
[41]F. Schubert, A. Peiffer, B. Koehler, and T. Sanderson, "The elastodynamic finite integration technique for waves in cylindrical geometries," J. Acoust. Soc. Am. **104**, 2604–2614 (1998).
[42]F. Schubert and B. Koehler, "Three-dimensional time domain modeling of ultrasonic wave propagation in concrete in explicit consideration of aggregates and porosity," J. Comput. Acoust. **9**, 1543–1560 (2001).
[43]F. Schubert, "Numerical time-domain modeling of linear and nonlinear ultrasonic wave propagation using finite integration techniques—Theory and applications," Ultrasonics **42**, 221–229 (2004).
[44]K. E. Rudd, K. R. Leonard, J. P. Bingham, and M. K. Hinders, "Simulation of guided waves in complex piping geometries using the elastodynamic finite integration technique," J. Acoust. Soc. Am. **121**, 1449–1458 (2007).
[45]K. E. Rudd, C. A. Bertoncini, and M. K. Hinders, "Simulations of ultrasonographic periodontal probe using the finite integration technique," Open Acoustics Journal **1**, 72–90 (2008).
[46]K. E. Rudd and M. K. Hinders, "Simulation of incident nonlinear sound beam 3D scattering from complex targets," J. Comput. Acoust. **16**, 427–445 (2008).
[47]*Corrosion of Aluminum and Aluminum Alloys*, edited by J. R. Davis (ASM International, Metals Park, OH, 1999).
[48]ASTM G 34-01, Standard Test Method for Exfoliation Corrosion Susceptibility in 2XXX and 7XXX Series Aluminum Alloys (EXCO Test), *American Society for Testing and Materials* (2006).

A study of coupled flexural-longitudinal wave motion in a periodic dual-beam structure with transverse connection

Yi Yun and Cheuk Ming Mak[a)]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

(Received 1 January 2009; revised 12 March 2009; accepted 21 April 2009)

A theoretical study of the multi-coupling flexural and longitudinal waves that propagate in a periodic dual-beam-type waveguide with structural connection branches is conducted. The analytical equations of the transfer matrix method are derived for the wave transmission with consideration given to the fully flexural and longitudinal motions that are tri-coupled at each connection. Based on this transfer matrix method, numerical calculation is performed to investigate the characteristic wave-types that propagate in a semi-infinite periodic structure. The complex wave-coupling phenomena in the periodically connected dual-beam waveguide are then analyzed numerically. Remarkably, it is found that three symmetric and three antisymmetric types of characteristic coupled waves propagate in a periodic structure. The numerical results show that the energy contribution of the coupled waves with respect to the source excitation depends on the forbidden band of the wave-types and on the energy ratios and combination of wave-types. This study promotes the fundamental understanding and prediction of coupled acoustic waves in multi-layered frame structures. The long-term significance is that it may lead to a more effective control method for structure-borne sound transmission in a multi-layered coupling structure.
 [DOI: 10.1121/1.3132706]

PACS number(s): 43.40.At, 43.40.Cw, 43.20.Bi [JGM] Pages: 114–121

I. INTRODUCTION

A number of building structures, bridges, container ship structures, and steel-reinforced concrete constructions are built from an assembly of a number of same or similar structure elements, of which the frames are typically coupled in an identical manner to form a so-called "spatially periodic structure." The excited vibration and transmission of the mechanical waves in these structures—from side to side in a bridge or layer to layer in a building—often give rise to structure-borne noise problems in the connected spaces and can sometimes even be harmful to the stability of the entire structure. When they propagate through frame structures that contain many connection branches, these structure-borne sound waves are coupled and reflected by each connector. The reflected and transmitted waves are then coupled and reflected again by other connectors. This process is physically repeated and sets off infinite multi-interactions between the coupling connections and propagating waves in a periodic structure, which forms the dispersion bands of structure-borne sound waves.

Early on, the dispersion bands of waves in periodic waveguides were studied for the electro-magnetic waves in solids,[1] thus promoting our basic understanding of the properties of conductors, semi-conductors, and the like. Since the late 1980s, the optical wave bands in media with periodical modulation have been extensively studied, and these studies have led to a number of practical applications, including the advanced design of photonic crystals[2] and waveguide devices.[3] All of these studies have brought researchers deeper insight into the dispersion properties of periodic structures and have helped them to develop methods for the theoretical calculation of wave propagation. In acoustics, the classical problem of plane sound wave transmission in one-dimensional periodic media can be tackled in an exact manner via the transfer matrix method.[4] The theoretical computations of band structures have also been well-documented for sound waves in periodic acoustic structures by Kushwaha and Mod.[5] Enhanced wave transmission was modeled in rib-reinforced floors about 50 years ago by using a beam that was periodically loaded with eccentric attachments because of wave coupling.[6] Four different methods of calculating the structure-borne sound propagation in beams with many non-resonant discontinuities were demonstrated by Heckl,[7] and three of these methods took the coupling between longitudinal and flexural waves into account. The fundamental and central ideas in the area of periodic system characterization was introduced by Mead.[8] In this context, a quadratic and well-posed spectral problem was studied to determine the wave propagation constants of a periodic system. This work was extended by Mead,[9] which proposed a second order matrix equation leading to the propagation constants of a periodic system. Several years ago, a mathematical model for the coupling of waves that propagate in a periodically supported Timoshenko beam was presented by Heckl.[10] Furthermore, the propagation characteristics of coupled longitudinal and flexural waves in beam-type transmission paths with asymmetric loads in the form of resonant columns were theoretically analyzed[11] and experimentally examined[12] by Friss and Ohlrich. However, little understanding of the fundamental physical propagation characteristics of the coupling acoustic

[a)]Author to whom correspondence should be addressed. Electronic mail: becmmak@polyu.edu.hk

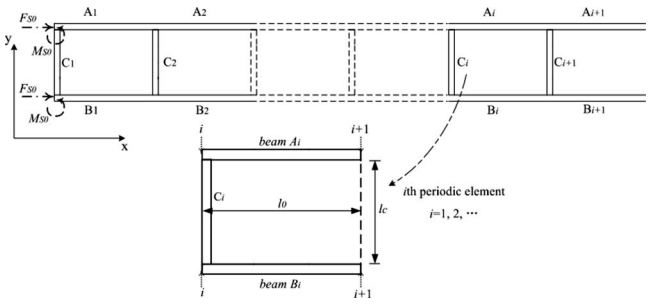


FIG. 1. Scheme of the semi-infinite periodic dual-beam structure and the excitations.

waves in multi-layered structures has been gained from these studies. It is because they are commonly concerned with models of a single-channel waveguide that comprises the independent beam-type components or uncoupled wave transmission path in the structure.

Therefore, the analytical study reported in this paper investigated the characteristics of the multi-coupling flexural and longitudinal waves that propagate in a periodic dual-beam-type waveguide with structural connection branches. The propagation of waves in a semi-two-dimensional system was adopted because the coupling interaction between the waves at the connections and transmission paths through a three-dimensional structure is so complicated that the theoretical predictions based on the series of approximations can be very different from the actual experimental observations. The transfer matrix method is developed by using the concept based on the propagation constants[8,9] of the waves in a periodic structure so that it avoids the problems from inversion of ill-conditioned matrices and the cumulative errors. The developed method is therefore explicit and appropriate for the calculation of the coupled waves in the periodic beam structure.

II. THEORETICAL MODEL AND ANALYTICAL FUNDAMENTALS

A. Simple model of a periodic dual-beam structure with a transverse connection

This paper examines the band structure of flexural-longitudinal wave propagation in a dual-beam coupling structure that is periodically connected with transverse branches. A simplified model is shown in Fig. 1. The structure-borne sound consists of the flexural waves and longitudinal waves that propagate in two horizontal beams—A and B are coupled at each connection with a vertical branch C_i. The beams and branches discussed theoretically are even, straight, isotropic, and homogeneous, and the following physical parameters are assumed. $\rho_{(1,2,3)}$=the density of beams A and B and branch C_i, $B_{(a,b,c)0}$=the bending stiffness of beams A and B and branch C_i, $E_{(A,B,C)}$=Young's modulus of beams A and B and branch C_i, $k_{(A,B,C)f}$=the flexural wave numbers of beams A and B and branch C_i, and $k_{(A,B,C)l}$=the longitudinal wave numbers corresponding to the acoustic speeds of the longitudinal wave $c_{(1,2,3)}$ of beams A and B and branch C_i. The characteristics of the wave-types and the energy transmission of the coupled flexural-longitudinal waves in a semi-infinite periodic dual-beam structure are calculated for analysis in a case study.

B. Wave transfer matrix and the propagation constants of the characteristic wave-types

In the analytical model of this research, the coupled wave components of the complex velocity (horizontal, vertical, and rotational) and force (horizontal, vertical, and moment) vectors are used for describing the coupled wave motions and response in the dual-beam structure. For mathematical derivation, all of the analytical equations in this paper are based on the harmonic wave of separate frequency ω_n with time dependence suppressed. Normally the longitudinal and flexural waves in a beam can be expressed in the form of the independent wave velocity components. In addition, the longitudinal-flexural waves can be described by the beam velocities and the corresponding forces caused by the wave motions. At the connections of every periodic element, the vector consisting of velocity and force components can be related to the vector of the flexural-longitudinal waves that propagate through the beams in a matrix form as follows:

$$\begin{bmatrix} V_n \\ F_n \end{bmatrix} = [S_{VF}][v_w], \tag{1}$$

$$[v_w] = \begin{bmatrix} v_{wA} \\ v_{wB} \end{bmatrix} = [S_{VF}]^{-1} \begin{bmatrix} V_n \\ F_n \end{bmatrix}, \tag{2}$$

where the velocity vectors of the flexural and longitudinal wave components are expressed as

$$[v_{wA}] = [v_{Af}^{+}\ v_{Afj}^{+}\ v_{Af}^{-}\ v_{Afj}^{-}\ v_{Al}^{+}\ v_{Al}^{-}]^T,$$

$$[v_{wB}] = [v_{Bf}^{+}\ v_{Bfj}^{+}\ v_{Bf}^{-}\ v_{Bfj}^{-}\ v_{Bl}^{+}\ v_{Bl}^{-}]^T,$$

of which "+" donates the wave components propagating in the positive x-direction, "−" donates the components going in the negative x-direction, f, fj, and l donate the propagating flexural, nearfield flexural, and longitudinal wave components, respectively. The vectors of the velocities and forces of the two beam channels, indicated by the subscripts a and b, are expressed as

$$[V_n] = [V_{ya}\ V_{yb}\ \omega_a\ \omega_b\ V_{xa}\ V_{xb}]^T,$$

$$[F_n] = [F_{ya}\ F_{yb}\ M_a\ M_b\ F_{xa}\ F_{xb}]^T,$$

where V_x, V_y, and ω are the x-degree, y-degree, and rotational velocities in the beam, and F_x, F_y, and M are the x-degree force, y-degree force, and moment acting on a beam. To describe the relationship between the independent flexural-longitudinal waves and the velocities-forces in two beams, the waves to motions-actions transfer matrix $[S_{VF}]$ takes on the matrix form:

$$[S_{VF}] = [S_{V1}\ S_{V2}\ S_{V3}\ V_{F1}\ S_{F2}\ S_{F3}]^T,$$

$$S_{Vi} = \begin{bmatrix} D_{Vai} & \mathbf{O}_{6\times1} \\ \mathbf{O}_{6\times1} & D_{Vbi} \end{bmatrix}, \quad S_{Fi} = \begin{bmatrix} D_{Fai} & \mathbf{O}_{6\times1} \\ \mathbf{O}_{6\times1} & D_{Fbi} \end{bmatrix}, \tag{3}$$

in which the matrix are derived by

$$D_{V(a,b)1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad D_{V(a,b)3} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix},$$

$$D_{Va2} = \begin{bmatrix} -jk_{Af} \\ -k_{Af} \\ jk_{Af} \\ k_{Af} \\ 0 \\ 0 \end{bmatrix}, \quad D_{Vb2} = \begin{bmatrix} -jk_{Bf} \\ -k_{Bf} \\ jk_{Bf} \\ k_{Bf} \\ 0 \\ 0 \end{bmatrix},$$

$$D_{Fa1} = \begin{bmatrix} jR_{AF} \\ -R_{AF} \\ -jR_{AF} \\ R_{AF} \\ 0 \\ 0 \end{bmatrix},$$

and

$$D_{Fb1} = \begin{bmatrix} jR_{BF} \\ -R_{BF} \\ -jR_{BF} \\ R_{BF} \\ 0 \\ 0 \end{bmatrix}, \quad D_{Fa2} = \begin{bmatrix} R_{AM} \\ -R_{AM} \\ R_{AM} \\ -R_{AM} \\ 0 \\ 0 \end{bmatrix},$$

$$D_{Fb2} = \begin{bmatrix} R_{BM} \\ -R_{BM} \\ R_{BM} \\ -R_{BM} \\ 0 \\ 0 \end{bmatrix}, \quad D_{Fa3} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ R_{Al} \\ -R_{Al} \end{bmatrix}, \quad D_{Fb3} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ R_{Bl} \\ -R_{Bl} \end{bmatrix},$$

$$R_{AM} = \frac{B_{a0}k_{Af}^2}{j\omega_n}, \quad R_{BM} = \frac{B_{b0}k_{Bf}^2}{j\omega_n}, \quad R_{AF} = \frac{B_{a0}k_{Af}^3}{j\omega_n},$$

$$R_{BF} = \frac{B_{b0}k_{Bf}^2}{j\omega_n}, \quad R_{Al} = \rho_1 c_1, \quad R_{Bl} = \rho_2 c_2.$$

It is clear that the y-degree velocities and forces, rotational velocities, and moments in the beams A and B resulted from the flexural wave motions, while the x-degree velocities and forces in two beams resulted from the longitudinal wave motions. It should be noted that the wave-coupling effect in a periodic dual-beam structure is caused by the vertical connections. By introducing the dynamic continuity conditions at the interfaces that are vertically connected with the branch beams, the relationship between the velocities and the forces of the coupled flexural and longitudinal waves at the connections of dual-beam structure can be characterized as a 12×12 coupling transfer matrix that can be expressed as

$$\begin{bmatrix} V_n \\ F_n \end{bmatrix}_i^+ = [W_C] \times \begin{bmatrix} V_n \\ F_n \end{bmatrix}_i^+, \tag{4}$$

where

$$\begin{bmatrix} V_n \\ F_n \end{bmatrix}_i^- \quad \text{and} \quad \begin{bmatrix} V_n \\ F_n \end{bmatrix}_i^+$$

denote the velocity and force vectors of the beam on the left and right side of the connection points on beam A-B with C_i. Based on the dynamic equilibrium on two sides of a branch C_i, the wave-coupling matrix $[W_C]$ is given by

$$[W_C] = \begin{bmatrix} I_6 & O_6 \\ -Z_{Cw} \times T_{Cv} & I_6 \end{bmatrix}, \tag{5}$$

in which the transfer elements are given by

$$Z_{Cw} = \begin{bmatrix} Z_{Cl} & O_{2\times4} \\ O_{4\times2} & Z_{Cf} \end{bmatrix},$$

$$Z_{Cl} = \begin{bmatrix} R_{cl} & -R_{cl} \\ -R_{cl}\phi_{Cl}^{-j} & R_{cl}\phi_{Cl}^{j} \end{bmatrix},$$

$$Z_{Cf} = \begin{bmatrix} \Omega_{Mc} & -\Omega_{Mc} & \Omega_{Mc} & -\Omega_{Mc} \\ -\phi_{Cf}^{-j}\Omega_{Mc} & \phi_{Cf}^{-1}\Omega_{Mc} & -\phi_{Cf}^{j}\Omega_{Mc} & \phi_{Cf}\Omega_{Mc} \\ j\Omega_{Fc} & -\Omega_{Fc} & -j\Omega_{Fc} & \Omega_{Fc} \\ -j\phi_{Cf}^{-j}\Omega_{Fc} & \phi_{Cf}^{-1}\Omega_{Fc} & j\phi_{Cf}^{j}\Omega_{Fc} & -\phi_{Cf}\Omega_{Fc} \end{bmatrix},$$

$$T_{Cf} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ \phi_{Cf}^{-j} & \phi_{Cf}^{-1} & \phi_{Cf}^{j} & \phi_{Cf} \\ -jk_{Cf} & -k_{Cf} & jk_{Cf} & k_{Cf} \\ -jk_{Cf}\phi_{Cf}^{-j} & -k_{Cf}\phi_{Cf}^{-1} & jk_{Cf}\phi_{Cf}^{j} & k_{Cf}\phi_{Cf} \end{bmatrix}^{-1},$$

$$T_{Cv} = \begin{bmatrix} T_{Cl} & O_{2\times4} \\ O_{4\times2} & T_{Cf} \end{bmatrix}, \quad T_{Cl} = \begin{bmatrix} 1 & 1 \\ \phi_{Cl}^{-j} & \phi_{Cl}^{j} \end{bmatrix}^{-1},$$

$$R_{cl} = \rho_3 c_3, \quad \phi_{Cl} = e^{k_{Cl}h}, \quad \phi_{Cf} = e^{k_{cf}h},$$

$$\Omega_{Mc} = \frac{B_{0C}k_{Cf}^2}{j\omega_n}, \quad \Omega_{Fc} = \frac{B_{0C}k_{Cf}^2}{j\omega_n}.$$

The new flexural and longitudinal waves that are excited in the branch will result in the velocities and forces acting on the two connection sides of the branch with the beams A and B, so that the flexural and longitudinal waves will be coupled there. Moreover, the transfer matrix of the longitudinal and flexural waves propagating in the continuous beam period (whose length=d) is given by

$$[v_w]_{i+1}^- = [P_{wv}][v_w]_i^+.$$

$$[P_{wv}] = \begin{pmatrix} P_{Af} & O & O & O \\ O & P_{Al} & O & O \\ O & O & P_{Bf} & O \\ O & O & O & P_{Bl} \end{pmatrix},$$

$$P_{Af} = \begin{pmatrix} e^{-jk_{Af}d} & 0 & 0 & 0 \\ 0 & e^{-k_{Af}d} & 0 & 0 \\ 0 & 0 & e^{jk_{Af}d} & 0 \\ 0 & 0 & 0 & e^{k_{Af}d} \end{pmatrix},$$

$$P_{Bf} = \begin{pmatrix} e^{-jk_{Bf}d} & 0 & 0 & 0 \\ 0 & e^{-k_{Bf}d} & 0 & 0 \\ 0 & 0 & e^{jk_{Bf}d} & 0 \\ 0 & 0 & 0 & e^{k_{Bf}d} \end{pmatrix}, \tag{6}$$

$$P_{Al} = \begin{pmatrix} e^{-jk_{Al}d} & 0 \\ 0 & e^{jk_{Al}d} \end{pmatrix}, \quad P_{Bl} = \begin{pmatrix} e^{-jk_{Bl}d} & 0 \\ 0 & e^{jk_{Bl}d} \end{pmatrix}.$$

It is generally understood that the flexural and longitudal waves can propagate independently through the dual-beam part between the connnection branches without coupling. Therefore, the transfer relationship of the waves in the continuous beam part can be described by using the diagonal matrices. On the whole, the coupled wave transmission in the periodic structure can be expressed as

$$\begin{bmatrix} V_n \\ F_n \end{bmatrix}_{i+1}^{-} = [U_e] \times \begin{bmatrix} V_n \\ F_n \end{bmatrix}_{i}^{-}, \tag{7}$$

where the entire periodic transfer matrix is given

$$[U_e] = [S_{VF}][P_{wv}][S_{VF}]^{-1}[W_C].$$

According to the Bloch wave theory,[13] for a linear acoustic system, when the acoustic waves are propagating through a semi-infinite one-dimensional periodic structure, the wave motions can be described as the characteristic wave-types of Bloch waves. Then the relationship between velocity vector $[V_n]_i$ and force vector $[F_n]_i$ at the two periodic connection points nearby satisfies the form

$$[V_n]_{(i)}^{-} = \sum_{j=1}^{N} v_{j,(i)}[\xi_{jn}], \quad [F_n]_{(i)}^{-} = \sum_{j=1}^{N} f_{j,(i)}[\zeta_{jn}], \tag{8}$$

$$\begin{bmatrix} v \\ f \end{bmatrix}_{(i+1)} = e^{\mu} \begin{bmatrix} v \\ f \end{bmatrix}_{(i)}. \tag{9}$$

This represents a problem on the eigenvalue vector for the transfer matrix $[U_e]$, where $\mu_j = \pm(\mu_{jR} + j \cdot \mu_{jI})$ are the pair of jth eigen values—the frequency-dependent complex propagation constants for the corresponding pair of the N characteristic wave-types ($N=6$ for this periodic structure). Correspondingly the characteristic wave-types are formulated by the eigenvectors $[\xi_{jn}\zeta_{jn}]^T$, which take on the form $[\xi_{jn}]=[X^j_{Vxa}, X^j_{Vxb}, X^j_{\omega a}, X^j_{\omega b}, X^j_{Vya}, X^j_{Vyb}]^T$ and $[\zeta_{jn}]=[X^j_{Fxa}, X^j_{Fxb}, X^j_{Ma}, X^j_{Mb}, X^j_{Fya}, X^j_{Fyb}]^T$. As the "attenuation constant" of the coupled wave type, the real part μ_{jR} expresses the exponential decay rate for the jth characteristic wave-type that propagates through a periodic beam element, whereas the imaginary part μ_{jI} is defined as the "phase constant," of which the cosine value describes the phase transfer of the jth characteristic wave-type that propagates through each element. If the propagation constants of the positive-going wave-types are defined as $\mu_j = \mu_{jR} + \cdot \mu_{jI} (0 \le |\mu_{jI}| < 2\pi)$, then, correspondingly, the real and imaginary parts of the propagation constants ought to be negative. For the frequency-dependent wave propagation in a periodic structure, the frequency domain is classified into pass bands, i.e., frequencies at which the coupled waves travel through the periodic structure with little loss, and forbidden bands, i.e., frequencies at which the coupled waves propagating in the periodic structure are evanescent. In an ideal case as the damping factor is negligible, a pair (positive and negative-going) of characteristic wave-types yield up to a pair of pure imaginary propagation constants at any frequency within the pass bands, which indicates that the coupled waves propagating through the periodic structure will not decay. On the other side, the real parts of propagation constants are nonzero at the frequencies within the forbidden bands, which indicates that the coupled waves will be attenuated as propagating through the periodic structure. The zone of larger attenuation constants, i.e., $|\mu_{jR}|$ means that the corresponding wave-type is in the stronger forbidden band of the periodic structure. In the semi-infinite structure, only the positive-going propagation constant $\mu_j = -|\mu_{jR}| - j|\mu_{jI}|$ is the reasonable solution, because neither $+|\mu_{jI}|$ nor $+|\mu_{jR}|$ for a negative-going wave-type is physically possible for the phase retardation and energy decay in propagation.

III. ANALYSIS AND DISCUSSION

A. Settings and parameters used in computation

The numerical analysis and choice of the physical parameters for a semi-infinite periodic structure were designed to investigate and reveal the coupling effects of wave propagation. All of the computations and matrix manipulations were conducted using MATLAB. Aluminum was chosen as the beam material, of which Young's modulus is $E_0 = 6.9 \times 10^{10}$ N/m^2 with loss factor $\eta = 0.002$ and density $\rho_0 = 2700$ kg/m^3. The two equal beams are semi-infinite along the x-direction and periodically connected by the transverse beams with a same rectangular cross-section. The thickness and width of the beam cross-section are $h_0 = 11$ mm and $d_0 = 50$ mm, respectively, the periodic element length is $l_0 = 550$ mm and the length of transverse connection beam is $l_c = 500$ mm. The results in the frequency domain computed for the analysis and the discussion herein of the characteristic coupled waves are normalized by using the non-dimensional frequency parameter Ω_n,[11] which is given by

$$\Omega_n = (k_f l_0)^2 = \omega_n (12\rho_0/E_0)^{1/2}(l_0/h_0), \tag{10}$$

where k_f is the real wave number for the free flexural waves that are propagating in beams A and B.

B. Propagation constants and the nature of the wave-types

Basically, there are six characteristic wave-types for the coupled waves that propagate in the semi-infinite periodic waveguide, which all contain both positive-going and negative-going longitudinal and flexural wave motions in the beams because of the multi-coupling at the beam connections. These characteristic wave-types can be divided into two groups—symmetric and antisymmetric types—based on the different phase relationships of the wave motions between beams A and B. Herein the symmetric wave-types are named because the phase differences of the y-degree velocities between beam A and beam B are π and the phase differences of the x-degree velocities between two beams are 0. They are like the mirror images from the symmetry axis of the dual-beam structure in x-direction. For the motions of antisymmetric wave-types, the phase differences of the y-degree velocities between beam A and beam B are 0 and the phase differences of the x-degree velocities between two beams are π. They are like the inverted images from the symmetry axis of the dual-beam structure in x-direction. A further step to describe the propagation characteristics of the wave-types is to use the dispersion of the propagation constants. The computed results for the attenuation constants μ_R and $\cos(\mu_I)$ for the characteristic wave-types in the periodic beam structure are plotted in Figs. 2(a)–2(c). It can be seen from Fig. 2(a) that the symmetric flexural-longitudinal wave motion is governed by two types: α-I and α-II. It is found that more attenuation zones belong to wave-type α-I, and they fall off slowly and have broad forbidden bands. Those that belong to wave-type α-II, which, for the most part, belong to the frequency region below $\Omega_n=330$, have pass bands, but two strong forbidden bands from nearly $\Omega_n=25-47$ and 133–177, where the attenuation constants of α-II fall off rapidly and have sharp peaks at around two significant symmetric resonant modes of the connecting beam branch. It should be noted that the values of $\cos(\mu_I)$ are equal to 1 or -1 in most regions of the forbidden bands. It can be seen from Fig. 2(a) that the two curves of $\cos(\mu_I)$ for the two wave-types overlap at certain normalized frequencies where the attenuation constant is non-zero. This implies that the propagation constants almost become complex conjugates with the non-zero attenuation constant. Similarly, it can be seen from Fig. 2(b) that the antisymmetric flexural-longitudinal wave propagation is governed by two wave-types: β-I and β-II. It is found that more attenuation zones that correspond to the forbidden bands of type β-I fall off slowly and have broad bands. Those of type β-II in most regions below $\Omega_n=310$ have pass bands, but two significant forbidden bands from nearly 62–72 and 219–271, where the attenuation constants of β-II fall off rapidly and have two sharp peaks at around two strong antisymmetric resonant modes of the connecting beam branch.

Strong wave coupling occurs in the forbidden band gaps of the coupled longitudinal-flexural waves, as they are strongly attenuated through the periodic structure. In Fig. 2(c), the attenuation constants and $\cos(\mu_I)$ of the predominantly near-field wave-types are plotted as symmetric and antisymmetric types. For the predominantly near-field waves, the attenuation constants are obviously larger than those for the other wave-types, and all of the $\cos(\mu_I)$ values are almost equal to 1 or -1, which indicates that the energy of predominantly near-field waves decays dramatically as the waves propagate. As these two wave-types are in the strong forbid-

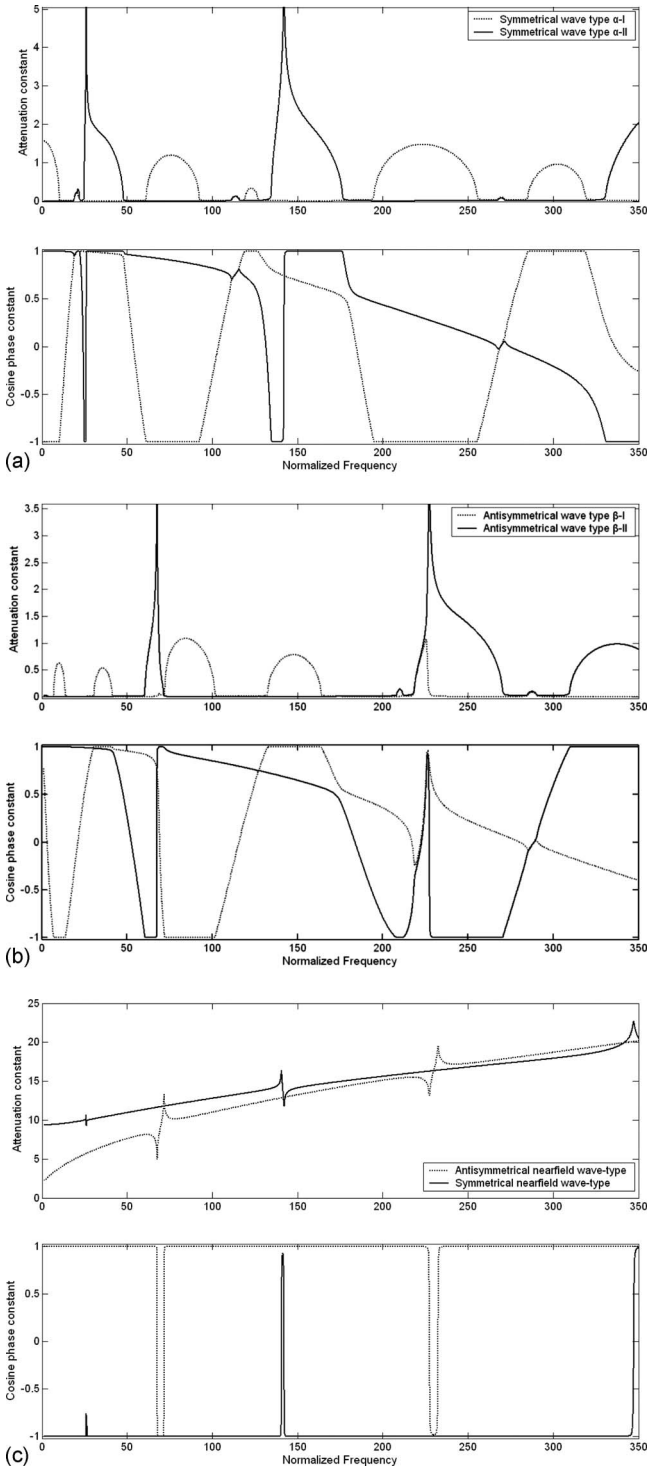


FIG. 2. Propagation constants of characteristic wave-types. (a) μ_R and $\cos(\mu_I)$ of the symmetric flexural-longitudinal wave-types: α-I and α-II. (b) μ_R and $\cos(\mu_I)$ of the antisymmetric flexural-longitudinal wave-types: β-I and β-II. (c) μ_R and $\cos(\mu_I)$ of the symmetric and antisymmetric predominantly near-field wave-types.

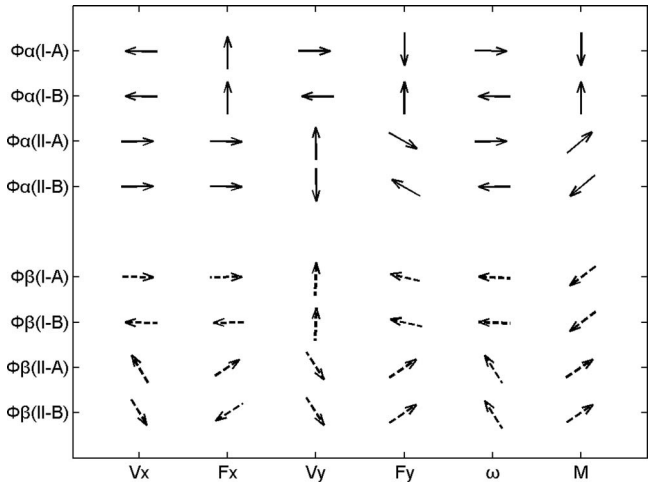


FIG. 3. Phase relationship of characteristic flexural-longitudinal wave vectors in normalized frequency $\Omega_n = 150$.

den band regions, they can be ignored in the consideration of structure-borne sound transmission through a periodic structure.

Figure 3 shows the indicative results of the phase behavior of force-velocity vectors and further illustrates the phase relationship of the motions and actions between two beams. The phase vectors of the coupled flexural-longitudinal waves at the connection between beams A and B are chosen in the normalized frequency $\Omega_n = 150$. Remarkably, it can be observed that for symmetric wave-types α-I and II, the phases of y-degree velocity V_y and rotational velocity ω of beam A are the reverse of those of beam B, and the phase of longitudinal velocity V_x of beam A is the same as that of beam B. In contrast, for antisymmetric wave-types β-I and II, it is found that the phases of y-degree velocity V_y and rotational velocity ω of beam A are the same as those of beam B, whereas the phase of longitudinal velocity V_x of beam A is the reverse of that of beam B. As the frequency is chosen from the pass bands of wave-types α-I and β-II, the phase vectors of their forces and moments point in different directions than their velocities and rotational velocities, which, in total, results in the positive energy flow constantly propagating along the periodic beams. However, the frequency is in the forbidden bands of wave-types α-I and β-II, and almost all of the phase vectors of their forces and moments are perpendicular to the phase vectors of their velocity fields, which indicates that the energy flow cannot continuously propagate through the periodic structure because of energy losses.

C. Excited waves in a semi-infinite periodic structure

In this section, the effect of wave coupling on the response of an ideal semi-infinite periodic structure to two synchronous point excitations is investigated via simulation using the analytical transfer matrix method. Two types of harmonic source excitations that synchronously act on the left side of the semi-infinite beams A and B (along the x-axis) are considered. They are defined as the standardized synchronous longitudinal (x-degree) forces of amplitude $F_{S0} = E_0 S_0$ and the synchronous moments of amplitude $M_{S0} = (E_0 I_0)/l_0$, where I_0 is the second moment of area of the beam elements and S_0 is its cross-sectional area. The normalized force/moment levels of the propagating wave-types through the first to fifth beam elements are plotted in Figs. 4(a) and 5(a) in the conditions of being excited by the longitudinal forces and moments, which are defined as $\hat{f}_{in}^{j} = 20\log(|\mathbf{f}_{i,j}|/|F_{\text{excit}}|)$, where $\mathbf{f}_{i,j}$ corresponds to the normalized eigenvector ζ_{jn} satisfying the condition that $X_{FxA}^{j} = 1$ is being excited by the synchronous longitudinal forces, and $X_{MA}^{j} = 1$ is the excitation of the synchronous moments. In addition, the variations in the total flexural and longitudinal energy levels of every beam element are plotted in Figs. 4(b) and 5(b) in the forms given by

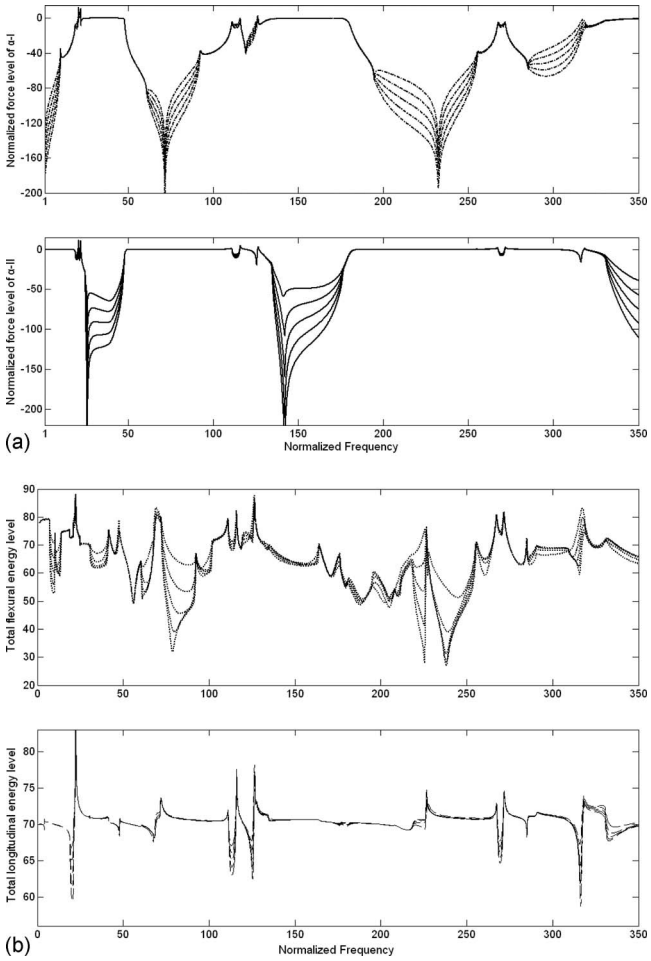


FIG. 4. Amplitude and energy transmission of coupled waves in response to the excitation of synchronous longitudinal forces. (a) The normalized force levels of the symmetrical wave-types α-I and α-II that propagate through the first to fifth beam elements. (b) The total longitudinal and flexural energy levels of the five beam elements for synchronous longitudinal force excitation.

$$LE_{\text{long}} = 10\log\left[\rho_0 S_0(|v_l^+|^2 + |v_l^-|^2)\int_{l_0}\cos^2 k_l x \cdot dx\right], \quad (11)$$

$$LE_{\text{flex}} = 10\log\rho_0 S_0\left[\frac{1}{2}(|v_f^+|^2 + |v_f^-|^2)l_0 + |v_f^+ v_f^-|\int_{l_0}\cos(2k_f x)\cdot dx\right], \quad (12)$$

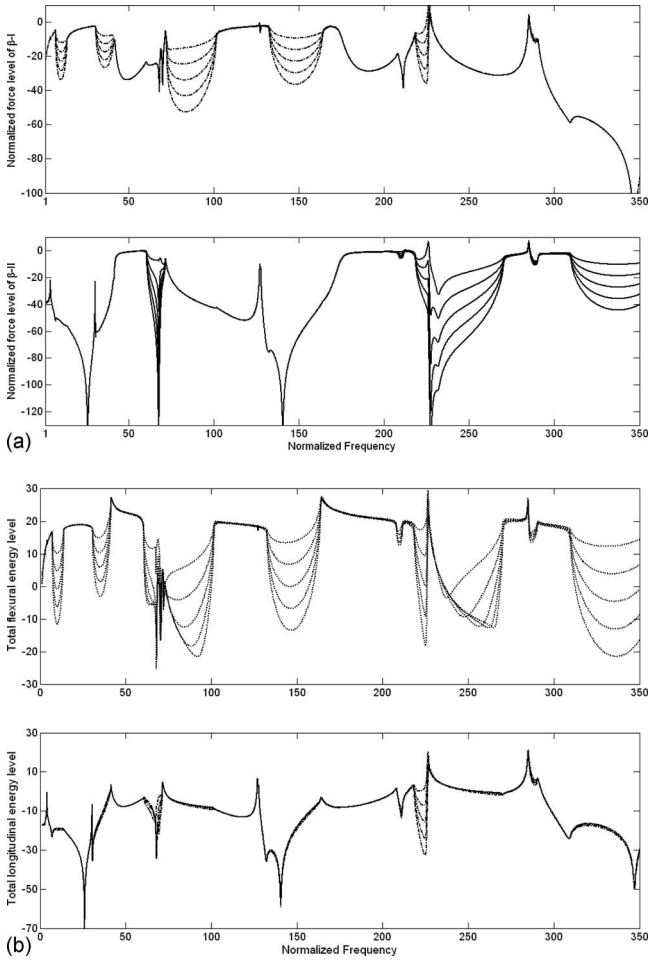


FIG. 5. Amplitude and energy transmission of coupled waves in response to the excitation of synchronous moments. (a) The normalized force levels of symmetrical wave-types β-I and β-II that propagate through the first to fifth beam elements. (b) The total longitudinal and flexural energy levels of the five beam elements for synchronous moment excitation.

Herein the energy unit –J is suppressed because of the use of standardized excitations.

For a periodic structure that is being excited by the synchronous longitudinal forces F_{S0}, the normalized force levels of the symmetrical wave-types α-I and α-II that propagate through the first to fifth beam elements are plotted and shown in Fig. 4(a). The near-field wave-types are neglected, as they decay significantly after propagating through a few elements, and the antisymmetric waves are omitted because they cannot be excited in this case. Notably, it can be seen from Fig. 4(a) that wave-type α-I is excited at a low level and is attenuated significantly within the frequency regions of about Ω_n=0–11, 65–90, 195–256, and Ω_n=285–320 (which belong to the main forbidden bands of α-I) as it propagates through the structure, whereas wave-type α-II is excited at a high level (i.e., the excited forces are near to the source excitation forces) and propagates through the structure without significant attenuation at those frequency regions. Similarly, wave-type α-II is excited at a low level and is attenuated significantly within the frequency regions of about Ω_n=25–48, 136–180, and Ω_n=330 and above (which belong to the strong forbidden bands of α-II) as it propagates through the structure, whereas wave-type α-I is excited at a high level and propagates through the structure without significant attenuation at those frequency regions. Figure 4(b) shows the total longitudinal and flexural energy levels of the five beam elements. It should be noted from this figure that the total longitudinal energy is excited at a considerable level (near to 73), and the waves propagate through the structure without significant energy loss at most frequency regions, except for certain narrow zones that belong to the small forbidden bands of wave-type α-I or α-II. Two prominent gaps in the curves of the total flexural energy level at the frequency regions that correspond to the strong forbidden bands of wave-type α-I can be observed, as the total flexural energy level is mainly due to the coupling effect of the structure. On the other hand, the total longitudinal energy level holds relatively steady at most frequencies, as the total longitudinal energy level is mainly due to the direct effect of the longitudinal source exciting forces. In fact, these two prominent gaps indicate that wave-type α-I contributes most of the energy to the total flexural energy level, compared with wave-type α-II, at those frequency regions. This means that the energy contribution of coupled waves with respect to source excitation depends not only on the forbidden band of the wave-types but also on the energy ratios and combination of wave-types.

For a structure that is being excited by synchronous longitudinal forces M_{S0}, the normalized force levels of symmetrical wave-types β-I and β-II that propagate through the first to fifth beam elements are plotted separately in Fig. 5(a). The near-field wave-types and the antisymmetric waves are again neglected. Notably, it can be seen from Fig. 5(a) that the excited wave type β-I is excited at a low level and is attenuated significantly within the frequency regions of about Ω_f=5–10, 30–40, 72–99, 132–164, and from 220 to 229, which belong to the main forbidden bands of β-I as it propagates through the structure. Besides, the wave-type β-II is excited at a high level where the excited moments are near the source excitation moments and the wave-type β-II propagates through the structure without significant attenuation at those frequency regions. Similarly, wave-type β-II is excited significantly at a low level and is attenuated strongly within the frequency regions of about Ω_f=11–21, 218–270, and Ω_f=311 and above (which belong to the strong forbidden bands of β-II) as it propagates through the structure, whereas wave-type β-I is excited at a high level and propagates through the periodic structure without significant energy loss at those frequency regions. Figure 5(b) shows the total longitudinal and flexural energy levels of the five beam elements for synchronous moment excitation. A comparison of the shapes of the curves of the total flexural energy level in Fig. 5(b) with those in Fig. 5(a) shows that the propagating flexural energy at frequencies approximately lower than 225 is mainly due to the transmission of wave-type β-I, whereas the propagating flexural energy at frequencies approximately higher than 225 is mainly due to the transmission of wave-type β-II. Figure 5 again illustrates that the energy contribution of coupled waves with respect to source excitation depends on the forbidden band of the wave-types and on the energy ratios and combination of wave-types.

IV. CONCLUSION

A new model based on a multi-coupling wave transfer matrix has been developed to study the phenomena of the coupled flexural-longitudinal waves that propagate in tri-coupled dual-channel periodic beam-type waveguides. A lightly damped semi-infinite structure that consists of two equally thin semi-infinite beams connected with resonant branches has been numerically analyzed. The connection branches are the beams perpendicularly connected at regular intervals. This type of waveguide can simulate a one- to two-dimensional model of a column-beam frame for modern steel-concrete buildings or bridges simply. The computed results of the complex propagation constants that govern the transmission of wave-types in periodic structures have clearly revealed the characteristics of pass and forbidden bands and the wave-coupling phenomena. It is found that there are six characteristic coupled wave-types that propagate through such a structure, and these can be divided into symmetric and antisymmetric groups of flexural-longitudinal and predominantly near-field characteristic wave-types. Their properties under different excitations are quantified from the computed transmission of the normalized amplitudes of the coupled wave-types together with the maximum flexural and longitudinal energies along the wave-carrying components. It has been revealed that the structure-borne sound energy from the synchronous longitudinal excitations at two beams mainly propagate through the periodic structure in the form of one or two types of symmetric characteristic coupled flexural-longitudinal waves. In contrast, the structure-borne sound energy from the synchronous rotational sources that excite dual-channel beams mainly propagate along the periodic structure in the form of one or two types of antisymmetric characteristic coupled flexural-longitudinal waves. These results demonstrate that the energy contribution of coupled waves with respect to source excitation depends on the forbidden band of the wave-types and on the energy ratios and combination of wave-types.

[1] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996).
[2] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystal-Molding the Flow of Light* (Princeton University Press, Princeton, NJ, 1995).
[3] M. Ibanescu, Y. Fink, S. Fan, E. L. Thomas, and J. D. Joannopoulos, "An all-dielectric coaxial waveguide," Science **289**, 415–419 (2000).
[4] C. H. Hodges and J. Woodhouse, "Vibration isolation from irregularity in a nearly periodic structure: Theory and measurements," J. Acoust. Soc. Am. **74**, 894–905 (1983).
[5] M. S. Kushwaha, "Classical band structure of periodic elastic composites," Int. J. Mod. Phys. B **10**, 977–1094 (1996).
[6] H. L. Müller, "Attenuation of bending waves caused by symmetrical and eccentric blocking masses," Dr.-Ing. dissertation, Institut für Technische Akustik der Technischen Universität, Berlin (1957).
[7] M. Heckl, "Structure-borne sound propagation on beams with many discontinuities," Acustica **81**, 439–449 (1995).
[8] D. J. Mead, "A general theory of harmonic wave propagation in linear periodic system with multiple coupling," J. Sound Vib. **27**, 235–260 (1973).
[9] D. J. Mead, "Wave propagation and natural modes in periodic systems: II. Multi-coupled systems, with and without damping," J. Sound Vib. **40**, 19–39 (1975).
[10] M. A. Heckl, "Coupled waves on a periodically supported Timoshenko beam," J. Sound Vib. **252**, 849–882 (2002).
[11] L. Friis and M. Ohlrich, "Coupling of flexural and longitudinal wave motion in a periodic structure with asymmetrically arranged transverse beams," J. Sound Vib. **118**, 3010–3020 (2005).
[12] L. Friis and M. Ohlrich, "Coupled flexural-longitudinal wave motion in a finite periodic structure with asymmetrically arranged transverse beams," J. Sound Vib. **118**, 3607–3618 (2005).
[13] H. Umezawa, *Advanced Field Theory* (AIP, New York, 1995).

Energy equipartition and frequency distribution in complex attachments

N. Roveri
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza," Via Eudossiana 18, 00184 Rome, Italy

A. Carcaterra[a)]
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza," Via Eudossiana 18, 00184 Rome, Italy and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

A. Akay
Department of Mechanical Engineering, Bilkent University, 06800 Bilkent, Ankara, Turkey and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 13 November 2008; revised 26 April 2009; accepted 11 May 2009)

As reported in several recent publications, an undamped simple oscillator with a complex attachment that consists of a set of undamped parallel resonators can exhibit unusual energy sharing properties. The conservative set of oscillators of the attachment can absorb nearly all the impulsive energy applied to the primary oscillator to which it is connected. The key factor in the ability of the attachment to absorb energy with near irreversibility correlates with the natural frequency distribution of the resonators within it. The reported results also show that a family of optimal frequency distributions can be determined on the basis of a variational approach, minimizing a certain functional related to the system response. The present paper establishes a link between these optimal frequency distributions and the energy equipartition principle: optimal frequency distributions are those that spread the injected energy as uniformly as possible over the degrees of freedom or over the modes of the system. Theoretical as well as numerical results presented support this point of view. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3147502]

PACS number(s): 43.40.At, 43.40.Kd, 43.40.Jc, 43.40.Tm [JJM] Pages: 122–128

I. INTRODUCTION

In the past decade the problem of energy sharing among a principal, or master, structure attached to a large number of resonators has been analyzed in some detail (viz., Refs. 1–3). Mechanism of energy sharing in a complex structure as depicted by the prototypical system described here brings out important fundamental issues in such systems and also has numerous engineering applications. Many engineering structures follow a similar paradigm where a population of resonators is attached to a principal structure. For example, structures such as a car body, airplane fuselage, or hull of a ship are coupled to a very large number of resonating interior components. Moreover, the fundamental aspects of the concept of complex attachments can be used for designing novel vibration absorbers.[4] The energy exchange that takes place in the complex system described here is substantially independent of any intrinsic damping in the system.[1–8]

Theoretical analyses that have been reported in a series of recent papers[9–13] describe how a complex attachment can rapidly and permanently absorb energy from a master structure. One of the basic findings of these investigations was the discovery of the significance of the distribution of the natural frequencies of the attached resonators in this energy transfer process. The reported analyses revealed the existence of a family of special frequency distributions that can lead to trapping of the energy within the attachment, leading to a phenomenon called near-irreversibility.[11] A common characteristic of this family of frequencies is the presence of a singularity or frequency concentration point in their distribution.[11–13]

A rather intriguing aspect of the energy-exchange phenomenon investigated here relates to the distribution of energy among the resonators and the frequencies of the oscillators. For most frequency distributions, the energy transferred to the attachment is largely confined to a limited number of resonators.[8,9] In these cases, after an interval, the duration of which is theoretically predicted in Ref. 8, the resonators become in-phase with one another and the energy is suddenly returned to the master. However, this energy return effect is not observed for those special frequency distributions introduced in Refs. 9 and 12, energy remains within the set of oscillators and is spread over the resonators rather uniformly.[9]

The link between the optimal frequency distributions defined in Ref. 12 and flow of energy from the master to the attached resonators is the subject of the present paper. Of particular interest is the idea that the optimal frequency distributions are akin to a requirement of energy equipartition among the degrees of freedom of the system, or over its modes, which maximizes the trapped energy within the at-

[a)]Author to whom correspondence should be addressed. Electronic mail: a.carcaterra@dma.ing.uniromal.it

 0001-4966/2009/126(1)/122/7/$25.00

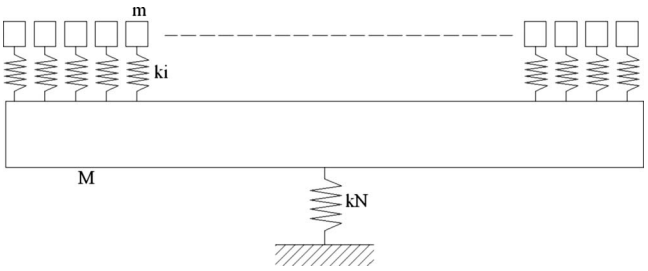


FIG. 1. Schematic of the master and the complex attachment.

tachment. When established, such a link can clarify how energy equipartition allows the master to keep only a small fraction of the total energy, thus having the attachment act as an effective vibration absorber.

The notion of energy equipartition in dynamics has deep roots and strong analogies in thermodynamics. For instance, in molecular mechanics, it is well known that the condition of thermalization, characterized by a uniform distribution of the energy among the molecules, is a condition to which physical systems approach as a consequence of the maximum entropy principle. Reaching thermalization implies that the system has gained its equilibrium and its macro-scale energy distribution has become stable. Based on this notion, this paper hypothesizes that if the oscillations of the attached resonators can reach a state of thermalization by a suitable selection of their natural frequencies, it can then be expected that the system will have a stable energy distribution. In other words, thermalization of the attached oscillators will avoid any periodic energy transfer between the master and the attached oscillators, *de facto* leading to an irreversible energy transfer as discussed in but following different criteria than those reported in Ref. 9 and in Refs. 11–13.

In Sec. II, the question of energy equipartition is considered using modal energies; in Sec. III, the problem is reexamined in terms of energy equipartition among the degrees of freedom of the system. As shown in Appendix, for the particular system investigated here, both forms of energy equipartition requirements are substantially the same.

II. FREQUENCY DISTRIBUTIONS THAT LEAD TO MODAL ENERGY EQUIPARTITION

The equations of motion of the system represented in Fig. 1 are

$$m\ddot{x}_j + k_j(x_j - x_N) = 0, \quad j = 1,2,\ldots,N-1,$$

$$M\ddot{x}_N + k_N x_N + \sum_{j=1}^{N-1} k_j(x_N - x_j) = 0, \tag{1}$$

where index N represents the master and $1,2,\ldots,N-1$ represent the oscillators of the attachment; m, and k_j are the mass and the stiffness of each oscillator of the attachment, and M and k_N represent the mass and the stiffness of the master, respectively; x_j is the displacement of the j-th oscillator. Expressing Eq. (1) in matrix form:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{0}, \tag{2}$$

where $\mathbf{M},\mathbf{K}$ are the mass the stiffness matrices. The use of modal coordinates $\boldsymbol{\eta}$ through the eigenvector matrix $\mathbf{U}$ produces

$$\mathbf{x} = \mathbf{U}\boldsymbol{\eta}. \tag{3}$$

Expressing the modal energies as

$$E_j = \tfrac{1}{2}(\dot{\eta}_j^2 + \omega_j^2 \eta_j^2) \tag{4}$$

where ω_j^2 are the eigenvalues of the system, for an initial impulse MV_0 imparted to the master, Eq. (4) takes the following form:

$$E_j = \frac{V_0^2}{2}(\Psi_{jN})^2, \tag{5}$$

where $\boldsymbol{\Psi} = \mathbf{U}^{-1}$. Modal energies depend explicitly on the system eigenvectors, and indirectly on the set of physical parameters m, k_j, M, and k_N of the system. M and k_N are given, as well as the total mass of the attachment $m(N-1)$, a small fraction of the master mass M. Therefore, E_j varies with the values of k_j $(j=1,2,\ldots,N-1)$ or equivalently depends on the set of the uncoupled natural frequencies $\Omega_j = \sqrt{k_j/m}$ of the attached oscillators.

Modal energy equipartition where the total energy E_{tot} spreads uniformly over the modes of the system can be expressed as

$$E_j(\Omega_1,\ldots,\Omega_{N-1}) = E_j(\boldsymbol{\Omega}) = \frac{E_{\text{tot}}}{N}, \quad j = 1,2,\ldots,N. \tag{6}$$

The frequencies Ω_j that lead to modal equipartition can be obtained by applying a least squares procedure to minimize the error function ε:

$$\varepsilon(\boldsymbol{\Omega}) = \sum_{j=1}^{N}\left[E_j(\Omega) - \frac{E_{\text{tot}}}{N}\right]^2. \tag{7}$$

The algorithm starts with an initial guess $\boldsymbol{\Omega}^{\text{in}}$ for the frequency distribution and stops when a specified convergence criterion is satisfied.

For the three different initial guesses shown in Fig. 2, the final distribution obtained through the minimization algorithm is the same for each, as shown in Fig. 3. Apparently, the results do not depend on the initial estimate for the frequency distribution. The optimal distribution is characterized by an inflection point in the neighborhood of the master frequency where its slope is close to zero. As a consequence, the modal density has a sharp peak around the master frequency, same as for those obtained in Ref. 12, but in this case using different optimization criteria.

In Fig. 4, the modal energy spectra corresponding to the three initial guess distributions are plotted, while the flat line represents the spectrum related to the optimal distribution that produces equipartition of the modal energies, determined through minimization of ε.

In Figs. 5 and 6, energy-time histories of the master are plotted for linear and optimal distributions, respectively (time is normalized with respect to the uncoupled natural

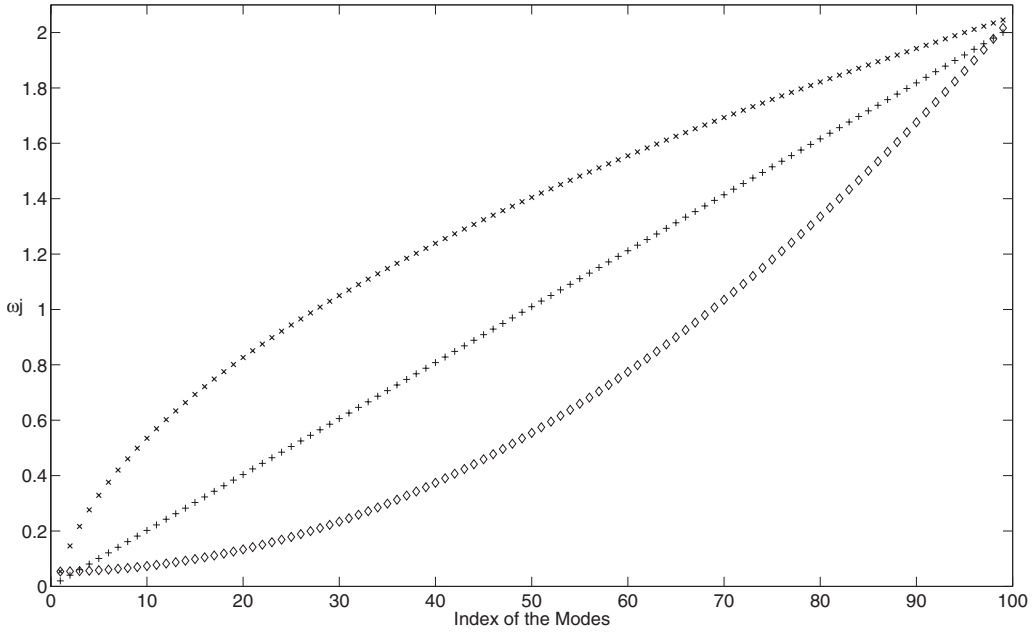


FIG. 2. Initial guesses for frequency distributions of the secondary structure with 99 degrees of freedom: +, ×, and ◇ are for the linear, exponential, and quadratic distributions, respectively.

frequency of the master). These figures demonstrate how an attachment with the optimal frequency distribution is able to minimize the energy stored in the master: after an initial transient, the secondary structure acts as an apparent damper and absorbs, almost completely, the total energy in the system.

III. ENERGY EQUIPARTITION AMONG THE OSCILLATORS OF THE ATTACHMENT

The theoretical developments in this section attempt to provide insight to the notion that frequency distributions leading to energy equipartition have particular forms. As shown in Sec. IV and further explained below, the main characteristic of these frequency distributions is the presence of a minimum slope around the master frequency that also corresponds to a large peak in the associated modal density.

The following theoretical analysis considers the requirement for energy equipartition among the oscillators of the attachment instead of among the modal energies of the system, which was considered in Sec. II. The connection between these two approaches will be discussed later in this section.

As shown in Ref. 12, Eq. (1) can be approximated by a continuous distribution of oscillators attached to the master, replacing the summation by an integral, and the index i by a continuous variable ξ:

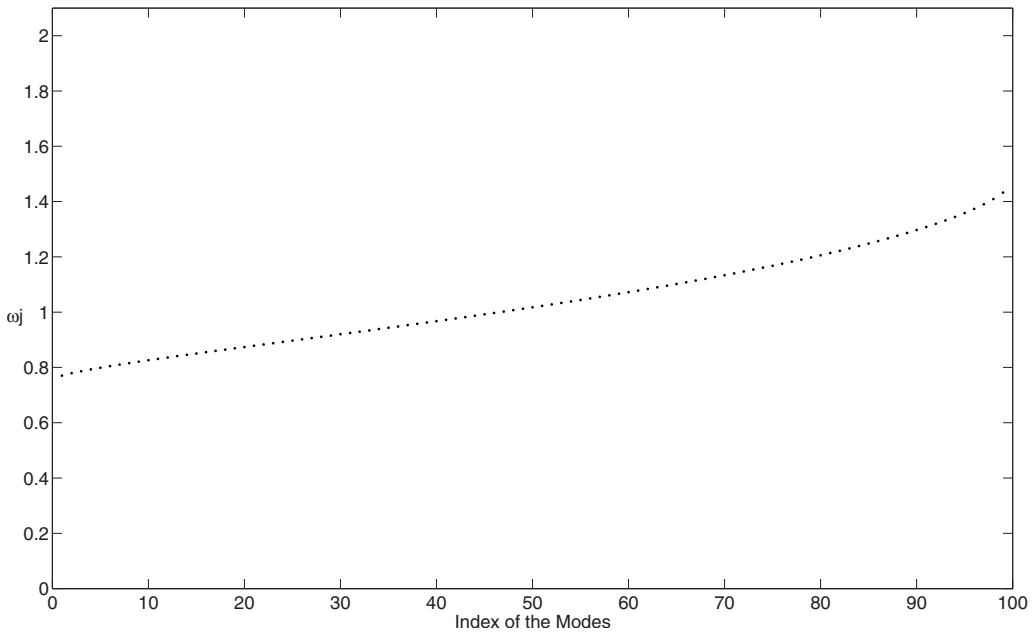


FIG. 3. Optimal frequency distribution in the attachment.

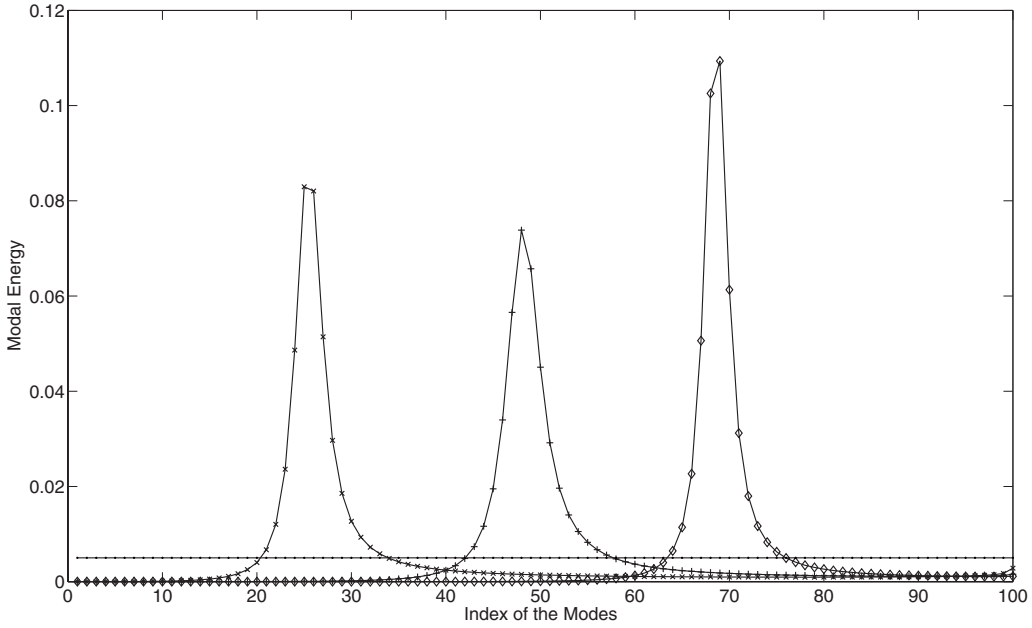


FIG. 4. Modal energy spectra related to the selected frequency distributions; +, ×, and ◇ are for the linear, exponential, and quadratic distributions, respectively.

$$M\ddot{x} + k_N x + \int_0^1 k(\xi)[x - x(\xi)]d\xi = f,$$

$$m(\xi)\ddot{x}(\xi) - k(\xi)[x - x(\xi)] = 0. \tag{8}$$

A detailed discussion about the meaning and the limits of this approximation is given in Refs. 8 and 12. Physically, such an approximation implies that the finite set of resonators is replaced by an infinite set, and thus as N approaches infinity, the frequency gap between neighboring resonators vanishes. Expressing $x(\xi)$ in terms of x in the second equation of Eq. (8), and substituting it into the first equation in Eq. (8), leads to an explicit expression for x, producing an approximate frequency domain counterpart of Eqs. (8), see Ref. 12:

$$-M\omega^2 X + k_N X + j\omega C_{\text{eq}}(\omega)X = F,$$

$$-\rho_0\omega^2 X(\xi) + k(\xi)[X(\xi) - X] = 0, \tag{9}$$

where F, X, and $X(\xi)$ are the Fourier transforms of f, x, and $x(\xi)$, respectively, and a uniform mass distribution $m(\xi)=\rho_0$ is assumed. The equivalent damping is represented as

$$C_{\text{eq}}(\omega) = \rho_0 \frac{\pi}{4}\omega^2 \left.\frac{1}{d\omega_n(\xi)/d\xi}\right|_{\omega_n=\omega_M},$$

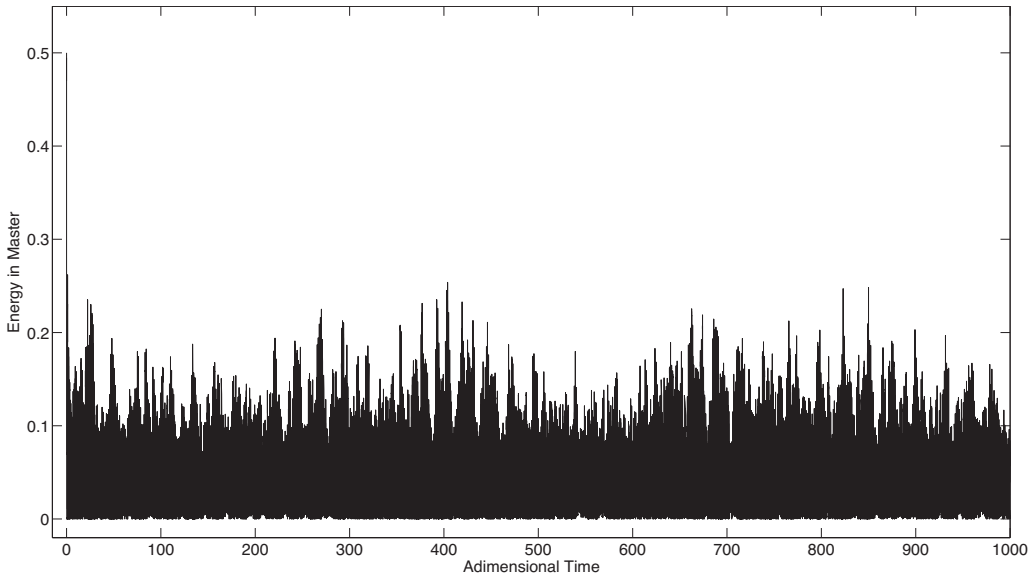


FIG. 5. Time history of the master energy plotted for linear distribution of the uncoupled frequencies of the satellite resonators, $N=100$; time is normalized with respect to the highest modal period: T_1. The non-dimensional energy of the complete system is 0.5.

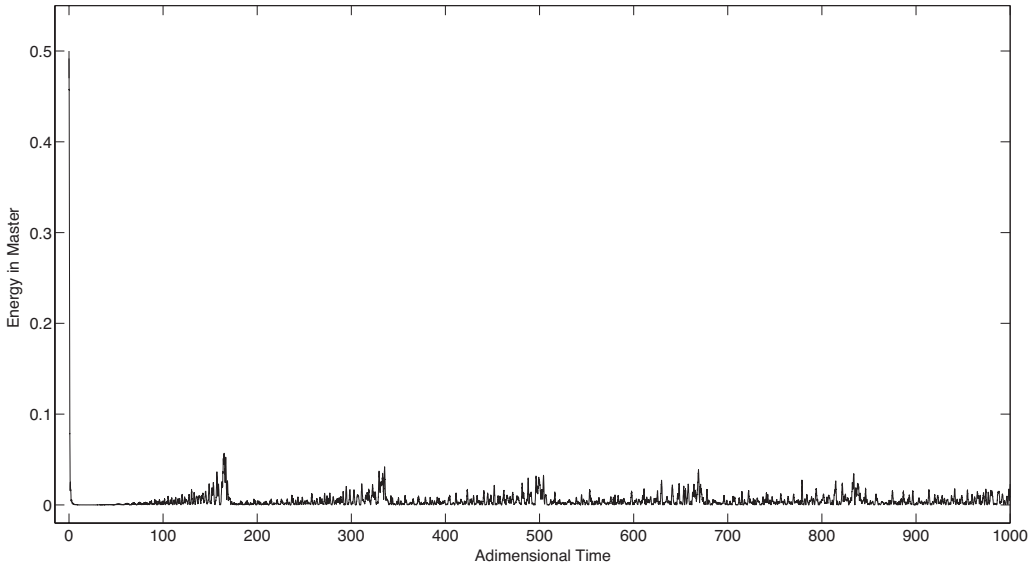


FIG. 6. Time history of the master energy plotted for optimal frequency distribution of the uncoupled frequencies of the satellite resonators, $N=100$.

$$\text{with } \omega_n(\xi) = \sqrt{\frac{k(\xi)}{\rho_0}}$$

as the natural frequency distribution within the attachment. Solutions to Eq. (9) are

$$X = \frac{F}{-M\omega^2 + k_N + j\omega C_{\text{eq}}(\omega)},$$

$$X(\xi) = \frac{\omega_n^2(\xi)}{\omega_n^2(\xi) - \omega^2} \frac{F}{-M\omega^2 + k_N + j\omega C_{\text{eq}}(\omega)}.$$

Expressing the impulsive force as $F=MV_0$, with $\omega_M = \sqrt{k_N/M}$ and $\mu=\rho_0/M$, the total energy distribution $E(\xi)$ within the attachment is found by doubling the potential energy at any ξ:

$$E(\xi) = k(\xi)|X(\xi) - X|^2 = \frac{1}{\omega_n^2} MV_0^2 \mu \frac{\omega^4 \omega_n^2 \omega_M^2}{(\omega_M^2 - \omega^2)^2 + \left[\frac{\omega C_{\text{eq}}(\omega_M)}{M}\right]^2} \frac{1}{(\omega_n^2 - \omega^2)^2}$$

or in non-dimensional form:

$$e(\xi) = \frac{E(\xi)}{\frac{1}{\omega_n^2} mV_0^2} = \frac{\omega^4 \omega_n^2 \omega_M^2}{(\omega_M^2 - \omega^2)^2 + \left[\frac{\omega C_{\text{eq}}(\omega)}{M}\right]^2} \frac{1}{(\omega_n^2 - \omega^2)^2}, \tag{10}$$

where $\omega_n(\xi)$ is replaced for simplicity by ω_n. Equation (10) expresses the energy distribution in the attachment at any frequency ω and depends directly on the frequency distribution $\omega_n(\xi)$ and on its derivative $d\omega_n(\xi)/d\xi|_{\omega_n=\omega}$ through the expression for C_{eq}. The total energy over a frequency bandwidth B then becomes

$$\bar{e}(\xi) = \int_B \frac{\omega^4 \omega_n^2 \omega_M^2}{(\omega_M^2 - \omega^2)^2 + \left[\frac{\omega C_{\text{eq}}(\omega)}{M}\right]^2} \frac{1}{(\omega_n^2 - \omega^2)^2} d\omega. \tag{11}$$

Invoking Parseval's theorem for the equivalence of frequency- and time-averaging, energy expression in Eq. (11) can also be interpreted as the time average energy of an elemental oscillator located at ξ. Requiring $\bar{e}(\xi)$ to be independent of ξ, with a constant value $(\bar{e}_0)$ across the attachment, is equivalent to having the energy equally spread over all degrees of freedom $x(\xi)$:

$$\int_B \frac{\omega^4 \omega_n^2 \omega_M^2}{(\omega_M^2 - \omega^2)^2 + \left[\frac{\omega C_{\text{eq}}(\omega)}{M}\right]^2} \frac{1}{(\omega_n^2 - \omega^2)^2} d\omega = \bar{e}_0. \tag{12}$$

The functional relationship in Eq. (12) can be solved for $\omega_n(\xi)$ numerically within the bandwidth B. However, a reduced form of Eq. (12) reveals special properties of the solution $\omega_n(\xi)$ around the master frequency. Considering a narrow bandwidth B about ω_M and retaining only the zero-order term of the integrand yields

$$\frac{B\omega_M^6 \omega_n^2}{\left[\frac{\omega_M C_{\text{eq}}(\omega_M)}{M}\right]^2} \frac{1}{(\omega_n^2 - \omega_M^2)^2} \approx e_0, \tag{13}$$

which is valid for $\omega_n(\xi) \in B$, i.e., for $\omega_n(\xi)$ close to ω_M.

Local properties of the frequency distribution can be investigated in terms of the related natural frequency (or modal) density n in the attachment. Considering the number $dN=Nd\xi$ of natural frequencies within the interval $d\omega_n$, the associated modal density becomes $n(\omega_n)=dN/d\omega_n = Nd\xi/d\omega_n$, which appears explicitly in Eq. (13) through the relationship $\omega_M C_{\text{eq}}(\omega_M)/M = \mu(\pi/4)\omega_M^3 n(\omega_M)/N$. The corresponding expression for $n(\omega_M)$ from Eq. (13) then follows as:

$$n(\omega_M) = \frac{4BN\omega_n}{\pi\mu\bar{e}_0|\omega_n^2 - \omega_M^2|}. \tag{14}$$

The modal density at the master frequency as given by Eq. (14) is almost singular since $\omega_n(\xi) \approx \omega_M$. This is exactly the same property of those frequency distributions found in Refs. 9–13 that makes the set of attached resonators a highly effective vibration absorber. When a modal density has such a singularity as that in Eq. (14), expression (10) can be used to show how the master energy vanishes, letting the energy injected into the system almost completely migrate toward the attachment.

Finally, it can be shown that the energy equipartition requirement is the same as the requirement that the modal energies and the time (or frequency) averages of the oscillator energies are equal. The energy of oscillator j, expressed as twice that of its kinetic energy, $\Gamma_j = m\overline{\dot{x}_j^2(t)}$, where the bar represents the time average, is also equivalent to its average value in the frequency domain, as shown in Eq. (11). Noting that the modal energy expression E_j in Eq. (4) or Eq. (5) is also independent of time, a comparison of E_j and Γ_j can be made using the coordinate transformation in Eq. (3). For the special system under consideration here, where a large number of oscillators are attached to the master in a parallel manner with a total mass small compared to that of the master, modes are localized. For such systems the matrix $\mathbf{U}$ is almost diagonal and the physical and the modal coordinates lead to energies E_j and Γ_j that are substantially similar. As expected, numerical results also show a strong mode localization with an almost diagonal form for the eigenvector matrix $\mathbf{U}$. A proof of the equivalence between E_j and Γ_j for the system described here is presented in Appendix.

IV. DISCUSSION AND CONCLUSIONS

Earlier studies had shown that vibration energy of a structure can be absorbed nearly irreversibly by a complex attachment that consists of a large number of simple oscillators with the requirement that the attached oscillators possess a particular frequency distribution. These frequency distributions were shown to have a higher modal density around the natural frequency of the master structure. Their distributions were obtained through a variational approach that minimizes the energy associated with the master structure.

The particular form of the frequency distribution deserves a comment on why the frequencies are densely distributed around the frequency to be suppressed and not collocated with it. Selecting the uncoupled frequencies of all the attached oscillators to be the same as that of the natural frequency of the master amounts to constructing a classical tuned absorber that has two degrees of freedom. Considering that the proposed system is conservative, in such a case, an impulse applied to the master would produce a response characterized by two close natural frequencies resulting in a modulation that represents a periodic energy exchange between the master and the satellites that move in unison. However, satellites that are nearly-resonating with the master allow a strong coupling and avoid the simple beat phenomenon described above. The out-of-phase responses are produced by the spread of the resonator frequencies in a small band around the master frequency. The consequence is that soon after they absorb the initial impulse, the oscillators rapidly develop an out-of phase-motion and their total reaction on the master vanishes because of the incoherence of their phases. In a sense, the optimality of the frequency distribution is driven by a compromise between a near-resonant condition and an out-of-phase requirement, leading to the typical frequency form described in this and previous papers.

This paper shows how the frequency distributions obtained previously using a variational approach that minimizes the master energy also result from or are equivalent to an energy equipartition requirement within the system. Finally, an unexpected but significant result for systems as that considered here is that the requirement of energy equipartition among the modal energies is the same as an equipartition among the physical degrees of freedom.

In conclusion, the energy equipartition requirement on the prototypical system described here stores most of the energy in the attachment, leaving $1/N$ of the total energy in the master, making the attachment an effective energy sink that produces a high damping effect in the master motion.

APPENDIX: EQUIPARTITION AMONG THE MODES AND AMONG THE OSCILLATORS

Displacement and velocity in expressions for modal energies E_j are

$$x_i(t) = \sum_{j=1}^{N} \frac{U_{ij}\sqrt{2E_j}}{\omega_j} \sin(\omega_j t),$$

$$\dot{x}_i(t) = \sum_{j=1}^{N} U_{ij}\sqrt{2E_j} \cos(\omega_j t). \tag{A1}$$

The energy of the master is given as

$$E_N(t) = \tfrac{1}{2}M(\dot{x}_N^2(t) + \omega_M^2 x_N^2(t)).$$

The time-averaged E_N becomes

$$[E_N] = \lim_{t\to\infty} \frac{1}{T}\int_0^T E_N(t)dt = \frac{1}{2}M([\dot{x}_N^2] + \omega_M^2[x_N^2]),$$

which through Eq. (A1) becomes

$$[E_N] = E_{\text{tot}}\sum_{j=1}^{N} \frac{U_{Nj}^2(\psi_{jN})^2}{2}\left(1 + \frac{\omega_M^2}{\omega_j^2}\right),$$

where $E_{\text{tot}} = MV_0^2/2$ is the total energy of the system.

The time-average of energy for each resonator of the attachment can be expressed as twice its mean kinetic energy:

$$[E_i] = m\sum_{j=1}^{N} U_{ij}^2 E_j.$$

Orthonormality conditions permits expressing E_j in terms of U_{Nj}:

$$\mathbf{U}^T\mathbf{M}\mathbf{U} = \mathbf{I} \rightarrow (\mathbf{U}^{-1})^T = \mathbf{M}\mathbf{U} \rightarrow [\mathbf{U}^{-1}]_{jN} = M[\mathbf{U}]_{Nj},$$

which when substituted into Eq. (5) yields

$$E_j = ME_{\text{tot}}U_{Nj}^2.$$

Thus, the mean time energies can be expressed as

$$[E_N] = \frac{1}{2}\sum_{j=1}^{N} \frac{E_j^2}{E_{\text{tot}}}\left(1 + \frac{\omega_M^2}{\omega_j^2}\right),$$

$$[E_i] = n\sum_{j=1}^{N} U_{ij}^2 E_i. \tag{A2}$$

If energy is equally distributed among all the modes such that $E_j = E_{\text{tot}}/N$, for the master-attachment system with an optimal frequency distribution described in Sec. II, the first equation of Eq. (A2) becomes

$$[E_N] = \frac{1}{2}\frac{E_{\text{tot}}}{N^2}\left(N + \sum_{j=1}^{N} \frac{\omega_M^2}{\omega_j^2}\right).$$

The natural frequencies, as shown in Secs. II and III, have values close to the master frequency, thus allowing an approximation of the summation $\Sigma_{j=1}^{N}\omega_M^2/\omega_j^2$ by N, yielding

$$[E_N] \approx \frac{E_{\text{tot}}}{N}. \tag{A3}$$

Analogously from the second equation of Eq. (A2), one obtains

$$[E_i] = \frac{E_{\text{tot}}}{N}\left(m\sum_{j=1}^{N-1} U_{ij}^2 + \frac{\alpha}{N}\right)$$

Using the orthonormality conditions, $m\Sigma_{j=1}^{N-1}U_{ij}^2 = 1 - (1/N)$:

$$[E_i] \approx \frac{E_{\text{tot}}}{N}. \tag{A4}$$

Equations (A3) and (A4) show that the optimal frequency distribution produces energy equipartition over the modes as well as over the resonators of the attachment.

[1] A. D. Pierce, V. W. Sparrow, and D. A. Russel, "Fundamental structural-acoustic idealization for structure with fuzzy internals," J. Vibr. Acoust. **117**, 339–348 (1995).
[2] M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).
[3] R. L. Weaver, "The effect of an undamped finite degree of freedom 'fuzzy' substructure: Numerical solution and theoretical discussion," J. Acoust. Soc. Am. **101**, 3159–3164 (1996).
[4] A. Carcaterra and A. Akay, "Damping device," International Patent No. WO 2006/103291 A1.
[5] R. L. Weaver, "Mean and mean square responses of a prototypical master/fuzzy system," J. Acoust. Soc. Am. **101**, 1441–1449 (1997).
[6] R. L. Weaver, "Equipartition and mean square response in large undamped structures," J. Acoust. Soc. Am. **110**, 894–903 (2001).
[7] G. Maidanik, "Induced damping by a nearly continuous distribution of a nearly undamped oscillators: linear analysis," J. Sound Vib. **240**, 717–731 (2001).
[8] A. Carcaterra and A. Akay, "Transient energy exchange between a primary structure and a set of oscillators: Return time and apparent damping," J. Acoust. Soc. Am. **115**, 683–696 (2004).
[9] I. M. Koç, A. Carcaterra, Z. Xu, and A. Akay, "Energy sinks: Vibration absorption by an optimal set of undamped oscillators," J. Acoust. Soc. Am. **118**, 3031–3042 (2005).
[10] A. Akay, Z. Xu, A. Carcaterra, and I. M. Koç, "Experiments on vibration absorption using energy sinks," J. Acoust. Soc. Am. **118**, 3043–3049 (2005).
[11] A. Carcaterra, A. Akay, and I. M. Koç, "Nearly irreversible energy trapping by an undamped continuous structure with singularity points in its modal density," J. Acoust. Soc. Am. **119**, 2141–2149 (2006).
[12] A. Carcaterra and A. Akay, "Theoretical foundation of apparent damping and energy irreversible energy exchange in linear conservative dynamical systems," J. Acoust. Soc. Am. **121**, 1971–1982 (2007).
[13] A. Carcaterra, A. Akay, and F. Lenti, "Pseudo-damping in undamped plates and shells," J. Acoust. Soc. Am. **122**, 804–813 (2007).

Shaping of a system's frequency response using an array of subordinate oscillators

Joseph F. Vignola[a] and John A. Judge
Department of Mechanical Engineering, The Catholic University of America, 620 Michigan Avenue, NE, Washington, DC 20064

Andrew J. Kurdila
Department of Mechanical Engineering, Virginia Polytechnic University and State University, Blacksburg, Virginia 24061

(Received 12 December 2008; revised 9 April 2009; accepted 5 May 2009)

The frequency response of an oscillating structure can be tailored by attaching one or more subordinate oscillators. This paper shows how the magnitude and phase of the frequency response can be deliberately shaped by prescribing the distributions of the dynamic properties in an array of such subordinate oscillators. Exact analytic governing equations of motion are derived for the coupled system composed of the primary system and the subordinate array. For a relatively small number (<100) of attached oscillators whose total mass is small ($<1\%$) relative to the primary structure, it is possible to engineer frequency-response functions of the primary oscillator to have, for example, nearly linear phase or constant amplitude over a frequency band of interest. The frequency range over which response shaping is achieved is determined by the band of the attached oscillators. It is shown that the common analytic methodology for designing a dynamic vibration absorber represents the limiting case of a single oscillator in the subordinate set. Moreover, increasing the number of subordinate oscillators (without increasing the total added mass) offers a number of advantages in reshaping the dominant system's frequency response.
 [DOI: 10.1121/1.3143783]

PACS number(s): 43.40.At, 43.40.Kd, 43.40.Tm [DF] Pages: 129–139

I. INTRODUCTION

For some time there has been interest in understanding a class of systems in which the response of some dominant structure is altered by the presence of a number of small substructures. The addition of numerous small substructures can create apparent damping[1] and dramatically alter the nature of the dominant structure's resonance. Various approaches have been used to model the added complexity. The field of fuzzy structures (see Refs. 2–4) has emerged as one attempt at defining a systematic approach to modeling complex systems, in which small-scale features of a system are treated with imprecise statistical or "fuzzy" representations.[5] An alternative approach is to prescribe functional distributions for parameters defining the small-scale features and treat the system classically. Here, we follow the latter approach and show that particular parameter distributions can be chosen that allow the spectral response of the system to be tailored in desirable ways.

Suppose we consider a linear, time-invariant system that is complex in the sense that a large number of degrees of freedom is required to characterize every constituent subsystem. We further restrict ourselves to the case where the system is composed of a primary or host structure, modeled as a simple oscillator, that is connected to a large number of additional simple oscillators whose frequencies are of the same order as that of the primary structure. We refer to the collection of additional masses as a subordinate oscillator array (SOA) and will use a precise description of the properties within this subsystem. For practical purposes, we are interested in the case in which the mass of each of the individual oscillators composing the SOA is much less than the mass of the primary structure. Suppose μ is the ratio of the *total* mass in the subordinate system normalized to the mass of the primary oscillator. We examine the behavior of systems for which the mass ratio μ is between 0 and 1. Various aspects of this problem have been studied within the vibrations and acoustics community.[1,3,6–10] References 1, 3, 8, and 10 focus on the characterization of effective damping, the study of the asymptotic behavior of effective damping as the dimension of the subordinate array approaches infinity, or phenomenological ideas such as the "return time" after which energy returns from the subordinate array to the primary structure due to coherent motion of the subordinate oscillators. References 6 and 7 discuss several convergence issues associated with passing from a discrete formulation of the system dynamics to a continuous parametrization.

This paper addresses a distinct, previously unexplored aspect of this problem. We show how prescribing the distribution of the dynamic properties of the oscillators in the SOA can be used to design and tailor the overall system response. In Sec. II, we derive, in closed form, a transfer function and frequency-response function (FRF) relating the motion of the primary oscillator to an applied force. In Sec. III, we define functional distributions for the mass and stiff-

[a]Author to whom correspondence should be addressed. Electronic mail: vignola@cua.edu

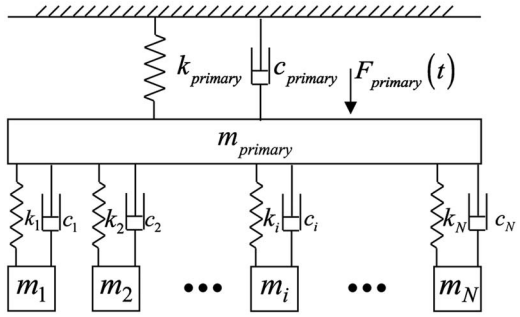


FIG. 1. Model of a dominant oscillator with SOA attached.

ness of the subordinate oscillators and show their effect on the spectrum of the response. In Sec. IV, we relate the possible changes in the system's frequency spectra to the flow of energy between the primary and subordinate elements. The effect of the subordinate array on the system response depends on the mass ratio; we explore various response regimes that change with μ in Sec. V. We close with a discussion of how the number and bandwidth of oscillators in the subordinate array relates to the smoothness of the spectra and the exchange of energy between elements, followed by general conclusions about the applicability of this concept to oscillatory systems.

II. DISCRETE MODEL AND EQUATIONS OF MOTION

The system is assumed to be linear and time invariant. We derive transfer functions and FRFs in terms of properties of the primary structure and a distribution of the properties of the subordinate oscillators. This analysis may be viewed as a generalization of the so-called dynamic vibration absorber (DVA) that is described in many classical texts.[11–13] Attributes of an oscillator with a SOA will be discussed in relation to an oscillator with a DVA.

A classical approach is used to model a system composed of a primary single degree of freedom (SDOF) oscillator with an array of N smaller attached oscillators, as shown in Fig. 1. Individual elements are modeled as simple mass-spring-dampers. Each element of the array has a mass of m_n and is supported by a spring with stiffness k_n and a viscous damper with coefficient c_n. The primary oscillator has mass of m_{primary} and is supported by a spring of stiffness k_{primary} and a viscous damper with coefficient c_{primary}. Equations of motion for each of the oscillators are found using Newton's second law. When we restrict ourselves to the case where the external force is only applied to the primary oscillator, the equation of motion the nth subordinate oscillator is

$$m_n\ddot{x}_n + c_n\dot{x}_n - c_n\dot{x}_{\text{primary}} + k_n x_n - k_n x_{\text{primary}} = 0 \tag{1}$$

and the equation of motion for the primary oscillator becomes

$$m_{\text{primary}}\ddot{x}_{\text{primary}} + \left(\sum_{n=1}^{N} c_n + c_{\text{primary}}\right)\dot{x}_{\text{primary}} - \sum_{n=1}^{N} c_n\dot{x}_n + \left(\sum_{n=1}^{N} k_n + k_{\text{primary}}\right)x_{\text{primary}} - \sum_{n=1}^{N} k_n x_n = F_{\text{primary}}(t). \tag{2}$$

These $N+1$ equations can be written in matrix form as

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{F}(t). \tag{3}$$

The inertia matrix, $\mathbf{M}$, is diagonal and composed solely of the masses of the system elements, while the damping and stiffness matrices, $\mathbf{C}$ and $\mathbf{K}$, have off-diagonal terms coupling the motion of the various oscillators.

$$\mathbf{M} = \begin{pmatrix} \underset{n=1\cdots N}{\text{diag}}\{m_n\} & \mathbf{0} \\ \mathbf{0} & m_{\text{primary}} \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} \underset{n=1\cdots N}{\text{diag}}\{c_n\} & -\mathbf{c} \\ -\mathbf{c}^T & c_{\text{primary}} \end{pmatrix},$$

$$\mathbf{K} = \begin{pmatrix} \underset{n=1\cdots N}{\text{diag}}\{k_n\} & -\mathbf{k} \\ -\mathbf{k}^T & k_{\text{primary}} \end{pmatrix}. \tag{4}$$

In Equation (4), $\mathbf{c}$ and $\mathbf{k}$ are $N\times 1$ column vectors, while $\mathbf{c}^T$ and $\mathbf{k}^T$ are their transposes. The force and displacement vectors are given by

$$\mathbf{F}(t) = \begin{pmatrix} \mathbf{0} \\ F_{\text{primary}}(t) \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_{\text{sub}} \\ x_{\text{primary}} \end{pmatrix}. \tag{5}$$

Taking the Laplace transform of Eq. (3) yields

$$\begin{bmatrix} \underset{n=1\cdots N}{\text{diag}}\{s^2m_n + sc_n + k_n\} & -s\mathbf{c} - \mathbf{k} \\ -s\mathbf{c}^T - \mathbf{k}^T & s^2m_{\text{primary}} + s\sum_{n=1}^{N+1} c_n + \sum_{n=1}^{N+1} k_n \end{bmatrix} \times \begin{pmatrix} \mathbf{x}_{\text{sub}}(s) \\ X_{\text{primary}}(s) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ F_{\text{primary}}(s) \end{pmatrix}. \tag{6}$$

Here the $N+1$ terms correspond to the primary quantity. It should be recognized that $\text{diag}_{n=1\cdots N}\{s^2m_n+sc_n+k_n\}$ is easily invertible because it is a diagonal matrix. Equation (6) can be solved for the transform of the motion of the SOA elements [the vector $\mathbf{x}_{\text{sub}}(s)$] in terms of the motion of the primary oscillator

$$\mathbf{x}_{\text{sub}}(s) = \left(\underset{n=1\cdots N}{\text{diag}}\{s^2m_n + sc_n + k_n\}\right)^{-1}(s\mathbf{c} + \mathbf{k})x_{\text{primary}}(s), \tag{7}$$

allowing derivation of the transfer function relating applied force to primary oscillator displacement:

$$\frac{X_{\text{primary}}(s)}{F_{\text{primary}}(s)} = \left(\left(s^2 m_{\text{primary}} + s\left(\sum_{n=1}^{N} c_n + c_{\text{primary}} \right) + \sum_{n=1}^{N} k_n + k_{\text{primary}} \right) - \sum_{n=1}^{N} \left(\frac{(sc_n + k_n)^2}{s^2 m_n + sc_n + k_n} \right) \right)^{-1}. \tag{8}$$

Equation (8) is closed form expression that depends on the mass, damping, and stiffness of each element of the system.

The FRF is found by setting $s=i\omega$ and combining the summations in the denominator:

$$\frac{X_{\text{primary}}(\omega)}{F_{\text{primary}}(\omega)} = \left(-\omega^2 m_{\text{primary}} + i\omega c_{\text{primary}} + k_{\text{primary}} + \sum_{n=1}^{N} \left(i\omega c_n + k_n - \frac{(i\omega c_n + k_n)^2}{-\omega^2 m_n + i\omega c_n + k_n} \right) \right)^{-1}. \tag{9}$$

This expression can be non-dimensionalized by multiplying both sides by the primary stiffness and defining a non-dimensional frequency $\Omega = \omega/\omega_{\text{primary}} = \omega\sqrt{m_{\text{primary}}/k_{\text{primary}}}$,

$$\frac{X_{\text{primary}}(\Omega) k_{\text{primary}}}{F_{\text{primary}}(\Omega)} = \left(1 - \Omega^2 + \frac{i\Omega}{Q_{\text{primary}}} + \sum_{n=1}^{N} \alpha_n \left(\frac{-\Omega^2 \left(1 + \frac{i\Omega}{\beta_n Q_n} \right)}{1 - \left(\frac{\Omega}{\beta_n} \right)^2 + \frac{i\Omega}{\beta_n Q_n}} \right) \right)^{-1}. \tag{10}$$

Equation (10) characterizes the spectra in terms of the non-dimensional frequency and the non-dimensional property distributions within the SOA elements.

$$\alpha_n = \frac{m_n}{m_{\text{primary}}}, \quad \beta_n = \sqrt{\frac{\gamma_n}{\alpha_n}}, \quad \gamma_n = \frac{k_n}{k_{\text{primary}}}, \quad Q_n = \frac{\sqrt{m_n k_n}}{c_n}. \tag{11}$$

The ratio of the total mass of all subordinate oscillators to the mass of the primary oscillator is $\mu = (\Sigma m_n)/m_{\text{primary}} = \Sigma \alpha_n$. The parameter β_n is the non-dimensional frequency an individual subordinate oscillator would have if isolated from the rest of the system. Note that only two of the three non-dimensional distribution parameter sets α_n, β_n, and γ_n appear in Eq. (10) because of the interdependence on the distribution parameters.

In the limiting case of $N=1$ (a single subordinate oscillator), the above analysis reduces to the well known system commonly referred to as a DVA or a tuned-mass damper (TMD). This example is treated in many classic texts dating back to Timoshenko[11] and Den Hartog[12] as well modern books including Inman.[13] In the case of finite damping ($Q_n < \infty$), the DVA cannot completely suppress vibration of the primary structure as is expected in the limiting case of no damping ($Q_n = \infty$). That is, when $N=1$, $\beta_1 = 1$, and $\Omega = 1$, the FRF in Eq. (10) still has a finite value. However, vibration on a primary structure can be significantly reduced in a frequency band determined by the ratio of the primary and subordinate masses. Other authors including Zuo and Nayfeh[14,15] investigated the use of multiple-tuned mass damper system for vibration isolation and shown the distribution of properties within the system can be designed to improve performance.

III. PROPERTY DISTRIBUTIONS

The FRF for the primary oscillator, Eq. (10), is given as a function of the distributions of frequencies (β_n), masses (α_n), and quality factors (Q_n) within the SOA. The shape of this FRF can be controlled by deliberate selection of distributions for each of these parameters. For the purpose of illustrating the impact of the distributions, a three parameter function given by Eq. (12) is used to define β_n, the distribution of the isolated natural frequencies of the subordinate oscillators. A linear function will define α_n, the distribution of non-dimensional masses within the SOA. A pair of related functions is used to generate the first and second halves of the β_n distribution which is anti-symmetric about the center value:

$$\beta_n = \begin{cases} \frac{\Delta}{2}\left(\left(\frac{2(n-1)}{N-1} \right)^p - 1 \right) + 1 & \text{for } n \le \frac{N}{2} \\ \frac{\Delta}{2}\left(1 - \left(\frac{2(N-n)}{N-1} \right)^p \right) + 1 & \text{for } n \ge \frac{(N+1)}{2}. \end{cases} \tag{12}$$

The normalized frequencies β_n form a band of width Δ centered at 1. Thus, the actual subordinate oscillator frequencies form a band that surrounds the natural frequency that the primary oscillator would have in the absence of the SOA. This distribution is shown in Fig. 2 for a variety of values of p, for the case $N=50$. When the exponent p becomes very small all the subordinate oscillators will have nearly the same isolated natural frequency. In the limit as $p \to 0$, all the subordinate oscillators move in unison, corresponding to a DVA with mass $\Sigma_{n=1}^{N} m_n$. When $p \to \infty$, the system approaches the case where there are only two elements in the SOA and their isolated natural frequencies correspond to the boundaries of the band, $1-\Delta/2$ and $1+\Delta/2$.

The non-dimensional mass distribution, α_n, is given by

$$\alpha_n = \frac{\mu}{N}\left(q\left(\frac{n-1}{N-1} \right) - \frac{q}{2} + 1 \right), \tag{13}$$

where μ is the ratio of the total mass in the subordinate subsystem to the mass of the primary element, and q is a mass slope parameter. This linear distribution is also shown in Fig. 2.

When the mass ratio is very small ($\mu < \Delta/2eQ_{\text{primary}}$, where e=natural log base), the SOA has little impact on the FRF of the primary oscillator. As the mass in the SOA is increased, the magnitude and phase of the FRF change qualitatively from that of a SDOF system. Figure 3 shows an example of this variation as the mass ratio μ increases, for both the SOA system ($N>1$) and the DVM case ($N=1$). The

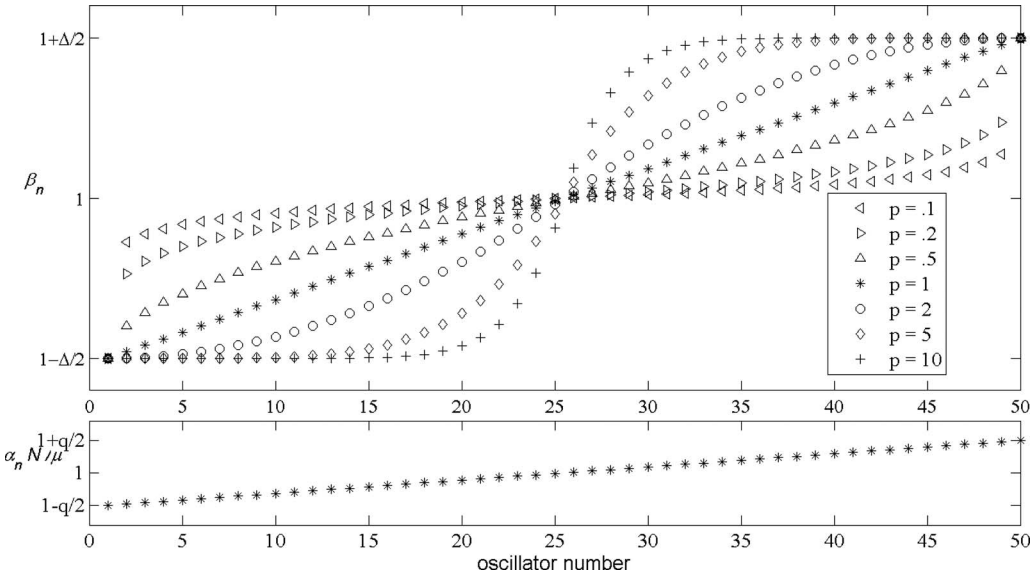


FIG. 2. The family of frequency distributions β_n form a band on monotonically increasing values that surround the natural frequency of the primary oscillator (top). The non-dimensional distribution mass for each element in SOA is defined to be linear (bottom).

frequency band in which the attachments affect the response of the primary oscillator has a different character for the two systems. Inman[13] discussed how, for a DVM, the width of the band of vibration suppression is dependent on the mass ratio. This is because the splitting of the two natural frequencies of the two-DOF system increases as the mass of the attachment increases. For an oscillator with a SOA, initially the bandwidth increases with increasing mass ratio and has the appearance of added damping. This apparent- or pseudo-damping has been described by a number of authors including Carcaterra and co-workers,[1,16–19] Maidanik,[7] and Strasberg and Feit.[8,9] When the mass ratio reaches about $\mu \approx \Delta^2/10$, a new region becomes evident in the spectrum, whose bandwidth is equal to the band of the SOA.

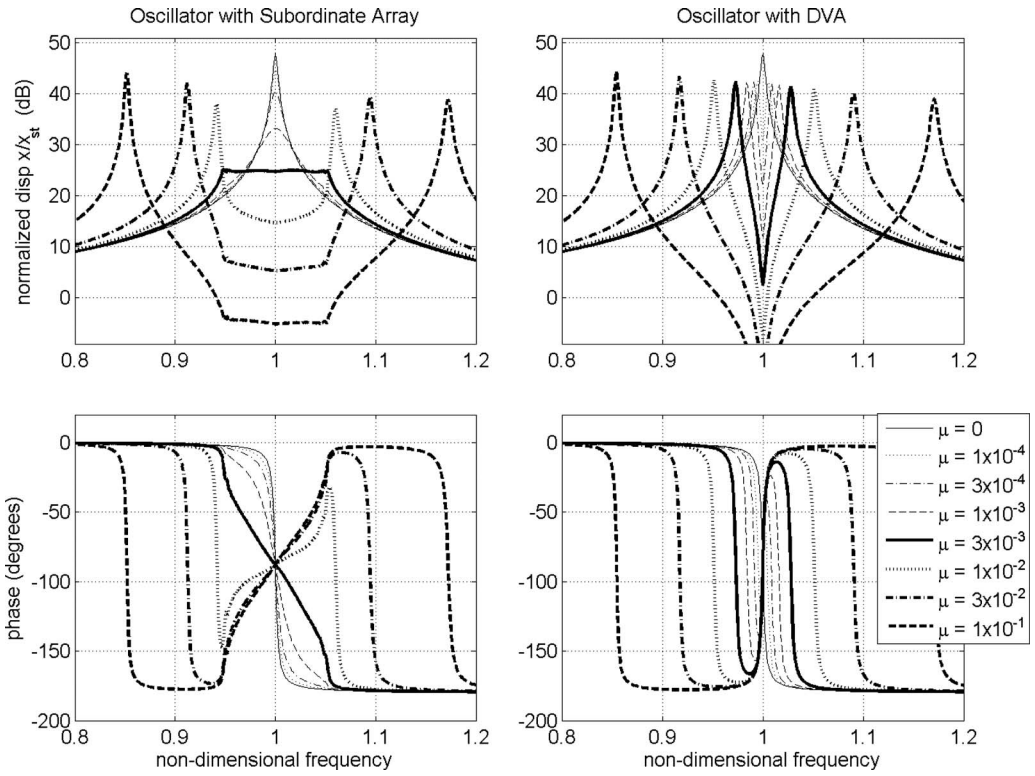


FIG. 3. Comparison of the dynamic displacement, X, normalized by the static displacement, X_{st}, magnitude (upper) and phase (lower) of the FRF of an oscillator with a subordinate array (left) and a DVA (right) as the mass ratio μ is increased. In this example the $\omega_{primary}$ is centered in the band of the subordinate oscillators and $\omega_{primary}=\omega_{DVM}$, quality factor $Q_n=250$ for all elements, the fractional bandwidth $\Delta=0.1$, the number of oscillators in the SOA, $N=50$. The p parameter in the subordinate oscillator frequency distribution [Eq. (12)] and the mass distribution slope q were chosen to achieve a flat response within the band, and are 0.815 and −0.25, respectively.

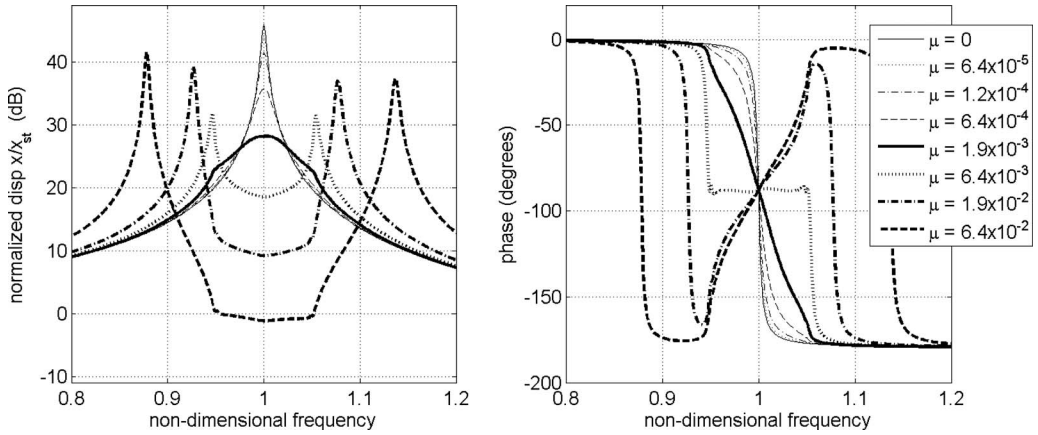


FIG. 4. Sequence of FRF magnitude and phase spectra for an oscillator with attached subordinate array with quality factor lowered to $Q_n=200$. Spectra with a nearly linear phase ($\mu=0.0019$) and nearly constant phase ($\mu=0.0064$) can be seen. As in Fig. 3, the mass distribution slope is $q=-0.25$, the parameter $p=0.815$, and number of oscillators, $N=50$.

The sequence of spectra shown in Fig. 3 indicates that the spectral response develops as the mass ratio increases. The sequence includes one case ($\mu=0.003$) that is nearly spectrally flat in the band of the SOA (in this case, a bandwidth of 10% of the center frequency). Figure 4 shows a slightly modified sequence of spectra where the quality factor of each individual oscillator has been lowered to 200. In this sequence, spectra with nearly linear ($\mu=0.0019$) and constant ($\mu=0.0064$) phase are seen. The nearly flat frequency response can be produced over a range of frequencies. These plots suggest that the SOA cannot only attenuate a resonant response of a structure at a specific frequency like a DVA but can do so over a frequency band defined by the band of the SOA. For each of these FRFs, the number of oscillators in the subordinate system is chosen such that the modal overlap parameter, $\eta \geqslant 2$. This parameter is discussed in the paper by Strasberg and Feit[8] and the book by Lyon[20] and is defined as the ratio of the average width of a spectral peak divided by the number of peaks in a band, and can be expressed as $\eta=N-1/\bar{Q}\Delta$. This condition, $\eta \geqslant 2$, roughly speaking, can also be seen as the minimum number of DOFs so that the spectrum is smooth. The variation of the dynamic response of the primary oscillator with increasing mass ratio is explored further in Sec. IV.

IV. PARTITIONING OF ENERGY

The fraction of energy dissipated by the primary oscillator versus subordinate oscillators is affected by the mass ratio and the other parameters of the system. Consider a scenario in which an impulsive force imparts an initial velocity equal to $1/m_{\text{primary}}(1N\cdot s)$ to the primary oscillator. The total energy introduced into the system via the impulsive force is

$$E_{\text{int}} = \frac{1}{2m_{\text{primary}}}(1N\cdot s)^2. \tag{14}$$

Over time, this energy is distributed to all elements of the system and subsequently is dissipated by the various dampers in the system. For the purpose of characterizing the various effects the SOA can have on the system response, we plot the fraction of the total introduced energy that is dissipated by the primary damper, as a function of mass ratio μ. This energy ratio is calculated by first expressing the velocity spectrum as the product of $i\Omega$ and the displacement spectrum, $v(\Omega)=i\Omega x(\Omega)$ calculated from Eq. (11), and then integrating the product of this velocity spectrum and its complex conjugate: $\int_{-\infty}^{\infty} v(\Omega)v^*(\Omega)d\Omega$. By Parseval's theorem, this is equal to the time integral of the mean square velocity time history of the primary mass ($\int_{-\infty}^{\infty} v(t)^2 dt$). The instantaneous power dissipated by the primary damper equals the product of the force the damper applies on the primary mass $c_{\text{primary}}\dot{x}^2_{\text{primary}}$ and the velocity of the primary mass $\dot{x}_{\text{primary}}$ or $c_{\text{primary}}\dot{x}^2_{\text{primary}}$. The integral over all time equals the total energy dissipated by the primary damper and can be calculated as

$$E_{\text{diss}} = c_{\text{primary}}\int_{-\infty}^{\infty} v(t)^2 dt = c_{\text{primary}}\int_{-\infty}^{\infty} v(\Omega)v^*(\Omega)d\Omega. \tag{15}$$

The ratio of the total energy dissipated by the primary damper [Eq. (15)] to the energy introduced into the system by the impulsive excitation [Eq. (14)] is shown in Fig. 5 as a

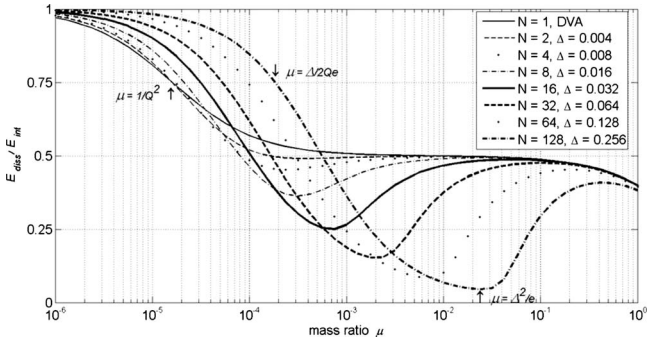


FIG. 5. A collection of curves indicating the fraction of energy introduced by impulsive excitation that is dissipated by the primary damper, as a function of mass ratio. Eight different cases are shown that range from the case of the DVA where $N=1$ to a SOA with a moderate number of elements and a bandwidth of a little more than 25%. In each case the quality factor Q_n is 250 for all elements. For each of these cases, the number of oscillators is chosen to be $N=2\Delta\cdot Q_{\text{primary}}$, which is minimum condition for a smooth spectral response.

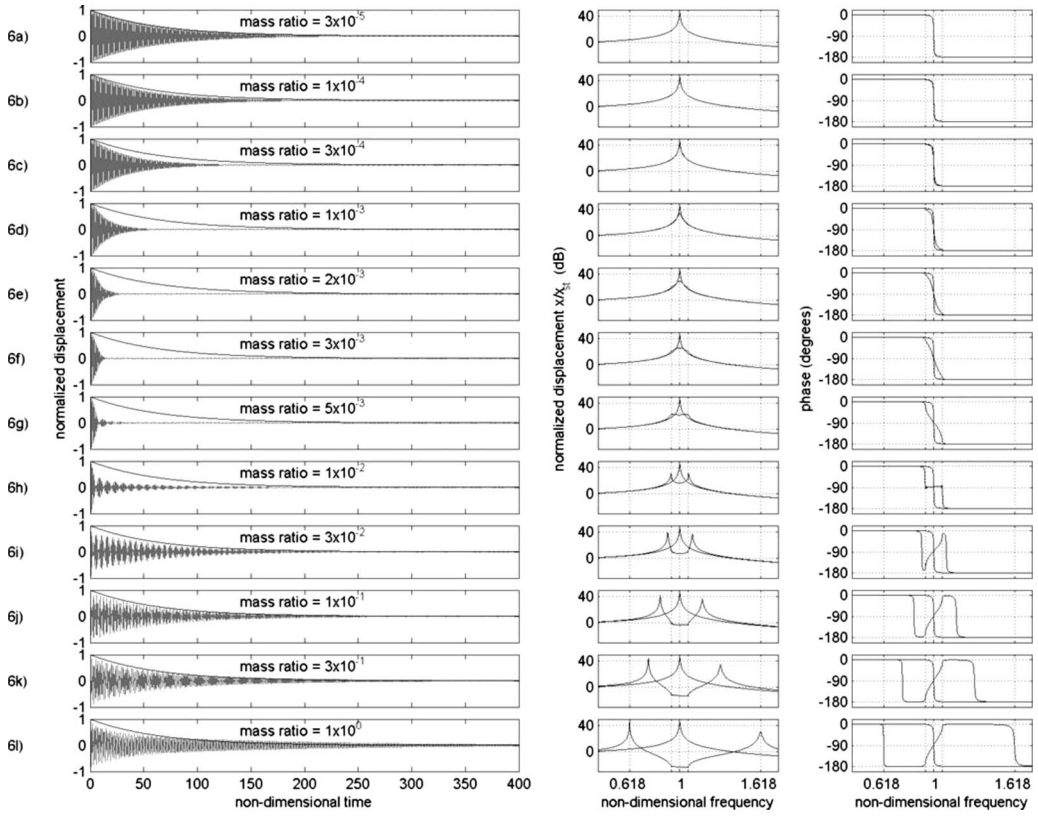


FIG. 6. Impulse responses in time and frequency domain over a range of mass ratios. For each of these plots, $Q=200$, $N=50$, $\Delta=12.5\%$, and $p=0.8$. In each time-history plot, there is an additional curve representing the envelope of the response of an unadorned oscillator (simple one-DOF system) with the same mass, stiffness, and damping. Likewise the magnitude and phase plots have a FRF for the same single degree of freedom oscillator for a reference. The outer tick marks in the frequency spectra indicate the band of a DVA with $\mu=1$. The inner tick marks define the band of the SOA.

function of the mass ratio for different numbers of elements and fractional bandwidths. For the case of the DVA ($N=1$), the attached element has little impact when the mass ratio μ is significantly less than $1/Q^2_{\text{primary}}$. As the mass ratio increases, more of the energy of the system is drawn into the subordinate oscillator, with just over half of the energy being drawn off by the DVA when the added mass is significant. For the SOA case ($N>1$), a region of even greater energy absorption develops at intermediate mass ratios. The curves for the SOA cases can be divided into a number of regimes, characterized by different types of behavior in both the time and frequency domains.

V. TIME RESPONSES AND REGIMES OF INTEREST

As the mass ratio μ changes, the character of the displacement response of the primary mass passes through distinct regimes. For the following discussion, we will assure that all the isolated oscillators, including the primary, have the same quality factor, $Q_n=Q$. A sequence of impulse responses is shown in Fig. 6 where the mass ratio is increased from a very small amount ($\mu=3\times10^{-5}$) to $\mu=1$ (where the total mass in the subordinate oscillators equals the mass of the primary oscillator).

A. SDOF regime ($\mu \ll \Delta/2eQ$ or $\mu \ll 1/Q^2$)

For very low mass ratios (when $\mu \ll \Delta/2eQ$), the subordinate substructure has negligible effect on the response of the primary, and the envelope of the decaying sinusoidal response matches that of the unadorned system [see Fig. 6(a)].

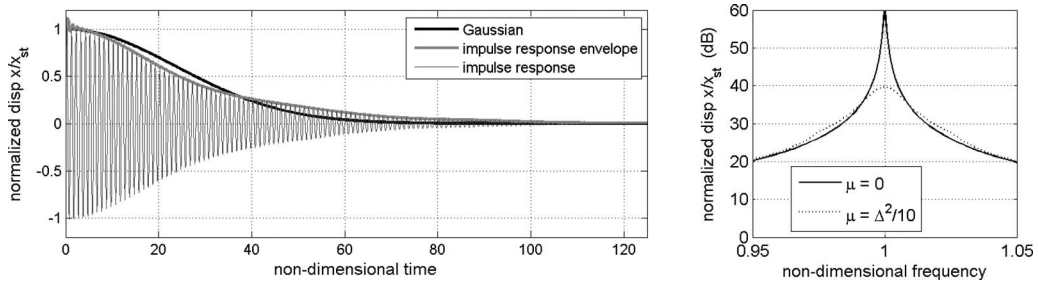


FIG. 7. The impulse response is shown in the time and frequency domains. The heavy gray line in the time plot (left) is the ringdown envelope (magnitude of the Hilbert transform) and is contrasted with a Gaussian function (the heavy black line). The FRF (right) of the adorned system is compared with a FRF of an unadorned oscillator. The system parameters for the adorned system here are mass ratio $\mu=\Delta^2/10$, $p=0.8$, $Q_{\text{primary}}=10^3$, $N=100$, and the fractional bandwidth $\Delta=0.05$.

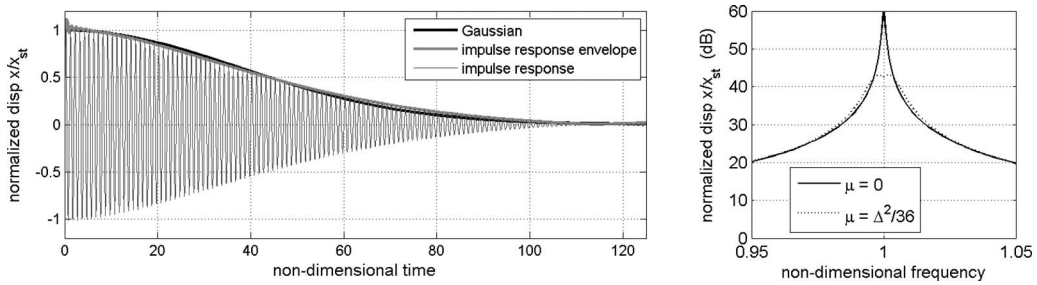


FIG. 8. The difference between the impulse response ringdown envelope and a Gaussian can be minimized by adjusting the frequency distribution within the SOA. In the case shown here the mass ratio is $\mu=\Delta^2/36$, $p=0.3$, $Q_{\text{primary}}=10^3$, $N=100$, and the fractional bandwidth $\Delta=0.05$. The lower value of the frequency distribution parameter p reduces the discontinuities that occur at the edges of the band of the SOA.

B. Apparent damping regime ($\Delta/2eQ<\mu<\Delta^2/e$ and $N\gg1$)

As μ increases, so-called apparent or pseudo-damping is observed. The envelope of the decaying sinusoidal response remains exponential but the time constant decreases as μ is increased. In this range, the magnitude and phase response are nominally the same as that of a SDOF oscillator with increased damping. An empirical expression for the apparent quality factor in this regime is given below, which is independent of any physical damping

$$Q_{\text{apparent}}=2\pi\frac{\text{energy stored}}{\text{energy transferred into SOA}}=\frac{2\Delta}{\pi\mu}. \quad (16)$$

The total quality factor, Q_{total}, for the primary oscillator, when $\mu<\Delta^2/e$, results from the combination of energy transfer and classical dissipation, and is expressed as

$$Q_{\text{total}}=\left(\frac{1}{Q_{\text{primary}}}+\frac{1}{Q_{\text{apparent}}}\right)^{-1}=Q_{\text{primary}}\frac{\pi\mu}{\pi\mu+2\Delta Q_{\text{primary}}}. \quad (17)$$

Examples of the behavior in the apparent damping regime can be seen in Figs. 6(b)–6(d). When the mass ratio satisfies $\mu<\Delta^2/e$, energy drawn out of the primary oscillator does not return if $Q_{\text{primary}}<Q_{\text{apparent}}$.

C. Rapid energy sink regime ($\mu\approx\Delta^2/e$ and $N\gg1$)

The case where the mechanical energy is most rapidly drawn out of the primary oscillator is often the regime of greatest interest. In this regime, the frequency spectra can take on a variety of shapes, including spectra corresponding to a Gaussian, a linear phase filter, a flat bandpass filter, a sinc function, or a constant phase filter. This regime is characterized by a rapid transfer of energy into the subordinate system that takes place within a period of time approximately equal to the reciprocal of the bandwidth of the subordinate set.

1. Gaussian envelope behavior ($\mu\approx\Delta^2/10$ and $N\gg1$)

As the mass ratio approaches $\mu\approx\Delta^2/10$, the time domain envelope (which was exponential for lower mass ratios) transitions into an approximate Gaussian envelope. The Gaussian has the unique property of being its own Fourier transform ($F\{e^{-(t/r)^2}\}=\tau\sqrt{\pi}e^{-(\omega r/2)^2}$). The corresponding FRF is also a nearly Gaussian function centered at the resonant frequency of the primary oscillator. Figures 6(e) and 7 show that the time domain ringdown inflects and is no longer exponential. A more nearly Gaussian ringdown envelope can be created by altering the frequency distribution within the SOA (see Fig. 8).

2. Linear phase behavior ($\mu\approx\Delta^2/5$ and $N\gg1$)

A dynamic response with a nearly linear phase can be created when the mass ratio satisfies $\mu\approx\Delta^2/5$ and $N\gg1$. Comparing Fig. 6(f) to the plots above and below it shows that, in this particular case, energy is drawn out of the primary oscillator as rapidly as possible and does not subsequently return. Figure 9 illustrates this behavior for six different SOA bandwidths.

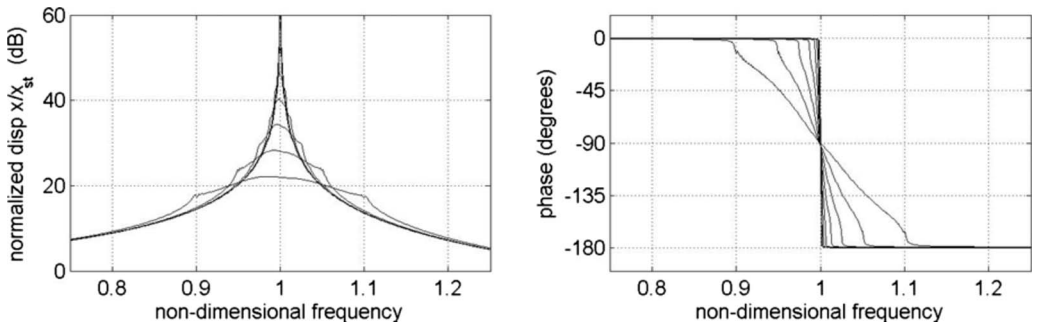


FIG. 9. A sequence of spectra with nearly linear phase response is shown for six SOA bandwidths ($\Delta=0.625\%, 1.25\%, 2.5\%, 5\%, 10\%, 20\%$). For this sequence the mass ratio $\mu=\Delta^2/5$, $p=0.8$, $N=100$, $Q_{\text{primary}}=N/2\Delta$.

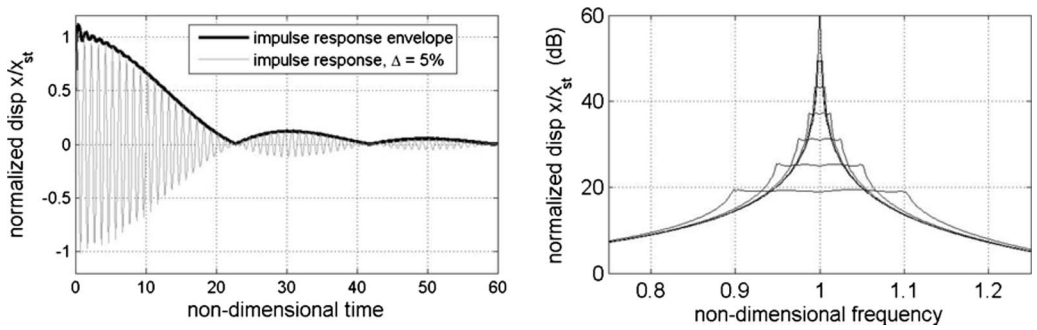


FIG. 10. A sequence of spectra with nearly flat in-band response. For this sequence $\Delta=0.625\%, 1.25\%, 2.5\%, 5\%, 10\%, 20\%$, and the mass ratio $\mu=\Delta^2/3.5$, $p=0.8$, $N=100$, it should be noted that the quality factor used for this sequence is lowered to $Q_{\text{primary}}=N/3\Delta$ (this smoothes the in-band response). The widest band response has a peak-to-peak flatness <0.5 dB and the in-band flatness of the narrower band systems is proportional.

3. Flat bandpass filter behavior ($\mu\approx\Delta^2/3.5$ and $N\gg1$)

Systems with spectra featuring flat response over some frequency band have applications in a number of areas such as transduction, filtering, and vibration isolation. Such a response is seen in Figs. 6(g) and 10. As the time histories in these plots show, energy is initially drawn away from the primary oscillator even faster than for lower mass ratios, but it subsequently returns from the SOA to the primary oscillator, and continues to pass back and forth prior to being completely dissipated. The non-dimensional time domain envelope equals zero at approximately $T=t\omega_n=1/\Delta$ and integer multiples. Note that this time period is not related to the energy return time discussed by Carcaterra,[1] which is addressed in Sec. VI.

4. Sinc-function envelope behavior ($\mu\approx\Delta^2/2.85$ and $N\gg1$)

A function commonly encountered in Fourier analysis and digital signal processing is the sinc function,[21] ($\text{sinc}(x)=\sin x/x$). The time domain envelope of the response of the primary oscillator approaches a rectified sinc function as the mass ratio is increased to $\mu=\Delta^2/2.85$. As the starting and ending frequencies of the SOA distribution begin to dominate the spectral response, the envelopes of the time histories can be seen as a beating between these two frequencies. This beating will persist for all further increases in the mass ratio (Fig. 11).

5. Constant phase behavior ($\mu\approx\Delta^2/1.7$ and $N\gg1$)

As mass ratio is increased still further, nearly constant phase response appears next in the sequence [see Fig. 6(h)]. The spectra shown in Fig. 12 further illustrate this for four different bandwidths of the SOA.

D. Transition regime ($\mu\approx\Delta^2$ and $N\gg1$)

Increasing the mass ratio further results in the system beginning to behave somewhat like the two-DOF DVA [see Figs. 6(i) and 13]. The time envelope exhibits classical beating of two frequencies; however, the two frequencies are no longer the end frequencies of the SOA, and are instead determined only by the mass ratio. In this transition regime, the amplitude of the second beat is some certain fraction of the amplitude of the first beat dictated by the mass ratio, while the decay of the amplitude of subsequent beats is dictated solely by the physical damping in the system. The ratio of the amplitudes for the first and second beats rises with rising mass ratio, transitioning from sinc-function-like behavior at lower mass ratios to exponential-type decay at higher mass ratios.

E. DVA regime ($\mu\gg\Delta^2$)

The dynamics of an SOA adorned primary oscillator mimic that of a DVA when the mass ratio is substantially greater than the square of the fractional bandwidth [see Figs. 6(j)–6(l) and 14]. In this regime, the response appears as classical beating and the two frequencies are determined by the mass ratio. When the mass ratio $\mu=1$, the frequencies are

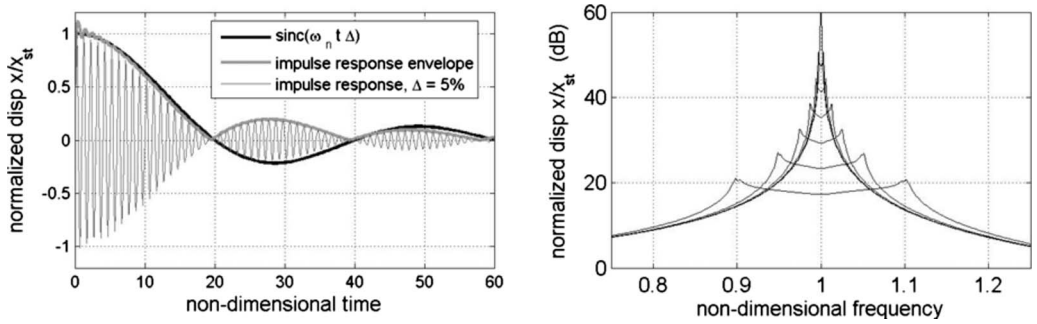


FIG. 11. The envelope of the time histories for systems $\mu=\Delta^2/2.85$, $p=0.8$, and $N=100$ have a rectified sinc-function form. For the sequence of spectra (right) $\Delta=0.625\%, 1.25\%, 2.5\%, 5\%, 10\%, 20\%$, and $Q_{\text{primary}}=N/2\Delta$.

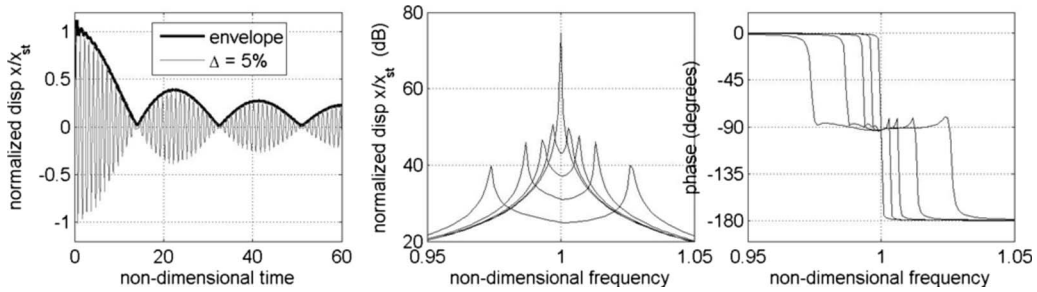


FIG. 12. The phase responses for systems $\mu=\Delta^2/1.7$, $p=0.8$, and $N=100$ have nearly a constant value over the band of the SOA. For this sequence of spectra $\Delta=0.625\%, 1.25\%, 2.5\%, 5\%$, and $Q_{\text{primary}}=N/2\Delta$ (the time history shown on the left is for the case $\Delta=5\%$).

defined by the golden ratio $(\pm 1+\sqrt{5})/2 \approx (0.618, 1.618)$ (the outer tick marks in the frequency spectra in Fig. 6 are placed at these frequencies). As in the previous regime, the response exhibits beating of two frequencies, and the decay of the amplitudes is governed by the amount of physical damping in the system. The attenuation of motion at the center frequency is substantially less than that of a damped DVA but attenuation is seen over a wider frequency band.

VI. REVERSIBILITY AND ITS RELATIONSHIP TO THE DENSITY OF STATES

Given an impulse excitation, and a small mass ratio ($\mu < \Delta^2/10$), energy moves solely from the primary oscillator into the SOA. This can be seen in the examples discussed in Secs. V A and V B and the first part of Sec. V C that includes the Gaussian regimes. These results correspond to the downward sloping portion of the curves in Fig. 5. In this regime, the rate of energy flow from the primary oscillator to the SOA is greater than the rate of dissipation. This can be seen by recognizing that $2\pi/Q_{\text{primary}}$ is the amount of energy removed by classical dissipation per cycle divided by the total energy in the primary oscillator and that $2\pi/Q_{\text{apparent}}$ is the amount of energy transferred to the SOA per cycle normalized by the same total energy. Carcaterra *et al.*[17] referred to this irreversibility as "energy trapping." As the mass ratio increases beyond this point, the mechanical energy moves back and forth between the between the primary oscillator and the SOA. Another way to consider this is to recognize that after an impulse excitation the SOA elements initially move coherently. This motion lags the motion of the primary oscillator which ensures that the energy flow is in the direction of the SOA. Because each vibrates at a different frequency, the individual elements of the SOA lose coherence as time passes and the net force they apply on the primary becomes negligible, implying that energy is temporally trapped in the SOA. At some time later, the motion of the elements of the SOA once again becomes coherent but now leads the motion of the primary oscillator, returning energy to the primary oscillator. This is only seen when $Q_{\text{primary}} > Q_{\text{apparent}}$. The time required for the SOA elements to go from coherent to incoherent and back to coherent is the coherence time, $\tau_{\text{coherence}}$ (also called the Heisenberg time or return time). For a linear distribution ($p=1$), the coherence time is the reciprocal of the difference between the frequencies of any two adjacent elements in the SOA. This can also be expressed in terms of the number of oscillators and the total bandwidth of the SOA as $\tau_{\text{coherence}} f_{\text{primary}} = (N-1)/\Delta$. This expression can also be seen as the density of states within the SOA. When the distribution of frequencies is not uniform ($p \neq 1$), the oscillators do not return to complete coherence, yet the time of greatest coherence can be expressed as $\tau_{\text{coherence}} f_{\text{primary}} = (N-1)/p\Delta$. Figure 15 shows that the return of energy from the SOA is only visible as a discrete event when two conditions are met: (1) the coherence time *longer* than the time it takes for apparent damping to transfer most of the system's energy to the SOA ($2\Delta/(\pi^2\mu) > \tau_{\text{coherence}} f_{\text{primary}}$), and (2) the coherence time is *shorter* than the time required for physical damping to dissipate most of the energy in the system ($\tau_{\text{coherence}} f_{\text{primary}} > 4Q_{\text{primary}}/\pi$). If condition (1) is not met (the coherence time is too short), energy return *from* the SOA occurs while energy is still being transferred *to* the SOA. If condition (2) is not met (the coherence time is sufficiently long), there is no significant energy remaining in the system to be returned when the subordinate elements return to coherent motion.

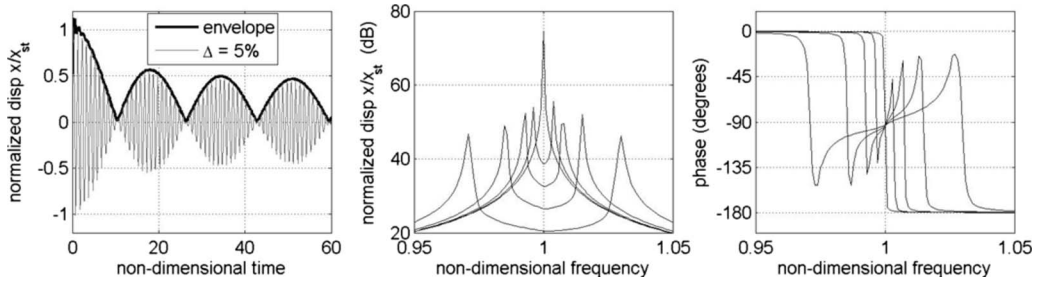


FIG. 13. The envelope of the time histories exhibiting beating at a frequency difference that exceeds the bandwidth of the SAO. A relatively larger fraction of mechanical energy is drawn out of the primary oscillator by the SOA between the first and second beats. The subsequent beat amplitude reductions are explained by the quality factors of the system. For this sequence the mass ratio $\mu=\Delta^2/1.0$, $p=0.8$, and $N=100$. For this sequence of spectra $\Delta=0.625\%, 1.25\%, 2.5\%, 5\%$, and $Q_{\text{primary}}=N/2\Delta$.

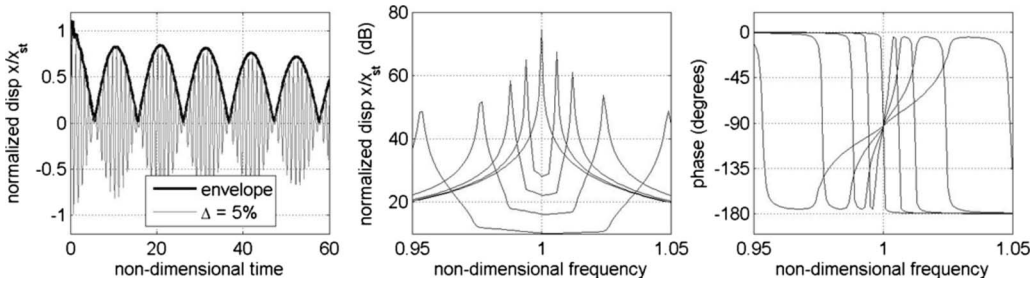


FIG. 14. The envelope of the time histories beat as a two-DOF system. For this sequence of spectra the mass ratio $\mu=\Delta^2/0.3$, $p=0.8$, and $N=100$. $\Delta=0.625\%, 1.25\%, 2.5\%, 5\%$, and $Q_{\text{primary}}=N/2\Delta$.

This example also illustrates how the modal overlap parameter η, discussed in Sec. III, affects the smoothness of the spectra. The return energy is significant when the mass ratio is in apparent damping regime $\Delta/2Qe<\mu<\Delta^2/10$ and the modal overlap parameter $\eta<2$. This condition also corresponds to the condition when there are not a sufficient number of DOFs to fill the band of the SOA and thereby smooth the spectral response of the primary oscillator. When the mass ratio of the SOA is large enough to affect the response and $\eta>2$, the spectra are smooth and can be engineered to achieve desired responses.

VII. CONCLUSIONS

The SOA can be seen not only as a generalization of a classical DVA which can provide very effective vibration isolation for a single frequency but also as a means to shape the frequency response of a resonant system. The tailored SOA can be used for vibration isolation and overcomes the inherent bandwidth limitations of a DVA by spreading resonant elements over a prescribed frequency band. However, as shown in this paper, the amplitude and phase spectra of the system can be tailored in a wide variety of ways, creating oscillatory systems that are useful for a variety of other applications. For instance, the approach can be used to design a linear attenuator over a defined band or to create transducers that are resonant over a wide frequency band without adding excessive damping.

This work has been pursued in the context of shaping the vibration response spectra of mechanical systems. However, one should recognize that the governing equation for an oscillating mechanical system is the same equation that governs other types of oscillator systems such as simple electrical circuits and acoustic systems. This implies that the responses of broad classes of systems can be manipulated using analogous subsystems. Additionally, the concept of the SOA can be implemented on the micro- or nano-scale to shape the spectra of high frequency micro- and nanomechanical resonators, which have a variety of potential applications in mass sensing, rf communications, and studies of fundamental physics. Such an implementation could be realized, in principle, in complementary metal oxide semiconductor or using other fabrication processes directly on chip.

The results shown in this paper were achieved using very simple property distributions in the subordinate array,

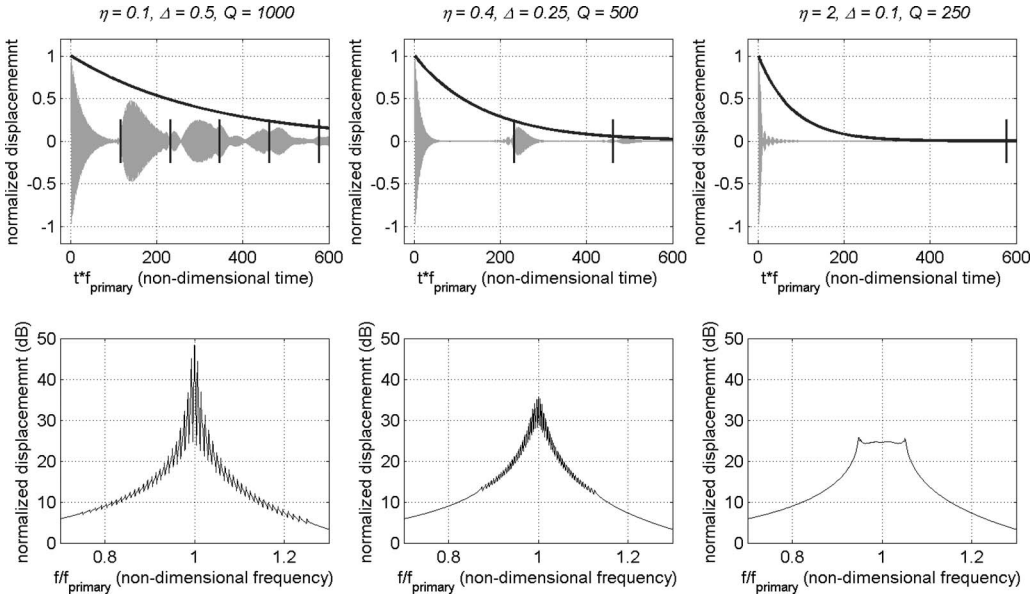


FIG. 15. Three examples for the response of primary oscillators with attached SOAs are shown in both time and frequency. The solid black line in the time histories is the ringdown envelope for the physical damping. In each case the mass ratio $\mu=3.16\times10^{-3}$, $N=50$, $p=0.85$, $q=-0.25$, The non-dimensional bandwidth for the three cases are $\Delta=0.5, 0.25, 0.1$ and the quality factors for all oscillators (primary and subordinate) are $Q=1000$, 500, and 250 for the three cases, respectively. The modal overlap η changes inversely as the bandwidth increases, and in these cases $\eta=0.1, 0.4, 2$. Below each time history, the corresponding spectrum illustrates how the rate of the energy return corresponds to smoothness in the frequency domain.

yet demonstrate a significant capability to shape the spectral response of an oscillatory system using only a small amount of added mass. For instance, a distribution with a total added mass ratio of ~0.3% produced flat frequency responses over the band of the SOA to within 0.5 dB. More general distributions of parameters allow even greater control over the shape of the frequency spectrum. Numerical optimization of a parameter distribution not tied to a specific functional form will allow creation of systems with highly tailored spectral response.

ACKNOWLEDGMENTS

The authors would like to thank John McCoy for introducing them to the class of systems discussed in this paper and for a number of helpful and enlightening conversations.

[1]A. Carcaterra and A. Akay, "Transient energy exchange between a primary structure and a set of oscillators: Return time and apparent damping," J. Acoust. Soc. Am. **115**, 683–696 (2004).
[2]M. Strasberg, "When is a 'fuzzy' structure not a fuzzy structure," J. Acoust. Soc. Am. **107**, 2885 (2000).
[3]A. D. Pierce, V. W. Sparrow, and D. A. Russell, "Fundamental structural-acoustic idealizations for structures with fuzzy internals," J. Vibr. Acoust. **117**, 339–348 (1995).
[4]R. Ohayon and C. Soize, *Structural Acoustics and Vibration* (Academic Press, San Diego, 1998).
[5]A. D. Pierce, "Fuzzy-structure acoustics," McGraw-Hill's Access Science Encyclopedia of Science & Technology Online, http://www.accessscience.com/content.aspx?id=757466 (Last viewed December 12, 2008).
[6]G. Maidanik, K. J. Becker, and L. J. Maga, "Replacement of a summation by an integration in structural acoustics," J. Sound Vib. **291**, 323–348 (2006).
[7]G. Maidanik, "Induced damping by a nearly continuous distribution of nearly undamped oscillators: Linear analysis," J. Sound Vib. **240**, 717–731 (2001).
[8]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).
[9]M. Strasberg, "Continuous structures as 'fuzzy' substructures," J. Acoust. Soc. Am. **100**, 3456–3459 (1996).
[10]R. J. Nagem, I. Veljkovic, and G. Sandri, "Vibration damping by a continuous distribution of undamped oscillators," J. Sound Vib. **207**, 429–434 (1997).
[11]S. P. Timoshenko, *Vibration Problems in Engineering* (D. Van Nostrand Company, New York, 1928).
[12]J. P. Den Hartog, *Mechanical Vibrations* (McGraw-Hill, New York, 1947).
[13]D. Inman, *Engineering Vibrations*, 3rd ed. (Prentice-Hall, Upper Saddle River, NJ, 2008).
[14]L. Zuo and S. A. Nayfeh, "Optimization of the individual stiffness and damping parameters in multiple-tuned-mass-damper systems," ASME J. Vibr. Acoust. **127**, 77–83 (2005).
[15]L. Zuo and S. A. Nayfeh, "Minimax optimization of multi-degree-of-freedom tuned-mass dampers," J. Sound Vib. **272**, 893–908 (2004).
[16]A. Carcaterra and A. Akay, "Theoretical foundations of apparent-damping phenomena and nearly irreversible energy exchange in linear conservative systems," J. Acoust. Soc. Am. **121**, 1971–1982 (2007).
[17]A. Carcaterra, A. Akay, and I. M. Koç, "Near-irreversibility in a conservative linear structure with singularity points in its modal density," J. Acoust. Soc. Am. **119**, 2141–2149 (2006).
[18]A. Carcaterra, A. Akay, and F. Lenti, "Pseudo-damping in undamped plates and shells," J. Acoust. Soc. Am. **122**, 804–813 (2007).
[19]A. Akay, Z. Xu, A. Carcaterra, and I. M. Koç, "Experiments on vibration absorption using energy sinks," J. Acoust. Soc. Am. **118**, 3043–3049 (2005).
[20]R. H. Lyon, *Statistical Energy Analysis of Dynamical Systems* (MIT, Cambridge, MA, 1975).
[21]R. Bracewell, *The Fourier Transform and Its Applications* (MGraw-Hill, New York, 1986).

A numerical study of defect detection in a plaster dome ceiling using structural acoustics

J. A. Bucaro[a),b)]
SET, Inc. @ the Naval Research Laboratory, 4555 Overlook Avenue, Washington, DC 20375

A. J. Romano, N. Valdivia, and B. H. Houston
Naval Research Laboratory, 4555 Overlook Avenue, Washington, DC 20375-5320

S. Dey[b)]
Global Strategies Group (North America), 2200 Defense Boulevard, Suite 405, Crofton, Maryland 21114

(Received 3 December 2008; revised 22 April 2009; accepted 23 April 2009)

A numerical study is carried out to evaluate the effectiveness of using measured surface displacements resulting from acoustic speaker excitation to detect and localize flaws in a domed, plaster ceiling. The response of the structure to an incident acoustic pressure is obtained at four frequencies between 100 and 400 Hz using a parallel *h-p* structural acoustic finite element-based code. Three ceiling conditions are modeled: the pristine ceiling considered rigidly attached to the domed-shape support, partial detachment of a segment of the plaster layer from the support, and an interior pocket of plaster deconsolidation modeled as a heavy fluid. Spatial maps of the normal displacement resulting from speaker excitation are interpreted with the help of predictions based on static analysis. It is found that acoustic speaker excitation can provide displacement levels readily detected by commercially available laser Doppler vibrometer systems. Further, it is concluded that for 1 in. thick plaster layers, detachment sizes as small as 4 cm are detectable by direct observation of the measured displacement maps. Finally, spatial structure differences are observed in the displacement maps beneath the two defect types, which may provide a wavenumber-based feature useful for distinguishing plaster detachment from other defects such as deconsolidation.
 [DOI: 10.1121/1.3133922]

PACS number(s): 43.40.Le, 43.40.Dx, 43.60.Rw [TDM] Pages: 140–148

I. INTRODUCTION

Interest in development of structural acoustic monitoring techniques to support efforts related to historic building assessment and restoration has been growing. This growth is in part associated with increasing regulations requiring Federal agencies to make every effort possible to rehabilitate historically significant buildings and structures. The scope of the interest can be appreciated, for example, by noting that there are over 1×10^6 commercial and institutional buildings erected over 40 years ago still in use in the United States, an age to which many local jurisdictions have lowered the threshold for applying the "historic" label. Further, the Department of Defense alone controls an estimated 90 000 structures built over 60 years ago, not a small number of which might be considered historic.

In a number of the truly historic buildings whose finish layers often consist of plaster, walls and/or ceilings often bear precious artwork such as mosaic or frescoed images. In the case of the latter, where the paintings were created on wet plaster, the current physical and mechanical conditions of the plaster layer(s) determine for the most part the near to midterm viability of the artwork. Particularly for the case of a ceiling, the development of defects such as detachment of the plaster layer from its supporting brick or mortar structure or pockets of plaster deconsolidation can eventually lead to catastrophic failure of the plaster ceiling and the unrecoverable loss of priceless artwork.

Several years ago, Vignola *et al.* (2005) reported the successful use of externally produced vibration to evaluate the state-of-health (SOH) of highly frescoed plaster walls in the U.S. Capitol building. In this technique, a low level broadband force was applied locally to the wall using an electro-dynamic shaker, and the resulting spatially dependent surface vibrations were mapped over the frescoed wall using a scanning laser Doppler vibrometer (SLDV). Direct observation of the scanned displacement maps allowed those authors to detect and locate plaster defects whose presence was subsequently confirmed by conventional tapping tests, a time consuming manual technique in which the restorer gently hits the frescoed plaster surface with the knuckles or a small hammer while sensing the resulting vibrations and sounds with the fingertips and ears. Several years earlier, Castellini *et al.* (2000) carried out similar measurements on artificially aged walls and icons in France and on actual historical frescoed walls in Italy. Although the shaker/SLDV-based plaster SOH assessments of Vignola *et al.* (2005) and Castellini *et al.* (2000) were quite successful, they were empirical in nature, and little supporting vibration analysis was reported, which would allow one to generalize these results to other situations.

[a)]Author to whom correspondence should be addressed. Electronic mail: bucaro@pa.nrl.navy.mil
[b)]On-site contractors at the Naval Research Laboratory, Washington, DC.

Here we consider a related approach in which a broadband acoustic speaker rather than a shaker is used to excite the structure. As in the shaker excitation case, the resulting vibrational response of the structure would be mapped over its surface using a SLDV. Use of a speaker is an especially attractive approach for ceilings (our particular interest here) in that both the excitation and the scanning could be carried out remotely eliminating the requirement for scaffolding. Such an acoustic speaker approach was first introduced and applied to frescoes by Castellini *et al.* (1994, 1996) and later applied by Tornari *et al.* (2001) to mosaics, ceramics, inlaid wood, and easel paintings. As in the case using shaker excitation, both these and subsequent studies using speaker excitation did not provide sufficient analytical results, which would allow one to predict the sensitivity of the technique to defect detail such as type and size and how well the approach might perform more generally.

In the present study, we use a simulated database of surface vibration generated using an advanced structural acoustic finite element-based code. In particular, we explore application of the speaker-based technique to what we consider to be a generic domed ceiling—a common structure found in historic buildings and residences—in which a plaster layer is attached to, and takes the shape of, a backing structure considered to be relatively rigid such as brick or masonry. Aided by static analysis, we address the effectiveness of using the "measured" surface displacements resulting from acoustic speaker excitation to detect and localize two different defect types: (A) detachment of the plaster layer from the supporting brick or masonry, and (B) an internal region in the plaster, which has become deconsolidated. Specific questions we wish to resolve are as follows: (1) Is acoustic excitation effective at producing readily measured surface displacements, which could be used in detecting these flaws? (2) In general, what issues and/or benefits are introduced by using acoustic speaker versus force actuator excitation? (3) Are typical defects of the type mentioned above detectable by straightforward observation of the acoustically excited spatial displacement maps and how does this depend on flaw size and ceiling thickness? (4) What is the minimum detectable defect size? (5) If a defect is detected, might one be able to differentiate between detachment and deconsolidation?

In Sec. II, we describe the ceiling geometry and the finite element-based structural acoustic code used to produce the ceiling response data, which forms the core of this study. In Sec. III, after establishing some baseline predictions using simple static models, we present the results of the dynamic simulations. Finally, in Sec. IV, we provide answers to the questions posed above.

II. NUMERICAL DATABASE

A. Ceiling structure

The ceiling structure [Fig. 1(a)] is modeled as a three-dimensional (3D) thin shell enclosing a volume of air. The geometry of the shell is that of half an ellipsoid with radii along x, y, and z directions given by 4.2164, 4.216, and 1.9304 m, respectively, with a uniform thickness t of 0.0254 m. The somewhat shallow 3D elliptical shape is a common dome geometry since a 13%–20% rise-to-diameter ratio guarantees virtually no bending moments. The shell is considered made of plaster, which is modeled as a linear viscoelastic material having density ρ of 1444 kg/m^3, Young's modulus E of 7 GPa, Poisson's ratio σ of 0.2, and uniform damping of 2%. The air inside has density 1.25 kg/m^3 and speed of sound 320 m/s. The ceiling base (where a real ceiling would meet a supporting wall) and the outer surface (where it would meet and attach to the masonry or brick foundation) are considered rigidly fixed, and the system is subjected to acoustic excitation from a point source located 0.5 m below the center of the base of the ellipsoidal volume [Fig. 1(b)].

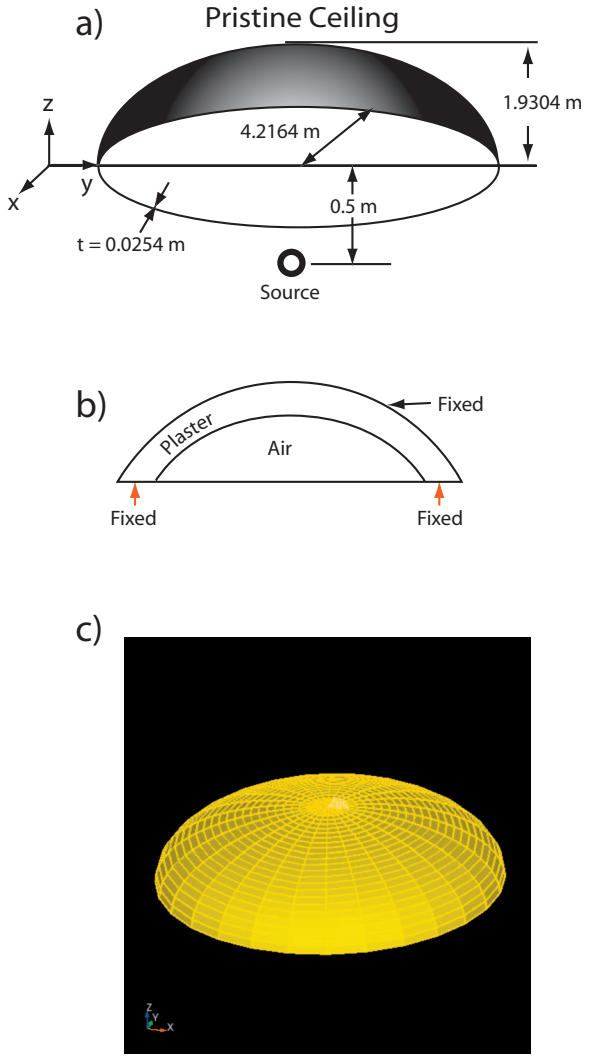


FIG. 1. (a) The geometry of the elliptical ceiling and location of the acoustic source (viewed from beneath the ceiling), (b) the boundary conditions assumed in the finite element calculation, and (c) the mesh used in conjunction with the *hp*-finite element code (viewed from above the ceiling).

B. Numerical code

The coupled structural acoustic response is computed using the parallel *hp*-version finite element code STARS3D (Dey *et al.*, 2001; Dey and Datta, 2006), which utilizes high-order basis functions to achieve accurate results. This is nec-

essary since in the frequency regime studied here (100–400 Hz), structural wavenumber-thickness products reach values as high as ~90. For models with fixed low-order *h*-approximations, one must use a spatial mesh with enough refinement to satisfy the requirement of a certain number of elements per wavelength. We use a *p*-version approach where the dispersion error can be controlled by increasing the polynomial degree of approximation (p) for a fixed spatial mesh refinement (h). Accordingly, we use a spatial mesh [Fig. 1(c)] that approximates the curved shell-geometry well and then execute a *p*-convergence study to determine the proper polynomial degree of approximation (p) to use with the mesh at hand. For the present model, a *p*-convergence analysis indicated numerical converged solutions at cubic ($p=3$) approximations. The numerical model is solved at 100, 200, 300, and 400 Hz to determine the complex-valued 3D displacement field on the surface of the ceiling in contact with the air inside.

C. Ceiling conditions

Three ceiling conditions are modeled. (1) Pristine ceiling: In this model, depicted in Fig. 1(b), the homogeneous ceiling is considered rigidly attached to the backing throughout the outer boundary. This implies no motion along the portions marked "fixed." (2) Partial detachment: In this model, a portion of the outer boundary of the ceiling is fully detached from the rigid backing and the boundary condition is considered to be free. The shape and location of this free patch, shown as the shaded area in Fig. 2(a), resulted from the desire to have both a non-axisymmetric defect shape and the ability to derive it based on the geometric model and mesh already used for the homogeneous ceiling finite element model. To achieve this in a straightforward manner we isolated a circular portion of the homogeneous mesh and selected one-quarter of it to be the defect patch region. Figure 2(b) illustrates the boundary condition for this case, which implies free motion of that portion of the boundary marked "free." (3) Embedded deconsolidation: In this model, depicted in Fig. 2(c), we consider a small pocket of heavy fluid with the same shape and location as that of the detached segment. The thickness of the pocket, whose center is located midway through the plaster layer, is one-third of the total ceiling thickness. The fluid defect, with density 1444 kg/m^3 and wave-speed 1175 m/s (compressibility =2 GPa), is taken to represent plaster deconsolidation at least in a simple sense.

III. ANALYSIS AND RESULTS

In order to establish a response baseline, we calculate the surface displacement level for the case of an infinite flat plaster plate of thickness t under the action of an applied static pressure. For pressure P the displacement, W, would be given by

$$W \approx -\frac{Pt(1-\sigma-\sigma^2)}{E}. \quad (1)$$

The finite element-based numerical simulation has modeled a pressure of 1 Pa at the source point (i.e., at a position 0.5 m

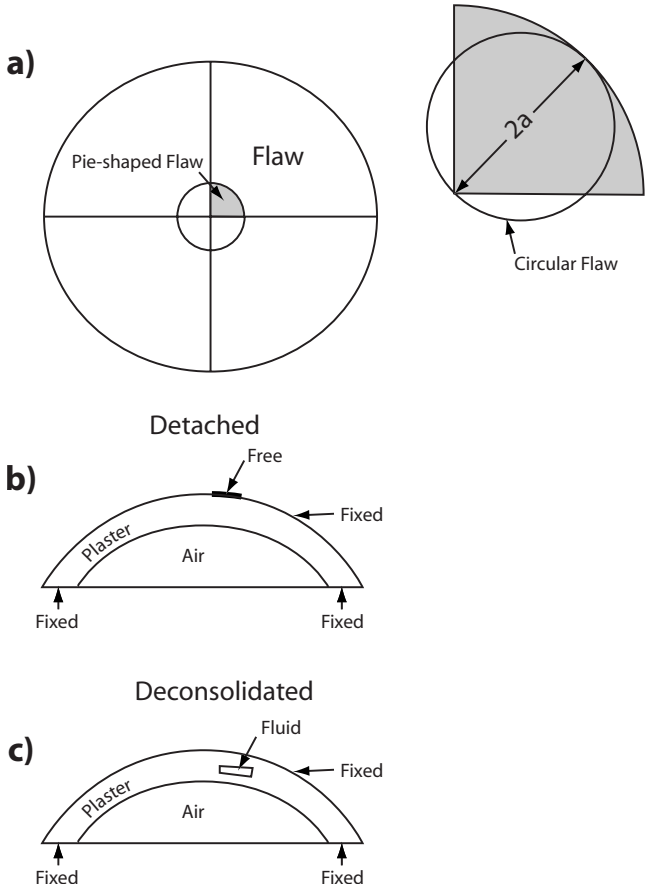


FIG. 2. (a) The geometry of the defect (plan view from above the ceiling). The insert shows a circular flaw having the same area as the pie-shaped flaw. (b) The boundary conditions for the ceiling with detached defect. (c) The boundary conditions for the ceiling with deconsolidated defect.

below the plane at which the ceiling begins), which with spherical spreading would become $(2.43)^{-1}$ Pa at a point at the plaster surface directly above the source. For this pressure, $E=7$ GPa and $t=2.54$ cm, using Eq. (1) we obtain $W \sim 0.8 \times 10^{-12}$ m.

With this base displacement level established, we now examine the normal displacement maps produced from the dynamic numerical simulation for the unflawed elliptical ceiling (Fig. 3). We consider only the normal displacements since in an actual experimental study the typical laser Doppler vibrometer would measure only this normal vibration component. As can be seen, at all four frequencies studied, the numerical computations predict somewhat complicated, nearly circularly symmetric response patterns.

We find that these circularly symmetric frequency dependent response patterns (Fig. 3) are, in fact, what one would expect for the pressure interference patterns produced and determined by reflections from the elliptical ceiling surface and the incident spherical wave. Further, the azimuthally dependent superposed weaker patterns are related to elastic wave effects in the plaster layer. To test this reasoning, we also used the finite element structural acoustic code to compute the pressure on the same elliptical surface for the case in which the plaster itself is rigid, and these are shown in Fig. 4. The computed pressure patterns are almost identical to those shown in Fig. 3 for the unflawed ceiling dis-

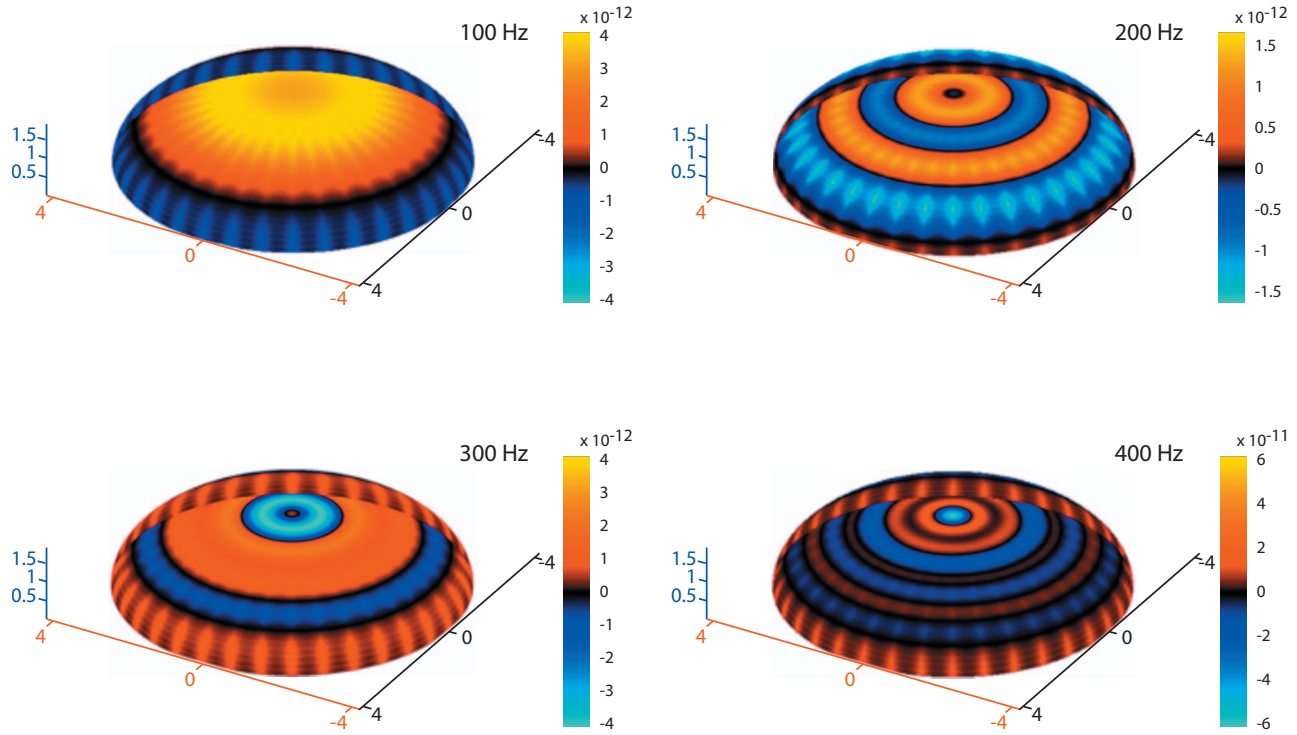


FIG. 3. Normal displacement levels for the unflawed ceiling calculated with the STARS3D code for four frequencies (displayed on the exposed surface of the ceiling).

placements; in addition, the weaker azimuthally dependent response is now missing. Further, we find that normalizing the displacement levels shown in Fig. 3 by the pressure levels in Fig. 4 produces average normalized responses over the ceiling very close to 3×10^{-12} m/Pa, which is the same value obtained for $W/P=t(1-\sigma-\sigma^2)/E$ from Eq. (1) using our specific ceiling parameters. These facts support our contention that the prominent ceiling response patterns are related to acoustic interference and focusing effects determined by the ceiling geometry while the much weaker structure is associated with elastic wave propagation and modal effects in the plaster layer.

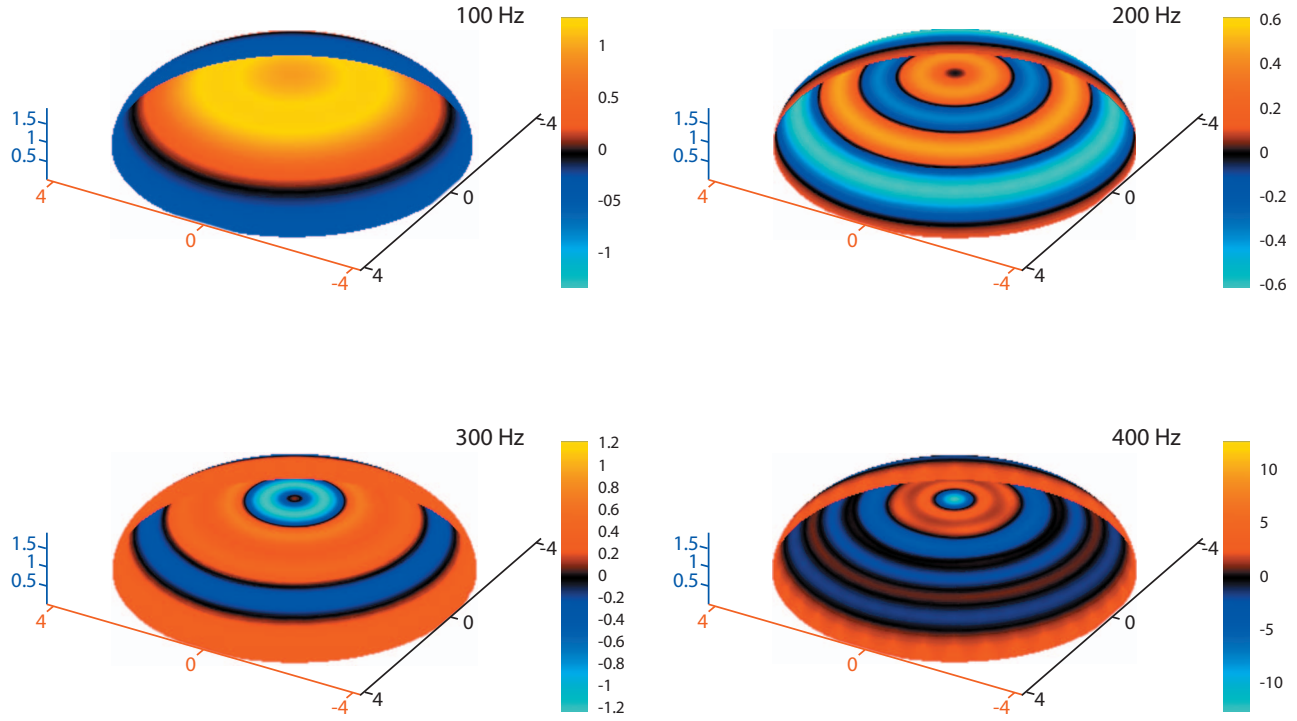


FIG. 4. Surface pressure levels for a perfectly rigid elliptical ceiling calculated with the STARS3D code for four frequencies (displayed on the exposed surface of the ceiling).

This is in contrast to the case in which shakers (Vignola *et al.*, 2005) are applied to the structure locally wherein elastic waves propagating in the plaster layer and modal responses are the cause of the observed frequency and spatially dependent displacement response functions. As is well illustrated here, in using an air-borne acoustic wave to excite the ceiling, not surprisingly the "architectural acoustics" will to a large extent determine the effective spatial distribution of the pressure applied to the ceiling, and one must take these effects into account when attempting to detect defects based on local variations in the acoustically-forced displacement. Of course, these patterns would be affected by both additional returns in an actual room enclosed by walls and by changes in the position of the source. These pressure interference effects could be problematic in that they can serve to confuse attempts to detect and localize a defect by direct observation of the abrupt increase in displacement level associated with a mobile flaw.

A. Detached defect

Before reviewing the numerically simulated dynamic response data associated with these flaws, we estimate the displacement response one would expect for acoustically-forced detached layers and how this might depend on defect size. This is of interest since defect displacement amplitudes sufficiently high compared to the levels produced on the nonflawed areas could be detected and localized simply by direct observation of the displacement maps themselves without further need of post-processing.

We are able to determine this in an approximate sense using available expressions in literature for the deflection under static loading of simple geometries obeying ideal, but we think relevant, boundary conditions. In particular, we consider a circular detached plate segment of radius a in an otherwise unflawed flat plaster plate of uniform thickness, t, under the action of a uniform static pressure, P, and for fixed boundary conditions over the circular edge of the defect. For this case, the center displacement, W_C, of the defect is given by Young (1989) as

$$W_C = -\frac{Pa^4}{64D}, \tag{2}$$

where D is the flexural rigidity given by

$$D = \frac{Et^3}{12(1-\sigma^2)}. \tag{3}$$

For the above case, the ratio of the defect center displacement to the plaster layer displacement away from the defect is then given from Eqs. (1)–(3) as

$$\frac{W_C}{W} \approx \frac{3}{16}(1+\sigma)\left(\frac{a}{t}\right)^4. \tag{4}$$

We plot this expression as a function of a for three different plaster thicknesses in Fig. 5. In order that the defect displacement be clearly pronounced above any spatial structure in the plaster layer displacement maps (such as seen in Fig. 2), we require the ratio W_C/W to be large, say, ≥ 4. Equation (4) in turn requires that

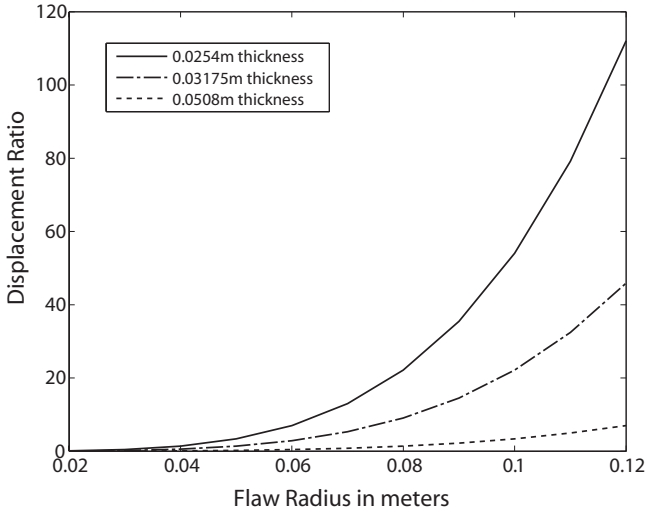


FIG. 5. The ratio of detached flaw displacement to that of the healthy ceiling versus the circular flaw radius computed from the static expression of Eq. (4) for three plaster thicknesses.

$$\frac{a}{t} \geq 2. \tag{5}$$

The minimum directly visible defect size defined by Eq. (5) is then seen to be directly proportional to the plaster thickness, t. For our 2.5 cm thick plaster ceiling, Eq. (5) implies that defect sizes (i.e., $2a$) of about 10 cm or greater should be directly visible in the scanned displacement data. The corollary is that much smaller defects ($\ll$10 cm) will not stand out against the background displacement, and post-processing methods such as flexural inversion (Bucaro *et al.*, 2004) will be required for their detection. The analysis so far pertains to defect size requirements necessary for sufficient contrast between the background ceiling displacements and those associated with the defect. We will discuss in Sec. IV minimum displacement levels required for detection by a typical SLDV system.

The above argument is based on a static response analysis expected to be sufficient at frequencies well below the first mechanical resonance of the defect. If, however, the dynamic acoustic excitation excites a defect resonance, levels higher than those predicted above could be experienced. One can estimate the fundamental resonance frequency, f_R, of the disk-shaped defect using a lumped-mass approximation wherein the spring constant, k, would be given as $P\pi a^2/W_C$ and the mass, m, as $\pi\rho t a^2$. Then with the help of Eq. (2) we have

$$\omega_R = 2\pi f_R = \sqrt{\frac{k}{m}} = 4ta^{-2}\sqrt{\frac{E}{3\rho(1-\sigma^2)}}. \tag{6}$$

We plot this expression as a function of a in Fig. 6 for two plaster layer thicknesses. As can be seen, for the plaster thickness used in the simulation (t=0.0254 m), defect sizes with diameters of about 0.5 m would resonate within our current band and thus have higher displacement levels than predicted statically. However, relying on the use of resonant excitation to lower the minimum observable defect size ($\sim$10 cm for t=0.0254 m) predicted by Eq. (5) would re-

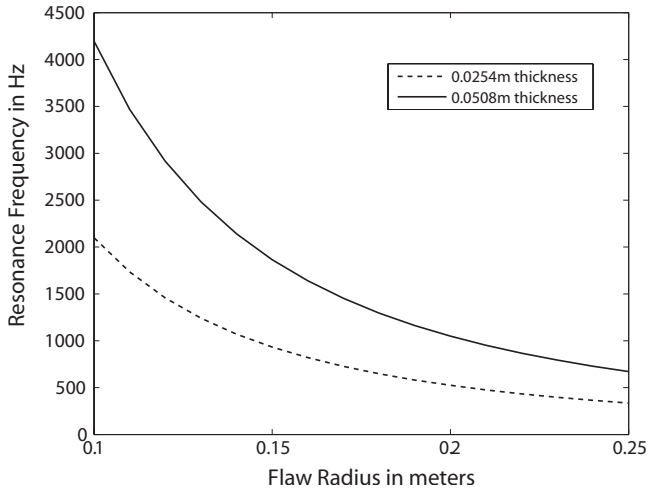


FIG. 6. The resonance frequency versus circular flaw radius computed using Eq. (6) based on a lumped parameter model for two plaster thicknesses.

quire very high frequencies as can be seen from Eq. (6), which gives $f_R > 8.4$ kHz for $2a < 10$ cm. We can obtain a rough check on the validity of this resonance analysis using the data reported in Vignola *et al.*, 2005 for the case of a shaker-excited 0.02 m thick plaster wall in the Brumidi Corridor of the U.S. Capitol building. In Fig. 4 of this reference, a 0.37 m diameter flaw is observed resonating at about 460 Hz, which is close to the 480 Hz predicted by Eq. (6) for this flaw size and wall thickness. We will discuss these resonance frequency estimates further in the discussion of the dynamic response.

For each of the four frequencies, we show in Fig. 7 the numerically computed dynamic displacement maps for the detached ceiling resulting from the point source excitation. We point out that displays using the same scale factor as in Fig. 3 (although not shown here) show almost identical interference patterns away from the defect region as expected. As can be seen in Fig. 7, the defect stands out clearly, i.e., displacements over the defect are much larger than those over the healthy plaster, and the normal plaster displacements appear as almost black on the linear scale over which these are displayed. Although the defect shape used in the finite element computations is not circular but quasi-pie-shaped (chosen for convenience to be consistent with the grid structure), we note that its area would be approximately that of a flat circular disk with radius 25 cm [see the insert in Fig. 2(a)]. For the latter, such a defect size has $a/t = 10$, which easily meets the requirement specified in Eq. (5) so that we should expect the displacement levels associated with the defect to dominate the spatial maps as indeed they do (Fig. 7). Using Fig. 3 to estimate the healthy plaster displacement near the flaw location, we find the contrast ratio to be between 1500 and 1800 over our frequency range in agreement with Eq. (4), which predicts a ratio of 2000. We also point out that for this circular disk, the displacement predicted by the static expression [Eq. (2)] is 6×10^{-9} m for a pressure of 1 Pa. This compares well with the dynamic displacement of 4.3×10^{-9} m at 100 Hz if we note from Fig. 4 that the dynamic pressure at the flaw is about 1 Pa.

In Table I we list the total normal force (real part) acting on the flaw as determined from integration of the pressures (shown in Fig. 4) over the area of the flaw. We then normalize the maximum detached flaw displacements [more clearly seen in the expanded display in Fig. 8(a)] by this force level and list these ratios in Table I. As can be seen in the table, for the three lower frequencies, these force-normalized displacement levels only increase weakly with frequency. However,

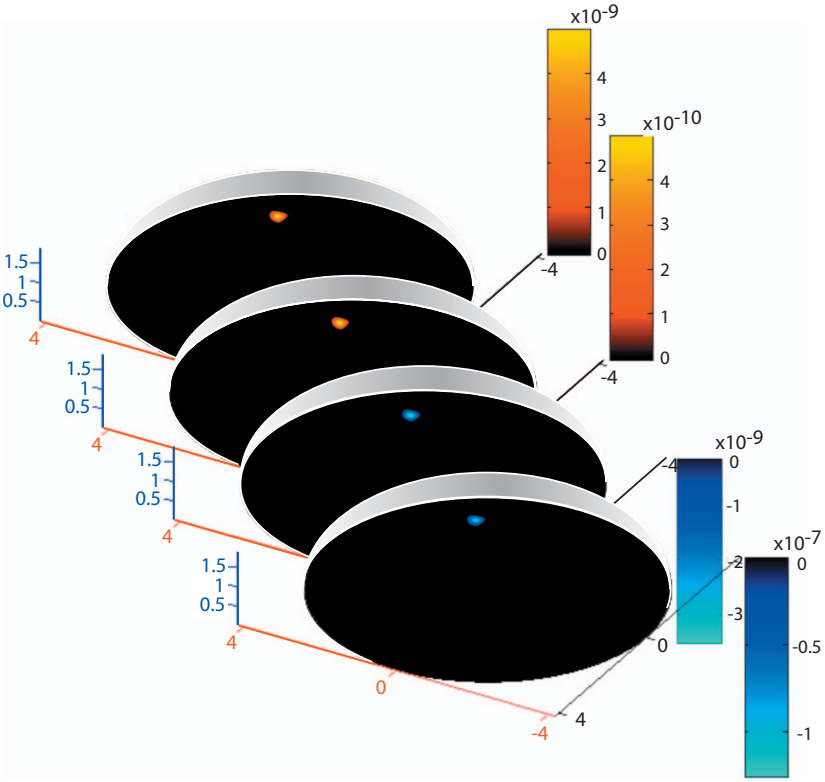


FIG. 7. Normal displacement levels for the flawed ceiling (detached segment) calculated with the STARS3D code for four frequencies (displayed on the exposed surface of the ceiling).

TABLE I. Dynamic displacements, total force acting on flaw, and wavelengths.

Frequency (Hz)	(a) Dynamic fluid flaw displacement ($m\times10^{-10}$)	(b) Dynamic detached flaw displacement ($m\times10^{-9}$)	(c) Total force on flaw (N)	(d) Force-normalized displacement (b)/(c) ($m\times10^{-9}/N$)	(e) λ flexural (m)	(f) λ acoustic (m)
100	1	4.3	0.2	22	1.8	3.44
200	1.2	0.48	0.02	24	1.27	1.72
300	2.2	−3.2	−0.1	32	1.04	1.14
400	110	−90	−0.85	106	0.9	0.86

as can be seen in column (d), the force-normalized response at 400 Hz is an order of magnitude higher. This would suggest that we are beginning to approach a resonance frequency. We say "suggest" because the actual response of the defect would not be simply proportional to the total force but would involve the spatial dependence of the force over the defect in some manner. Nonetheless, earlier using Eq. (6), we had estimated the fundamental resonance frequency of the defect as f_R=336 Hz (with a=25 cm, the radius of a disk whose area would be approximately that of our pie-shaped defect). This is seen to be consistent with the conjecture that at 400 Hz we are near the first resonance of the detached segment. For much larger defects, our band would be well above f_R; and unless higher order resonances were excited, the resulting displacement levels would be greatly reduced from the levels predicted below resonance.

Before leaving detached defects, we point out that oftentimes a plaster wall or ceiling is formed in three or more layers. For example, in many historic structures built in the latter part of the 19th century, three plaster layers were used consisting of the base or scratch-coat layer (also called Trullisatio), the middle or brown coat layer (also known as Arriccio), and the finish or white coat layer (also called Intonaco). By simple extension of the analysis leading up to Eq. (4), when the detachment involves *delamination* between two of these layers, one can show that an appropriate modification of Eq. (4) would be

$$\frac{W_C}{W}=\frac{3(1-\sigma^2)a^4}{tt_i^3}, \tag{7}$$

where t_i is to be interpreted as the distance between the outer plaster surface and the delamination and we have assumed that the elastic parameters and thicknesses of the layers are the same. Since t_i is by definition less than t, smaller delaminated sections can be visualized directly. For example, for delamination of the innermost or outermost layers, the discernable defect size shrinks by factors of 2.3 and 1.4, respectively.

B. Deconsolidation defect

Next, we consider the second defect type, viz., plaster deconsolidation here modeled as a heavy fluid pocket with a somewhat smaller compressibility than the consolidated plaster. The dynamic displacement maps calculated with the finite element code are shown in Fig. 8(b) in the area around the flaw. The displacement levels associated with the deconsolidated segment are again seen to be much larger than

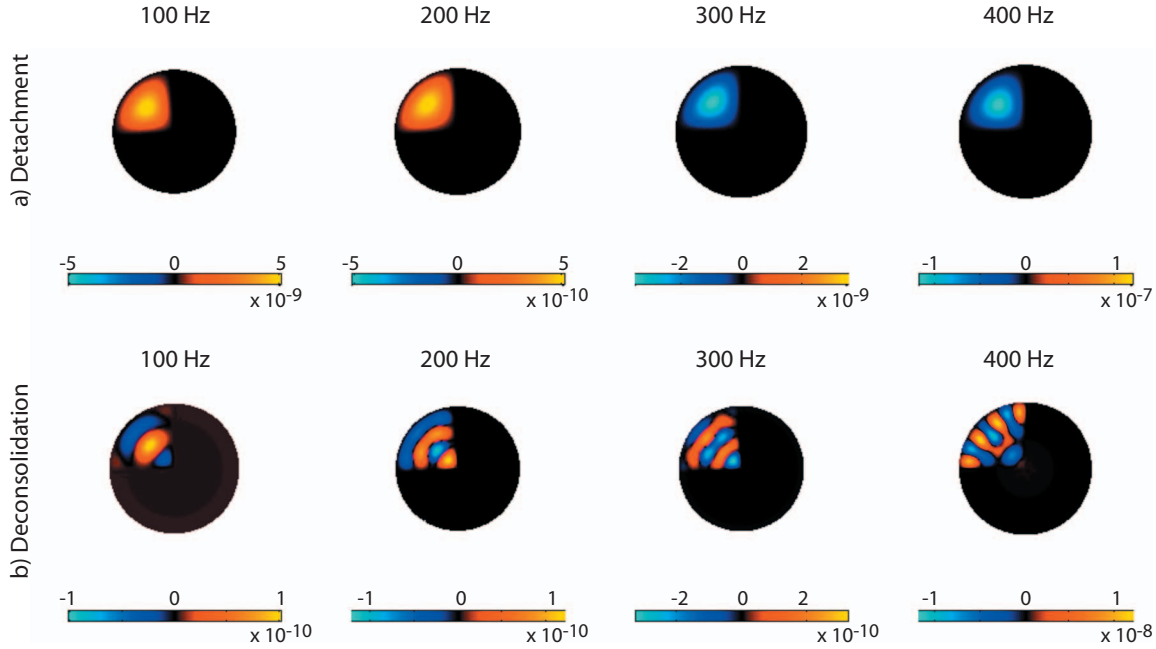


FIG. 8. Normal displacement levels for the flawed ceiling in an area around the flaw calculated with the STARS3D code for four frequencies (displayed on the exposed surface of the ceiling): (a) detached flaw and (b) deconsolidated flaw.

those associated with the unflawed regions of the ceiling although somewhat smaller than for the detached case. Thus this defect can also be detected and localized merely by observing the displacement map itself.

As can be seen in Fig. 8(b), the surface response for the fluid defect case oscillates spatially as one goes over the defect. This is in contrast to what we observed for the detached layer where the spatial response pattern fell off monotonically from the center of the flaw [Fig. 8(a)]. Although the origin of this effect is unclear at the present time, we point out that this feature may provide one the ability to distinguish between detached segments and deconsolidated defects, and this is currently under further investigation.

One explanation we considered for this spatial variation of the surface displacement across the deconsolidated pocket is the following. As shown in Table I, the flexural wavelength λ_{flex} in the 2.54 cm plaster plate is 0.9 m at 400 Hz. Beneath the deconsolidated segment, the plaster layer thickness is reduced by a factor of 3; and since λ_{flex} varies as $t^{1/2}$, it becomes ~ 0.5 m in the lower plaster layer. The defected segment is now of order λ_{flex} on a side, and this could result in a pseudostanding wave spatial modulation. However, closer examination of Fig. 8(b) indicates much larger wavenumbers (smaller wavelengths) associated with the displacements beneath the defect than expected for the flexural wavenumber at these frequencies. This mismatch would seem to indicate that the effect is tied more to the response of the fluid layer itself than the plaster layer above it.

IV. CONCLUSIONS

We have calculated the dynamic surface displacements resulting from acoustic speaker excitation of a flawed and unflawed plaster ceiling and have developed several expressions relating displacement levels to geometric and elastic parameters using static theory. The results support the following conclusions.

(1) We have demonstrated that even relatively small defects result in large displacement levels, which stand out against those of the healthy plaster. But can those displacement levels be detected by a typical laser vibrometer? We find that indeed even away from defect resonances, practical levels of speaker excitation produce easily measured defect surface vibration. For example, scaling our finite element-based simulations predicts that a 74 dB sound pressure level (SPL) speaker source (0.1 Pa) located ~ 2.4 m beneath the ceiling at 300 Hz produces a displacement on the order of 4.3×10^{-10} m (8×10^{-7} m/s velocity) beneath a detached pie-shaped segment of radius 0.5 m and area roughly 0.2 m^2. Commercially available laser vibrometers (e.g., the Polytec PSV 400) have minimum detectable levels $<5 \times 10^{-12}$ m/$\sqrt{\text{Hz}}$ in this frequency range. Over a 1 Hz bandwidth, this corresponds to a minimum measurable displacement of $<5 \times 10^{-12}$ m, which is about a factor of 100 below the predicted defect displacement levels mentioned above. Recognizing that the defect displacement levels are proportional to the square of the defect area [see Eq. (2)] implies that defects as small as 0.15 m ($2a$ for the circular defect) should have displacement levels, which are detectable with such a system. Increasing the source level to 100 dB SPL would allow 0.07 m flaws to be detected at 300 Hz. Also, since the Doppler vibrometer measures velocity, which is $\omega\times$ displacement, one could achieve still improved performance by increasing the frequency.

These minimum size estimates have assumed a plaster layer thickness of 0.0254 m, which is fairly common. Since detachment displacements are proportional to t^{-3} [see Eqs. (2) and (3)], even slightly thinner detached layers would have much higher displacements. For example, for a 25% thickness reduction, the detachment displacement levels increase by a factor of 2.4. Finally, we point out that if the delamination happens between the first and second or second and third layers, significantly smaller detachment sizes should have sufficient contrast and also be detectable [see Eq. (7)].

(2) Unlike the use of locally applied shaker excitation, the architectural acoustics of the room (walls and ceilings) must be taken into account. On the positive side, however, the plane-like waves from the speaker excite far fewer elastic excitations and modal responses in the plaster ceiling layer compared to that generated by the high spatial wavenumbers of the locally applied force of the shaker.

(3) For both the defect types and size simulated here (~ 0.5 m pie-shaped segment of detachment from the supporting structure and plaster deconsolidation), the displacement levels are considerably higher than those of the healthy plaster layer so that these flaws should be detectable by mere observation of the vibration maps.

(4) In general, the directly observable (sufficiently contrasted) minimum defect size is found to be less than two times the plaster layer thickness. For a typical plaster ceiling of 2.54 cm thickness, detachment segments with diameters (circular flaw) or radii (pie-shaped flaw) somewhat smaller than 10 cm should be sufficiently contrasted and thus directly observable. Much smaller sizes should nevertheless be accessible by post-processing of the displacement maps using, for example, successfully reported inversion operators (Bucaro *et al.*, 2004) or by increasing the acoustic frequency. For delamination in ceilings comprised of several layers, and probably for an internal pocket of deconsolidated plaster as well, these directly observable defect sizes are reduced considerably.

(5) In the frequency range studied here (100–400 Hz), the spatial structure in the displacement maps beneath the defect may provide a wavenumber-based feature, which could separate plaster detachment from the other types of defects such as deconsolidation.

ACKNOWLEDGMENTS

One of the authors (J.A.B.) would like to acknowledge many stimulating discussions with the late George W. Adams about plaster wall and ceiling "health" and to thank George posthumously for introducing him to the related work of

monitoring fresco-laden plaster at the U.S. Capitol building. This work was supported by the SERDP Sustainable Infrastructure Program and ONR.

Bucaro, J. A., Romano, A. J., Abraham, P., 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.

Castellini, P., Paone, N., and Tomasini, E. P. (**1994**). "Application of a laser Doppler vibrometer to nonintrusive diagnostic of frescoes damage," *Proceedings of the First International Conference on Vibration Measurements by Laser Techniques: Advances and Applications*, Ancona, Italy (SPIE, Bellingham, WA), Vol. **2358**, pp. 70–77.

Castellini, P., Paone, N., and Tomasini, E. P. (**1996**). "The laser Doppler vibrometer as an instrument for non-intrusive diagnostic of works of art: Application to fresco paintings," Opt. Lasers Eng. **25**, 227–246.

Castellini, P., Esposito, E., Paone, N., and Tomasini, E. P. (**2000**). "Non-invasive measurements of damage of frescoes paintings and icons by laser scanning vibrometer: Experimental results on artificial samples and real works of art," Measurement **28**, 33–45.

Dey, S., and Datta, D. K. (**2006**). "A parallel hp-FEM infrastructure for three-dimensional structural acoustics," Int. J. Numer. Methods Eng. **68**, 583–603.

Dey, S., Shirron, J. J., and Couchman, L. S. (**2001**). "Mid-frequency structural acoustic and vibration analysis in arbitrary, curved three-dimensional domains," Comput. Struct. **79**, 617–629.

Tornari, V., Bonarou, A., Castellini, P., Esposito, E., Osten, W., Kalms, M., Smyrnakis, N., and Sasinopulos, S. (**2001**). "Laser-based systems for the structural diagnostics of artworks: An application to XVII-century Byzantine icons," Proceedings of the Laser Techniques and Systems in Art Conservation Conference, pp. 172–183.

Vignola, J. F., Bucaro, J. A., Lemon, B. R., Adams, G. W., Kurdila, A. J., Marchetti, B., Esposito, E., Tomasini, E., Simpson, H. J., and Houston, B. H. (**2005**). "Locating faults in wall paintings at the U.S. Capitol by shaker-based laser vibrometry," APT Bulletin of the Journal of Preservation Technology **36**, 25–34.

Young, W. C. (**1989**). *Roark's Formulas for Stress and Strain* (McGraw-Hill, New York).

A Burton–Miller inverse boundary element method for near-field acoustic holography

D. J. Chappell[a)]
School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

P. J. Harris
School of Computing and Mathematical Sciences, University of Brighton, Lewes Road, Brighton BN2 4GJ, United Kingdom

(Received 14 October 2008; revised 22 April 2009; accepted 23 April 2009)

An inverse boundary element method based on the Burton–Miller integral equation is proposed for reconstructing the Neumann boundary data from pressure values on a conformal surface in the near-field of an arbitrary radiating object. The accuracy of the reconstruction is compared with that of a method based on the more commonly used Helmholtz integral equation. In particular, the behavior at characteristic frequencies, which are known to be problematic in the Helmholtz integral equation for the forward problem, is examined. The effect of regularization is considered, including the L-curve parameter selection method. Numerical computations are given for noisy data generated from an internal point source. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3133923]

PACS number(s): 43.40.Sk, 43.20.Rz, 43.40.Rj, 43.35.Sx [SFW] Pages: 149–157

I. INTRODUCTION

Near-field acoustic holography (NAH) was first documented in 1980[1,2] as a method for reconstructing acoustic radiation from vibrating structures based on acoustic pressures measured in a hologram plane. Traditionally, this was carried out via Fourier acoustics and so was only suitable for geometries which conform closely to one of the separable geometries of the acoustic wave equation, for example, an infinite plane, an infinite cylinder, or a sphere in a free field. Almost a decade later, an inverse boundary element method (IBEM) based on the Helmholtz integral equation was proposed for the study of arbitrary geometries.[3] A large number of publications on IBEM have since emerged, and more recently with the IBEM combined in hybrid methods which employ expansions of particular solutions of the Helmholtz equation to enrich the measured pressure data.[4,5]

A deficiency of these IBEMs is that the surface integral equation formulations fail to yield unique solutions for the associated forward exterior problems when the excitation frequencies coincide with the eigenfrequencies of the corresponding boundary value problems in the interior.[6] The values of the wave-number at these frequencies are known as characteristic wave-numbers. This deficiency is present in both the direct formulations of the Helmholtz integral equation, as employed in the references above, and the indirect formulations based on single or double layer potential representations of the pressure as employed in Ref. 7, for example. This is in contrast with the existence and uniqueness of solutions of the exterior Neumann problem for the Helmholtz equation (see, for example, Ref. 8) and so the deficiency is due to the integral equation formulation. A number of methods have been suggested for overcoming this problem, including modified Green's function methods[9] and supplementing the discretized system of equations with a few from the interior Helmholtz relationship.[10] However, one of the most robust and widely applicable methods is to take a linear combination of the surface Helmholtz equation and its normal derivative for direct formulations, or a combined single and double layer potential representation of the pressure in the indirect case. These techniques lead to uniquely solvable integral equations and in the direct formulation employed in this work, the method is attributed to Burton and Miller.[6]

There is a level of uncertainty as to whether the nonuniqueness of the surface integral equation described above is indeed problematic for the inverse problem. For example, in Ref. 4 it is inferred that there is likely to be a problem and the method of Schenk[10] is suggested, although not implemented, for its remedy. However, in Ref. 7 it is suggested that the inverse problem will encounter acceptable errors since it is ill-posed and so a regularized solution of minimal norm is sought, which is unique. These issues are addressed in this work by considering the behavior of the problem near the known characteristic wave-numbers for an acoustically radiating sphere when both the surface Helmholtz equation and the method of Burton and Miller are employed.

The regularization is based on the standard Tikhonov method as employed in Ref. 7. Since only computational test problems are considered in this work, exact solutions are available for waves generated by an internal point source. Hence the regularization parameter is chosen to minimize the l^2 relative error of the reconstructed solution over a range of parameters tested. This will also be used to study the suitability of the L-curve parameter selection method for NAH problems, which the present writer believes is the first time this has been done. The application and suitability of these

a)Author to whom correspondence should be addressed. Electronic mail: david.chappell@nottingham.ac.uk

methods will also be considered for the reconstruction of the Neumann boundary data on a cuboid of similar dimensions to a typical loudspeaker cabinet over a range of frequencies. This is of interest since this study was initially motivated by the application to acoustic radiation from loudspeakers.

II. BOUNDARY INTEGRAL FORMULATION

Let $\Omega \subset \mathbb{R}^3$ be a finite domain with boundary surface Γ. Let $\Omega_+ = \mathbb{R}^3 \backslash \bar{\Omega}$ denote the unbounded exterior domain, which is assumed to be filled with a homogeneous compressible acoustic medium with speed of sound c. For a time-harmonic disturbance of frequency ω, the sound pressure u satisfies the homogeneous Helmholtz equation in Ω_+,

$$\Delta u + k^2 u = 0, \tag{1}$$

where $k = \omega / c$ is the wave-number. Since this work considers an unbounded exterior domain, then u must also satisfy the Sommerfeld radiation condition

$$\lim_{R \to \infty} R \left\{ \frac{\partial u}{\partial R} - iku \right\} = 0. \tag{2}$$

A function u satisfying Eqs. (1) and (2) is called a radiating solution. The pressure data are assumed to be known at discrete points on a surface $\Gamma_0 \subset \Omega_+$. These data are usually obtained from measurements, although only test problems where the data are generated from a point source in Ω are considered here. The pressure data will then be used to recover the Neumann boundary data on Γ.

The solution of the related forward Neumann problem using a direct integral equation formulation is given by the Helmholtz integral equation

$$\int_\Gamma \left(u(\mathbf{y}) \frac{\partial G_k(\mathbf{x},\mathbf{y})}{\partial \mathbf{n}_y} - G_k(\mathbf{x},\mathbf{y}) \frac{\partial u(\mathbf{y})}{\partial \mathbf{n}_y} \right) d\Gamma_y = \begin{cases} \frac{1}{2} u(\mathbf{x}), & \mathbf{x} \in \Gamma \\ u(\mathbf{x}), & \mathbf{x} \in \Omega_+, \end{cases} \tag{3}$$

where G_k is the free space Green's function for Helmholtz equation in three dimensions given by

$$G_k(\mathbf{x},\mathbf{y}) = \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}. \tag{4}$$

In addition, $\mathbf{n}_y$ is the outward unit normal to Γ at $\mathbf{y} \in \Gamma$ and it is assumed that Γ is locally differentiable at $\mathbf{x}$. Hence given boundary data $\partial u / \partial \mathbf{n}$ on Γ, one can employ Eq. (3) for $\mathbf{x} \in \Gamma$ to obtain u on Γ and then evaluate u in Ω_+ using Eq. (3) with $\mathbf{x} \in \Omega_+$.

Let us introduce the single and double layer potential integral operators:

$$\mathcal{S}u(\mathbf{x}) := \int_\Gamma G_k(\mathbf{x},\mathbf{y}) u(\mathbf{y}) d\Gamma_y, \quad \mathbf{x} \in \Omega_+ \tag{5}$$

and

$$\mathcal{D}u(\mathbf{x}) := \int_\Gamma \frac{\partial G_k}{\partial \mathbf{n}_y}(\mathbf{x},\mathbf{y}) u(\mathbf{y}) d\Gamma_y, \quad \mathbf{x} \in \Omega_+, \tag{6}$$

respectively. The case $\mathbf{x} \in \Gamma$ is distinguished by denoting the single layer potential as V and the double layer potential as K. The boundary integral equations (3) may be written in terms of these integral operators. Hence the solution procedure described above may be expressed as

$$u(\mathbf{x}) = \left(\mathcal{D} \left[\left(-\frac{1}{2} I + K \right)^{-1} V \right] - \mathcal{S} \right) \frac{\partial u}{\partial \mathbf{n}}, \quad \mathbf{x} \in \Omega_+, \tag{7}$$

where I is the identity operator as usual.

Unfortunately, the procedure outlined above has a major drawback since for $\mathbf{x} \in \Gamma$, the integral equation (3) has nonunique solutions at certain values of the wave-number $k > 0$ (see Ref. 6), known as characteristic wave-numbers. The method of Burton and Miller is applied to avoid this problem due to its robustness and relatively wide applicability. It can be shown that a linear combination of Eq. (3) for $\mathbf{x} \in \Gamma$ and its normal derivative at $\mathbf{x}$ in the form

$$\left(-\frac{1}{2} I + K + \alpha \frac{\partial}{\partial \mathbf{n}_x} K \right) u(\mathbf{x}) = \left(\frac{\alpha}{2} I + V + \alpha \frac{\partial}{\partial \mathbf{n}_x} V \right) \frac{\partial u}{\partial \mathbf{n}_x}(\mathbf{x}), \tag{8}$$

where α is a coupling constant, has a unique solution for all real and positive wave-numbers provided that $\mathrm{Im}\, \alpha \neq 0$. In fact, it can be shown further that a choice of $\alpha = i\{\min(\mathrm{diam}(\Gamma), 1/k)\}$ is almost optimal in terms of approximately minimizing the condition number of the resulting integral operator[11] in the case where Γ is a sphere. Hence this is the choice of α taken in this work. The Burton–Miller formulation may be implemented in the solution procedure (7) by simply replacing the term in the square brackets with

$$\left(-\frac{1}{2} I + K + \alpha \frac{\partial}{\partial \mathbf{n}_x} K \right)^{-1} \left(\frac{\alpha}{2} I + V + \alpha \frac{\partial}{\partial \mathbf{n}_x} V \right). \tag{9}$$

III. BOUNDARY ELEMENT DISCRETIZATION

In this section the discretization of the integral equation formulation above using boundary elements is considered in order to obtain a linear matrix system. The integration procedures are then described and the solution of the resulting ill-posed inverse problem and its regularization are discussed.

A simple boundary element discretization Γ^* of Γ by flat triangular elements is employed, and the solution is approximated using piecewise constant collocation. The approximate solution $\hat{v} \approx \partial u / \partial \mathbf{n}$ may be expressed in the form

$$\hat{v}(\mathbf{y}) = \sum_{l=1}^{N} \hat{v}_l b_l(\mathbf{y}), \quad \mathbf{y} \in \Gamma^* \tag{10}$$

where b_l denotes the (piecewise constant) boundary element basis functions and N is the number of collocation points. The values $\hat{v}_l$ are to be determined and give $\hat{v}$ at the lth collocation point or node, which are taken at the centroid of each element as usual. Similarly, u is approximated on Γ by

$$\hat{u}(\mathbf{y}) = \sum_{l=1}^{N} \hat{u}_l b_l(\mathbf{y}), \quad \mathbf{y} \in \Gamma^*. \tag{11}$$

Let $\Gamma_0 \subset \Omega_+$ be conformal to Γ^* at a distance δ. Taking $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ to be the collocation points on Γ^*, then the points in $\Gamma_0 \subset \Omega_+$ at which the pressure data are known are given by $\mathbf{x}_j = \mathbf{y}_j + \delta \mathbf{n}_{y_j}$ for $j = 1, \ldots, N$. Substituting Eq. (10) into the definition of $\mathcal{S}$ [Eq. (5)] and evaluating at the exterior solution points yields the discrete operator $\mathcal{S}_N$ given by

$$\mathcal{S}_N \hat{v}(\mathbf{x}_j) = \sum_{l=1}^{N} \hat{v}_l \int_{\Gamma_l^*} G_k(\mathbf{x}_j, \mathbf{y}) d\Gamma_y \tag{12}$$

for $j = 1, \ldots, N$ and let $(S)_{l,j} = \int_{\Gamma_l^*} G_k(\mathbf{x}_j, \mathbf{y}) d\Gamma_y$ for $l, j = 1, \ldots, N$. Here Γ_l^* denotes the lth element of the boundary element discretization Γ^* where $b_l = 1$ (it is zero elsewhere) so that $\Gamma^* = \cup_{l=1}^{N} \Gamma_l^*$. Likewise, the double layer potential operator $\mathcal{D}$ may be discretized to give a discrete operator $\mathcal{D}_N$ and a related $N \times N$ matrix D. Substituting Eq. (10) into the definition of V and evaluating at the surface collocation points yields the discrete operator V_N given by

$$V_N \hat{v}(\mathbf{y}_j) = \sum_{l=1}^{N} \hat{v}_l \int_{\Gamma_l^*} G_k(\mathbf{y}_j, \mathbf{y}) d\Gamma_y \tag{13}$$

for $j = 1, \ldots, N$ and piecewise constant basis functions. Discretized forms of K and the normal derivatives of V and K may be written in a similar form. As above, in matrix-vector form these discretized operators appear as $N \times N$ matrices. Therefore following discretization, the term in square brackets in Eq. (7) may be written as an $N \times N$ matrix A, whether it is obtained from the surface Helmholtz or the Burton–Miller integral equation. Note that when k is a characteristic wave-number, the term in square brackets in Eq. (7) is not well defined since $(-I/2 + K)$ is not injective. However, the resulting matrix term is not singular due to the perturbation by the discretization, but will be very poorly conditioned.[12] Let $\mathbf{u}_0$ be the vector of known pressure data on Γ_0 and $\hat{\mathbf{v}} := [\hat{v}_1, \ldots, \hat{v}_N]^T$. Hence Eq. (7) or the Burton–Miller version of this equation may be discretized and written in the matrix-vector form

$$\mathbf{u}_0 = (DA - S)\hat{\mathbf{v}}. \tag{14}$$

Therefore the inverse problem is to obtain $\hat{\mathbf{v}}$ by inverting the $N \times N$ matrix $(DA - S)$.

A. Integration procedures

For the integrations to compute S and D, the integrands are continuous and so numerical integration based on standard Gaussian quadrature is employed. However, for the computation of A, a singularity subtraction procedure is employed in order to simplify the integration methods and provide efficient calculations for multiple frequencies. For the single layer potential operator this singularity subtraction is employed by making the splitting

$$Vu(\mathbf{x}) = \int_{\Gamma} \frac{(e^{ik|\mathbf{x}-\mathbf{y}|} - 1)}{4\pi|\mathbf{x}-\mathbf{y}|} u(\mathbf{y}) d\Gamma_y + \int_{\Gamma} \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|} u(\mathbf{y}) d\Gamma_y, \quad \mathbf{x} \in \Gamma. \tag{15}$$

It can be shown using Taylor's theorem that the first term is continuous and so may be evaluated using standard quadrature rules as above. Only the simpler second term now contains a singularity, which following discretization occurs in the case when $\mathbf{x}$ and $\mathbf{y}$ are in the same element. For discretization by flat triangular elements as described above this integration may be computed analytically and standard quadrature rules applied when $\mathbf{x}$ and $\mathbf{y}$ are in different elements.

A similar procedure may be carried out for the double layer potential K and the normal derivative of the single layer potential $(\partial/\partial \mathbf{n}_x)V$. In these cases the following splitting is applied:

$$\int_{\Gamma} (\mathbf{n} \cdot \nabla |\mathbf{x}-\mathbf{y}|) \frac{((ik|\mathbf{x}-\mathbf{y}| - 1)e^{ik|\mathbf{x}-\mathbf{y}|} + 1)}{4\pi|\mathbf{x}-\mathbf{y}|^2} u(\mathbf{y}) d\Gamma_y - \int_{\Gamma} \frac{(\mathbf{n} \cdot \nabla |\mathbf{x}-\mathbf{y}|)}{4\pi|\mathbf{x}-\mathbf{y}|^2} u(\mathbf{y}) d\Gamma_y, \quad \mathbf{x} \in \Gamma. \tag{16}$$

Note that the normal $\mathbf{n}$ and the gradient ∇ are taken with respect to $\mathbf{y}$ for K and with respect to $\mathbf{x}$ for $(\partial/\partial \mathbf{n}_x)V$. Again the first term is continuous and the second term is the only one to contain a singularity when $\mathbf{x}$ and $\mathbf{y}$ are in the same element. In this case the integral is simply zero for the flat triangular elements employed here. The situation is a little more complicated for the normal derivative of the double layer potential $(\partial/\partial \mathbf{n}_x)K$, not least since this contains a hypersingular kernel function. The following splitting is applied:

$$\frac{\partial}{\partial \mathbf{n}_x} Ku(\mathbf{x}) = \int_{\Gamma} \left\{ (\mathbf{n}_x \cdot \mathbf{n}_y) \frac{2(1 - ik|\mathbf{x}-\mathbf{y}|)e^{ik|\mathbf{x}-\mathbf{y}|} - 2 - k^2|\mathbf{x}-\mathbf{y}|^2}{8\pi|\mathbf{x}-\mathbf{y}|^2} + (\mathbf{n}_x \cdot \nabla_x |\mathbf{x}-\mathbf{y}|)(\mathbf{n}_y \cdot \nabla_y |\mathbf{x}-\mathbf{y}|) \right.$$
$$\left. \times \frac{(3 - 3ik|\mathbf{x}-\mathbf{y}| - k^2|\mathbf{x}-\mathbf{y}|^2)e^{ik|\mathbf{x}-\mathbf{y}|} - 3}{4\pi|\mathbf{x}-\mathbf{y}|^3} \right\} u(\mathbf{y}) d\Gamma_y + \text{⨍}\!\!\!\!\!\!-_{\Gamma} \frac{(\mathbf{n}_x \cdot \mathbf{n}_y) + 3(\mathbf{n}_x \cdot \nabla_x |\mathbf{x}-\mathbf{y}|)(\mathbf{n}_y \cdot \nabla_y |\mathbf{x}-\mathbf{y}|)}{4\pi|\mathbf{x}-\mathbf{y}|^3} u(\mathbf{y}) d\Gamma_y$$
$$+ k^2 \int_{\Gamma} \frac{1}{8\pi|\mathbf{x}-\mathbf{y}|} u(\mathbf{y}) d\Gamma_y, \quad \mathbf{x} \in \Gamma. \tag{17}$$

See the Appendix for details on how this splitting is derived using Taylor series so that the first integral is continuous, leaving just the two remaining integrals which contain a singularity when $\mathbf{x}$ and $\mathbf{y}$ are in the same element. The last integral is just a constant multiple of the singular integral from the single layer potential case and so may also be evaluated analytically for the above described discretization by flat triangular elements. The second integral is the most problematic since it contains a hypersingular integrand and hence may only be understood as a Hadamard finite part integral denoted by $\oint\!\!\!\!=$. However, this integral may still be evaluated using standard quadrature rules when $\mathbf{x}$ and $\mathbf{y}$ are in different elements. In Ref. 13 it is shown that

$$\oint\!\!\!\!=_{\Gamma} \frac{(\mathbf{n}_x \cdot \mathbf{n}_y) + 3(\mathbf{n}_x \cdot \nabla_x |\mathbf{x}-\mathbf{y}|)(\mathbf{n}_y \cdot \nabla_y |\mathbf{x}-\mathbf{y}|)}{4\pi |\mathbf{x}-\mathbf{y}|^3} d\Gamma_y = 0. \tag{18}$$

This implies that for discretization by piecewise constant basis functions, the second integral in Eq. (17) for $\mathbf{x}$ and $\mathbf{y}$ in the same element is given by the negative of the sum of the integrals when $\mathbf{y}$ is in each of the other elements. A related study is given in Ref. 14, where a closed analytical expression for the hypersingular integral over the singular element is derived.

B. Regularization

The operator expression in Eq. (7) is bounded from $L^2(\Gamma^*) \rightarrow L^2(\Gamma_0)$ when k is not a characteristic wave-number and otherwise is not well defined. For the Burton–Miller formulation (9) the corresponding operator expression is bounded from $L^2(\Gamma^*) \rightarrow L^2(\Gamma_0)$ for all wave-numbers. Therefore the inverse problem under consideration is ill-posed for $\partial u/\partial \mathbf{n}$ in $L^2(\Gamma)$ (see Ref. 12, Chap. 18). For experimental problems, the pressure measurements will contain errors. The ill-posedness of the problem means that these errors are amplified in the solutions, usually rendering them meaningless. For this reason it is necessary to consider appropriate regularization methods. As remarked in Ref. 15 many numerical regularization techniques, from a practical point of view, produce the same regularized solutions. One of the oldest and most widely used basic methods is Tikhonov regularization, which applied to Eq. (14) yields

$$B^* \mathbf{u}_0 = (B^* B + \lambda^2 I)\hat{\mathbf{v}}_\lambda \tag{19}$$

for $B = DA - S$. Here $\lambda > 0$ is a quantity to be determined called the regularization parameter and Eq. (19) has a unique solution provided B is bounded.[12] Note that this condition is not satisfied at characteristic wave-numbers when using the surface Helmholtz formulation. Other more sophisticated regularization methods are available, of particular note is the modified Tikhonov procedure introduced in Ref. 16, which incorporates a high-pass filter. However, the basic Tikhonov method is employed here for simplicity since the initial aim in this work is to provide a fair comparison between the Burton–Miller formulation and the surface Helmholtz integral equation. Since an exact solution is known for the test problems here, a good choice of regularization parameter is attainable by calculating the error in the approximate solution for a range of λ and then choosing the value to minimize this error. In the absence of an exact solution a number of methods are available to determine this parameter, which are briefly discussed below.

The data $\mathbf{u}_0$ in the test problems here will have noise added in order to replicate the noise in experimental observations. Let $\mathbf{u}_0^\varepsilon$ be the data with noise added such that

$$\|\mathbf{u}_0 - \mathbf{u}_0^\varepsilon\|_2 \leqslant \varepsilon, \tag{20}$$

then the following stability estimate holds:[17]

$$\|\hat{\mathbf{v}}_\lambda - \hat{\mathbf{v}}_\lambda^\varepsilon\|_2 \leqslant \frac{\varepsilon}{\lambda}, \tag{21}$$

where $\hat{\mathbf{v}}_\lambda^\varepsilon$ is the solution of Eq. (19) with $\mathbf{u}_0$ replaced by $\mathbf{u}_0^\varepsilon$. Adding regularization damps out contributions to the solution from the data errors reducing $\|\hat{\mathbf{v}}_\lambda^\varepsilon\|_2$. If too much regularization is imposed on the solution (i.e., λ is too large), then it will not fit the given data $\mathbf{u}_0^\varepsilon$ properly and the residual norm $\|B\hat{\mathbf{v}}_\lambda^\varepsilon - \mathbf{u}_0^\varepsilon\|_2$ will be large. If too little regularization is imposed on the solution (i.e., λ is too small), then the fit will be good, but the solution will be dominated by contributions from the data errors and $\|\hat{\mathbf{v}}_\lambda^\varepsilon\|_2$ will be large. Hence when choosing the parameter λ, it is natural to plot $\{\|B\hat{\mathbf{v}}_\lambda^\varepsilon - \mathbf{u}_0^\varepsilon\|_2, \|\hat{\mathbf{v}}_\lambda^\varepsilon\|_2\}$ for a range of $\lambda > 0$. This plot usually takes a distinctive "L" shape and is hence named the L-curve.[15] The L-curve shows the trade-off between two quantities that should be controlled. Corollary 7 of Ref. 15 shows that a good choice of the regularization parameter is given immediately to the right of the corner of the L-curve.

The choice of parameter obtained through comparison with the exact solution is compared with the L-curve to assess the suitability of this method for NAH problems, rather than more commonly used methods such as generalized cross validation (GCV) or the discrepancy principle.[16] One limitation of the L-curve method is that it fails when the solution is very smooth.[18] This limitation is not of concern here since data from experimental measurements will always contain some degree of (non-smooth) noise. The other main limitation of the L-curve method is that it may not pick a consistently optimal parameter as the size of the problem N increases.[19] However, since the size of NAH problems is usually fixed and related to the number of measurements taken, then this limitation is also not too worrying. The most notable alternative parameter selection method for NAH problems is GCV since like the L-curve method, it requires no prior knowledge of noise variance. Clearly this is advantageous when the data are obtained from experimental measurements. These methods are compared in Ref. 15, where it is deduced that the L-curve method is more robust to different types of noise. It has also been observed that for NAH problems, using the GCV in combination with standard Tikhonov regularization tends to under-smooth the solution, i.e., pick λ too small.[16]

IV. NUMERICAL RESULTS

Numerical results for acoustic radiation from two geometrically different test problems are considered. First, a unit sphere discretized by 320 flat triangular elements and second

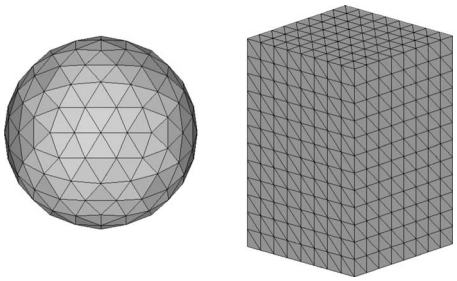


FIG. 1. The discretization of the sphere and cuboid by flat triangular elements.

a cuboid of similar dimensions to a typical loudspeaker cabinet ($0.28\times0.28\times0.42$ m^3) discretized by 1024 triangular elements. These meshes are shown in Fig. 1. The first example is considered since the exact locations of the characteristic wave-numbers are known for spheres and the second since it replicates a simplified version of the real world problem that motivates this study. In both cases the boundary data generated by a point source at $\mathbf{x}_p=(0,0,0.1)\in\Omega$ are reconstructed on Γ^*, where Ω is centered at the origin. The pressure data are therefore of the form

$$(\mathbf{u}_0)_j=\rho\frac{e^{ik|\mathbf{x}_p-\mathbf{x}_j|}}{|\mathbf{x}_p-\mathbf{x}_j|},\quad j=1,\ldots,N, \tag{22}$$

where $\rho\in\mathbb{C}$ is the strength of the source, which in these examples is arbitrarily taken to be $3-i$. The boundary data generated at $\mathbf{y}\in\Gamma^*$ may also be obtained for this point source by replacing $\mathbf{x}_j$ in Eq. (22) by $\mathbf{y}\in\Gamma^*$, differentiating in the direction of $\mathbf{n}_y$ and evaluating at $\mathbf{y}=\mathbf{y}_j$ for $j=1,\ldots,N$ to give

$$(\mathbf{v})_j=\rho\left(\frac{\mathbf{n}_{y_j}\cdot(\mathbf{x}_p-\mathbf{y}_j)}{|\mathbf{x}_p-\mathbf{y}_j|^3}(1-ik|\mathbf{x}_p-\mathbf{y}_j|)e^{ik|\mathbf{x}_p-\mathbf{y}_j|}\right),\quad j=1,\ldots,N. \tag{23}$$

Using this calculation it is possible to verify the accuracy of the approximate solutions and hence approximately determine an optimal regularization parameter λ for which the error is minimized over the values of λ tested.

A. Radiation from a sphere

Consider the distance δ between the triangulated sphere Γ^* and the conformal surface $\Gamma_0\subset\Omega_+$. In order to choose δ so that accurate results are obtained, the trade-off between the problem becoming increasingly ill-posed as δ gets larger (due to losing evanescent wave components) and the near singularity problem of the BEM formulation for δ very small must be considered. A sensible choice is to take δ as small as possible, but strictly larger than the maximum node-vertex distance Δx for each element. For this example $\Delta x=0.188$ to three significant figures, and so a choice of $\delta=0.19$ follows the above advice.

First consider the case $k=1$, where the Helmholtz integral equation is expected to work well since it is not close to a characteristic wave-number. The upper subplot of Fig. 2 shows the real part of the recovered vector $\hat{\mathbf{v}}_\lambda^\varepsilon$ of the approximate values of $\partial u/\partial\mathbf{n}$ at the BEM nodes, computed using the

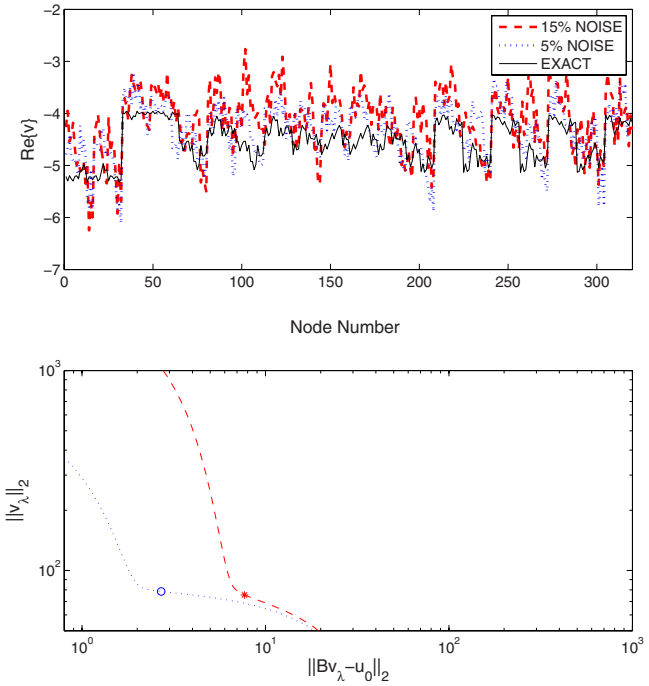


FIG. 2. (Color online) The real part of the reconstructed Neumann boundary data for the unit sphere compared against the exact values (upper) and the position of the minimum error solution on the L-curve (lower) using the surface Helmholtz integral formulation with $k=1$.

surface Helmholtz integral equation formulation. This is plotted against the exact values, computed using the point source expression (23). The plots show the recovered solution with λ taken to give the minimum l^2 relative error (of 50 logarithmically scaled values tested between 10^{-5} and 1) when 15% and 5% noises are added to the exterior pressure data. The l^2 relative error is calculated using

$$\sqrt{\frac{\sum_{j=1}^N|(\hat{\mathbf{v}}_\lambda^\varepsilon)_j-(\mathbf{v})_j|^2}{\sum_{j=1}^N|(\mathbf{v})_j|^2}}. \tag{24}$$

The lower subplot shows the L-curve plots for each of these recovered solutions with the asterisk (15% noise) and circle (5% noise) markers indicating the position on this curve of the minimum error solution. Figure 2 shows that reasonably good reconstructions for this problem are obtained. Table I gives the actual errors obtained and the values of λ used to achieve these errors. This also shows the results when the Burton–Miller formulation is employed, which here gave similar but slightly less accurate results. Note that the error and parameter values quoted in this paper are all given to three significant figures.

Now consider the case $k=\pi$, which corresponds to the first characteristic wave-number for the example of the unit sphere. Figure 3 gives the equivalent plot to Fig. 2 for the case $k=\pi$. Figure 4 gives the equivalent plot to Fig. 3 for the Burton–Miller integral equation formulation. Figures 3 and 4 show that both integral equation formulations give reasonably good reconstructions for this problem. The actual errors obtained are again summarized in Table I, along with the corresponding values of λ employed in the regularization. The error when using the Burton–Miller formulation is now

TABLE I. The minimum error in the reconstructed Neumann boundary data and the corresponding regularization parameters for the sphere over selected wave-numbers. Results are given for both the Helmholtz integral equation (H) and the Burton–Miller integral equation (BM) formulations.

k	Formulation	Noise (%)	λ	l^2 relative error
1	H	5	0.105	0.0970
	H	15	0.160	0.138
	BM	5	0.105	0.102
	BM	15	0.160	0.144
π	H	5	0.0869	0.0540
	H	15	0.0910	0.0988
	BM	5	0.0494	0.131
	BM	15	0.0791	0.184
3.19	H	5	0.0869	0.996
	H	15	0.160	0.998
	BM	5	0.0494	0.127
	BM	15	0.0791	0.183

approximately twice the error when using the Helmholtz integral equation. Also notice that more regularization is required when using noisier pressure data as expected and when using the Helmholtz formulation rather than the Burton–Miller formulation.

Initially these results seem to strongly suggest that the Burton–Miller formulation is not only unnecessary but also unfavorable. However, a key point to note is that since the sphere has been discretized by flat triangles, the positions of the characteristic frequencies are slightly perturbed from those of a sphere. This means that $k=\pi$ is not actually a characteristic frequency for this problem but is close to one. In the forward problem, the discretized equations become

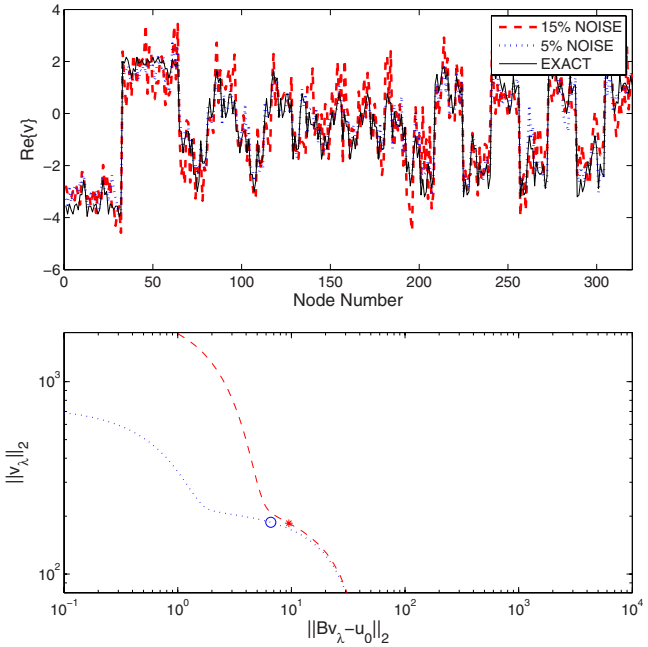


FIG. 3. (Color online) The real part of the reconstructed Neumann boundary data for the unit sphere compared against the exact values (upper) and the position of the minimum error solution on the L-curve (lower) using the surface Helmholtz integral formulation with $k=\pi$.

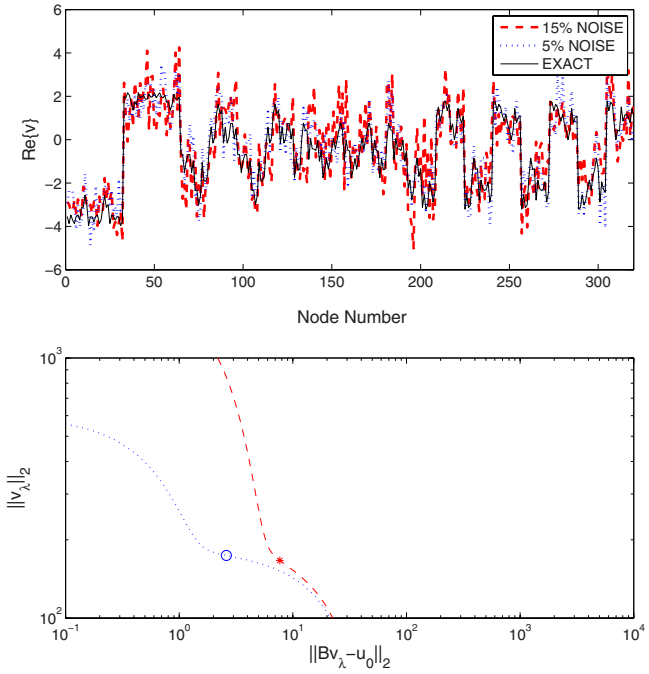


FIG. 4. (Color online) The real part of the reconstructed Neumann boundary data for the unit sphere compared against the exact values (upper) and the position of the minimum error solution on the L-curve (lower) using the Burton–Miller formulation with $k=\pi$.

ill-conditioned close to the characteristic frequencies (see, for example, Ref. 20). This extra ill-conditioning due to the characteristic frequency problem can be seen by comparing the L-curves in Figs. 3 and 4 since for the Burton–Miller formulation the optimum value of λ occurs just to the right of the L-curve corner as expected, but for the Helmholtz formulation a larger than expected choice of λ is required. The extra regularization required to attain the minimum error solution is due to the characteristic frequency related ill-conditioning and is not detected by the L-curve parameter selection method. This appears less significant when more noise is present as would be expected since the characteristic frequency related ill-conditioning becomes relatively less significant as a part of the total ill-conditioning of the problem. However, it will become more significant closer to the exact value of the characteristic frequency. It has been verified that the GCV also fails to detect the characteristic frequency related ill-conditioning, although the calculations are omitted. If λ is chosen according to the L-curve criterion in the Helmholtz formulation then for 5% noise, $\lambda=0.0373$ and the error is 0.136. For 15% noise, $\lambda=0.0610$ and the error is 0.161. In the light of these discussions the Burton–Miller formulation becomes more favorable.

A numerical study to find k for which $1/2$ is an eigenvalue of the discretized surface double layer potential K_N reveals, as a consequence of the Fredholm alternative, that the characteristic frequency $k=\pi$ for a sphere has been perturbed to approximately $k=3.19$ (to three significant figures) for the triangulated sphere. Table I gives the errors in the reconstructed solutions when both the Helmholtz and Burton–Miller formulations are employed. For the Helmholtz integral equation, the error when both 5% and 15% noises are added to the data is approximately 100%, which is

clearly unacceptably high. However, the Burton–Miller formulation gives a reasonably accurate reconstruction of the solution and the results are very similar to those shown for the $k=\pi$ case. For this reason the plot is omitted, but note that the L-curve method approximately selects the optimum regularization parameter.

The above results show that while the surface Helmholtz formulation may potentially give more accurate reconstructions away from characteristic wave-numbers, it is rendered useless when k coincides with a characteristic wave-number. Although these only form a discrete set of the set of positive real numbers, they are known to be dense as $k\to\infty$. Also, the closer k is to a characteristic wave-number, the more ill-conditioned the associated matrix equation becomes and so more regularization is required. The additional regularization required due to the characteristic wave-number problem is not detected by standard regularization parameter selection methods such as the L-curve method and GCV, and hence the reconstructed solutions are under-smoothed. The error in the reconstructed solutions using such parameter selection methods is therefore less predictable and in the example here is similar to the error when the Burton–Miller formulation is employed. Since the characteristic wave-numbers become closer together as k increases, these deficiencies of the surface Helmholtz formulation become more problematic for larger values of k. This may become increasingly important as higher frequency problems become more tractable for NAH due to increasing computer power, the development of more sophisticated measurement equipment, and the development of innovative hybrid methods such as in Ref. 5, for example. The above results also demonstrate that the L-curve parameter selection method works very well when the Burton–Miller formulation is employed.

B. Radiation from a cuboid

The Burton–Miller formulation for the example of the loudspeaker cabinet sized cuboid described above is now considered due to its robustness and the success of the L-curve method in choosing an optimal regularization parameter. A range of values for the wave-number k are considered in the range that can be modeled by the given pressure data. Since the maximum internodal distance between a given node and its four closest neighbors is ≈ 0.035, the maximum value of k that may be considered to maintain six nodes per wavelength is approximately $k=2\pi/0.21\approx 30$. As before δ is taken to be small but greater than the maximum node-vertex distance $\Delta x=0.0261$, to three significant figures, and so $\delta=0.03$.

Table II shows the l^2 relative errors in the regularized reconstructed solutions and the values of λ chosen to minimize these errors for the range of regularization parameters considered. These errors are shown for four different values of k and with both 5% and 15% noises added to the pressure data as before. The values of k are spaced over the range that can be modeled, and the choice $k=17.52$ corresponds to an approximate characteristic wave-number determined from a numerical study as before. Analogous plots to those for the sphere are given in Figs. 5–7, respectively.

TABLE II. The minimum error in the reconstructed Neumann boundary data and the corresponding regularization parameters using the Burton–Miller formulation for the cuboid over a range of wave-numbers.

k	Noise (%)	λ	l^2 relative error
1	5	0.0133	0.154
	15	0.0233	0.226
10	5	0.0115	0.110
	15	0.0184	0.178
17.52	5	0.00910	0.103
	15	0.0146	0.157
25	5	0.00719	0.102
	15	0.0115	0.166

Table II shows that the error in the reconstruction is reasonable over the range of wave-numbers considered. The larger errors for the lower values of the wave-number may be attributed to the relatively poor accuracy of the Burton–Miller formulation for low frequencies (below the first characteristic wave-number), see, for example, the calculations in Chap. 3 of Ref. 20. The table also reveals that less regularization is required as the wave-number increases, suggesting that the ill-posedness of the problem is becoming less severe. This could be attributed to the fact that more of the wave is captured in the distance between the object and the data points due to the shorter wavelength, enabling the solution of the inverse problem to be better determined.

Figures 5–7 display a selection of the tabulated results more clearly, with the case $k=1$ omitted due to its similarity to the plot for $k=10$. In particular, they show the impact of

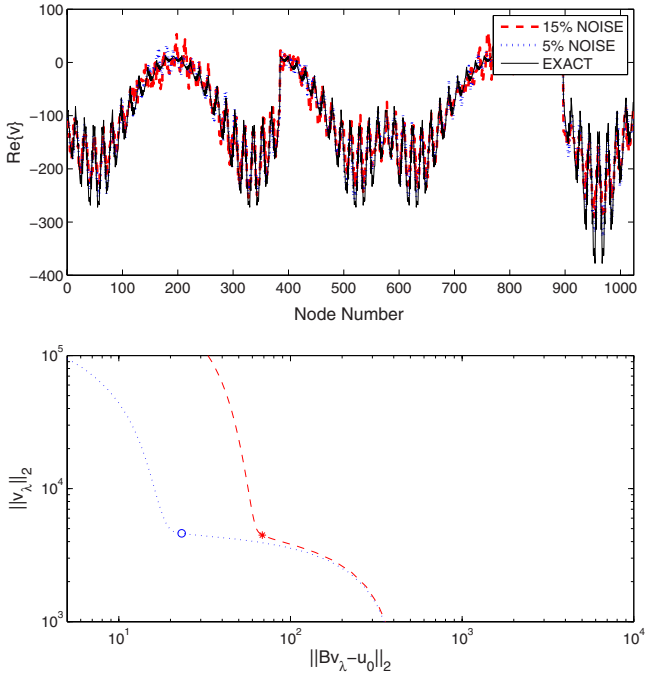


FIG. 5. (Color online) The real part of the reconstructed Neumann boundary data for the cuboid compared against the exact values (upper) and the position of the minimum error solution on the L-curve (lower) using the Burton–Miller formulation with $k=10$.

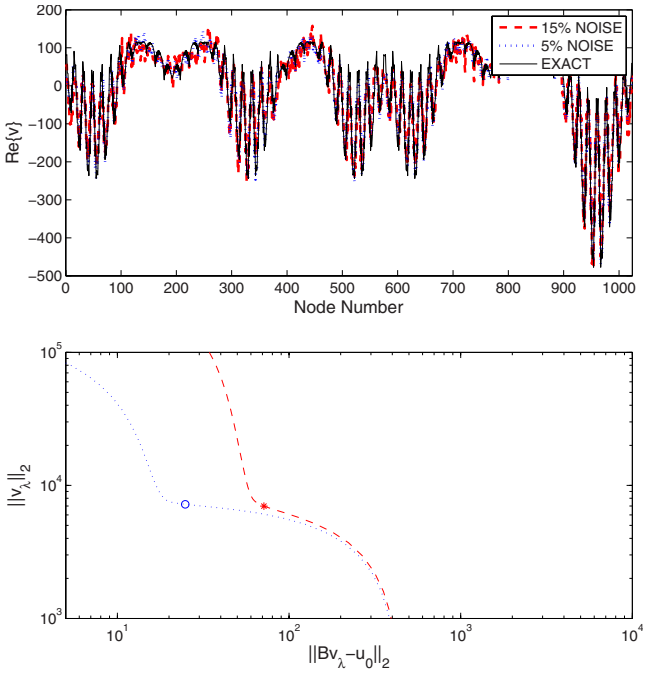


FIG. 6. (Color online) The real part of the reconstructed Neumann boundary data for the cuboid compared against the exact values (upper) and the position of the minimum error solution on the L-curve (lower) using the Burton–Miller formulation with $k=17.52$.

additional noise on the reconstructed solutions and the success of the L-curve method in identifying the approximately optimal regularization parameter. In all cases this optimal parameter is located just to the right of the corner in the L-curve.

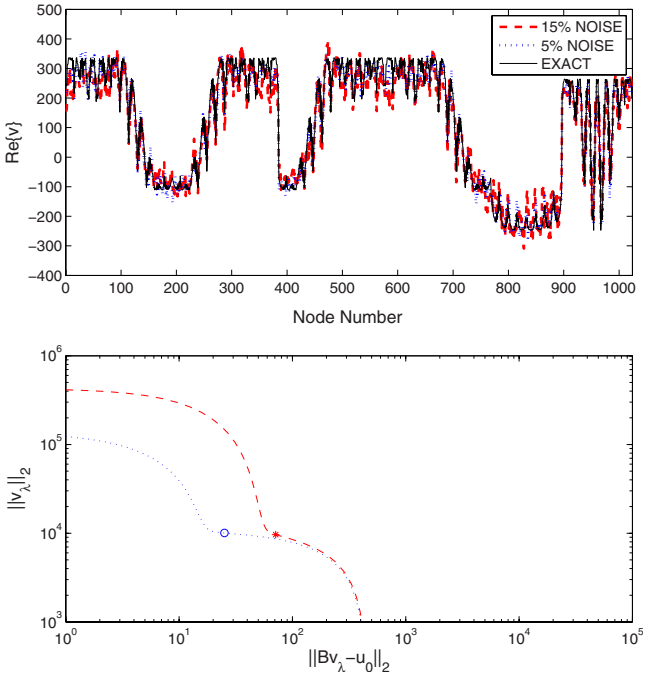


FIG. 7. (Color online) The real part of the reconstructed Neumann boundary data for the cuboid compared against the exact values (upper) and the position of the minimum error solution on the L-curve (lower) using the Burton–Miller formulation with $k=25$.

V. CONCLUSIONS

A Burton–Miller integral equation based IBEM was presented for application to NAH. Suitable methods for the integration and regularization were discussed. The method was able to accurately reconstruct the Neumann boundary data from pressure values on a conformal surface in the near-field of a radiating sphere and cuboid. In the former case, the characteristic wave-numbers are known analytically, and the deficiencies of the standard Helmholtz integral formulation were highlighted. The Burton–Miller formulation was shown to be more robust and reliably produced good reconstructions when combined with Tikhonov regularization and the L-curve parameter selection method. For the cuboid, this combination of methods was shown to produce good reconstructions over a range of wave-numbers. Future work incorporating the inverse Burton–Miller formulation described here may include the experimental verification of the method, for example, applied to acoustic radiation from loudspeakers, and the development of hybrid methods such as in Ref. 5.

ACKNOWLEDGMENT

The work carried out in this paper was partially funded by B&W Group Limited, Worthing, West Sussex, UK.

APPENDIX: DERIVATION OF THE SINGULARITY SUBTRACTION SPLITTING FOR THE HYPERSINGULAR INTEGRAL

In this appendix, the derivation of Eq. (17) is detailed. Taking the derivative inside the integration in the definition of $(\partial/\partial\mathbf{n}_x)Ku(\mathbf{x})$ and applying various differentiation rules yields the Hadamard finite part integral

$$\frac{\partial}{\partial\mathbf{n}_x}Ku(\mathbf{x}) = \oint\!\!\!\!\!\!=_\Gamma \left\{ (\mathbf{n}_x\cdot\mathbf{n}_y)\frac{(1-ik|\mathbf{x}-\mathbf{y}|)e^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|^3} + (\mathbf{n}_x\cdot\nabla_x|\mathbf{x}-\mathbf{y}|)(\mathbf{n}_y\cdot\nabla_y|\mathbf{x}-\mathbf{y}|) \times \frac{(3-3ik|\mathbf{x}-\mathbf{y}|-k^2|\mathbf{x}-\mathbf{y}|^2)e^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|^3} \right\} u(\mathbf{y})d\Gamma_y. \tag{A1}$$

Now one applies a singularity subtraction to each of the two terms in the integrand in Eq. (A1). Writing out the Taylor expansion of $e^{ik|\mathbf{x}-\mathbf{y}|}$ about 0 gives that

$$e^{ik|\mathbf{x}-\mathbf{y}|} = 1 + ik|\mathbf{x}-\mathbf{y}| - k^2|\mathbf{x}-\mathbf{y}|^2/2 + O(|\mathbf{x}-\mathbf{y}|^3).$$

Hence

$$(1-ik|\mathbf{x}-\mathbf{y}|)e^{ik|\mathbf{x}-\mathbf{y}|} - (1+ik|\mathbf{x}-\mathbf{y}| - k^2|\mathbf{x}-\mathbf{y}|^2/2) + ik|\mathbf{x}-\mathbf{y}|(1+ik|\mathbf{x}-\mathbf{y}|) = (1-ik|\mathbf{x}-\mathbf{y}|)e^{ik|\mathbf{x}-\mathbf{y}|} - 1 - k^2|\mathbf{x}-\mathbf{y}|^2/2 \tag{A2}$$

is $O(|\mathbf{x}-\mathbf{y}|^3)$ as $|\mathbf{x}-\mathbf{y}|\to 0$, and

$$(3-3ik|\mathbf{x}-\mathbf{y}| - k^2|\mathbf{x}-\mathbf{y}|^2)e^{ik|\mathbf{x}-\mathbf{y}|} - 3 \tag{A3}$$

is $O(|\mathbf{x}-\mathbf{y}|)$. Thus applying the subtractions (A2) and (A3) to the first and second terms in the integrand in Eq. (A1), respectively, yields a continuous integrand. Note that

$(\mathbf{n}\cdot\nabla|\mathbf{x}-\mathbf{y}|)$ is $O(|\mathbf{x}-\mathbf{y}|)$, where the surface normal and the differentiation can be at either $\mathbf{x}$ or $\mathbf{y}$ provided Γ is locally C^2 there (see, for example, Ref. 12, Lemma 6.15). The expression obtained is precisely the first integral in Eq. (17). The remainder of Eq. (17) is obtained by simply adding back the subtracted terms in new integrals in order to obtain $(\partial/\partial\mathbf{n}_x)Ku(\mathbf{x})$.

[1]E. G. Williams and J. D. Maynard, "Holographic imaging without the wavelength resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[2]E. G. Williams, J. D. Maynard, and E. Skudrzky, "Sound source reconstructions using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).
[3]B. K. Gardner and R. J. Bernhard, "A noise source identification technique using an inverse Helmholtz integral equation method," ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, 84–90 (1988).
[4]S. F. Wu and X. Zhao, "Combined Helmholtz equation-least squares method for reconstructing acoustic radiation from arbitrary shaped objects," J. Acoust. Soc. Am. **112**, 179–188 (2002).
[5]S. F. Wu, "Hybrid near-field acoustic holography," J. Acoust. Soc. Am. **115**, 207–217 (2004).
[6]A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[7]N. Valdivia and E. G. Williams, "Implicit methods of solution to integral formulations in boundary element based nearfield acoustic holography," J. Acoust. Soc. Am. **116**, 1559–1572 (2004).
[8]D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory* (Wiley, New York, 1983).
[9]D. S. Jones, "Integral equations for the exterior acoustic problem," Q. J. Mech. Appl. Math. **27**, 129–142 (1976).
[10]H. A. Schenk, "Improved integral equation formulations for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[11]S. Amini, "On the choice of coupling parameter in boundary integral formulations of the exterior acoustic problem," Appl. Anal. **35**, 75–92 (1989).
[12]R. Kress, *Linear Integral Equations*, 2nd ed. (Springer, New York, 1998).
[13]W. L. Meyer, W. A. Bell, and B. T. Zinn, "Boundary integral solutions of three dimensional acoustic radiation problems," J. Sound Vib. **59**, 245–262 (1978).
[14]A. Osterov and M. Ochmann, "A fast and stable numerical solution for acoustic boundary element equations combined with the Burton and Miller method for models consisting of constant elements," J. Comput. Acoust. **13**, 1–20 (2005).
[15]P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," SIAM Rev. **34**, 561–580 (1992).
[16]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[17]H. E. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Kluwer Academic, Dordrecht, 2000).
[18]M. Hanke, "Limitations of the L-curve method in ill-posed problems," BIT Num. Math. **36**, 287–301 (1996).
[19]C. R. Vogel, "Non-convergence of the L-curve regularization parameter selection method," Inverse Probl. **12**, 535–547 (1996).
[20]S. Amini, P. J. Harris, and D. T. Wilton, *Coupled Boundary and Finite Element Methods for the Solution of the Dynamic Fluid-Structure Interaction Problem*, Lecture Notes in Engineering Vol. **77** (Springer-Verlag, Berlin, 1992).

Reconstruction of sound source pressures in an enclosure using the phased beam tracing method

Cheol-Ho Jeong
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Jeong-Guon Ih[a)]
Department of Mechanical Engineering, Center for Noise and Vibration Control (NoViC), KAIST, Daejeon 305-701, Korea

(Received 15 September 2008; revised 15 April 2009; accepted 15 April 2009)

Source identification in an enclosure is not an easy task due to complicated wave interference and wall reflections, in particular, at mid-high frequencies. In this study, a phased beam tracing method was applied to the reconstruction of source pressures inside an enclosure at medium frequencies. First, surfaces of an extended source are divided into reasonably small segments. From each source segment, one beam is projected into the field and all emitted beams are traced. Radiated beams from the source reach array sensors after traveling various paths including the wall reflections. Collecting all the pressure histories at the field points, source-observer relations can be constructed in a matrix-vector form for each frequency. By multiplying the measured field data with the pseudo-inverse of the calculated transfer function, one obtains the distribution of source pressure. An omni-directional sphere and a cubic source in a rectangular enclosure were taken as examples in the simulation tests. A reconstruction error was investigated by Monte Carlo simulation in terms of field point locations. When the source information was reconstructed by the present method, it was shown that the sound power of the source in an enclosure could be estimated. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3132502]

PACS number(s): 43.40.Sk, 43.60.Jn, 43.55.Ka [NX] Pages: 158–166

I. INTRODUCTION

Acquiring information of acoustic parameters on a source surface such as normal velocity, acoustic pressure, impedance, intensity, and radiated sound power is very important in many acoustic problems. For example, for effective noise control, the identification of dominant vibro-acoustic sources takes the top priority in the early stage.

Many techniques have been developed for the accurate identification and ranking of the acoustic sources. The coherence method uses the statistical signal processing technique, in which the relation between the measured data at source and receiver positions is analyzed to identify the linear dependency between them.[1] This method is effective when the accurate numerical modeling of acoustic sources and transmission paths is difficult due to the complexity in boundary geometries and properties. In the sound intensity method, an acoustic energy flow in the sound field, usually at the near field of a radiator, is measured to identify the source and sink distribution around the source surface.[2] However, the measured intensity is indeed a field property, not a source property. Therefore, strong normal active intensity data near a radiating surface do not actually and necessarily mean that the nearest radiator surface portion corresponds to a strong contributor. Practicing this method requires substantial time and effort to scan the sound field, in particular, in the measurement for a large radiator or a fine spatial resolution. If tri-directional intensity data are measured at field points with a proper spatial resolution, reconstruction and visualization of the source might be possible. Owing to the recent progress, the near-field acoustic holography (NAH) becomes the most popular techniques in the source identification. Spatial Fourier transform of the spatial field data measured in the near-field to the wave number domain has been generally taken as a method to describe the sound propagation from a radiator to a regular-shaped hologram plane.[3] By virtue of the rapid development of the computational capability, the boundary element method (BEM) to find out the vibro-acoustic transfer function between the source surface and the field point has been widely used in the NAH.[4,5] The beam forming is also an array-based technique for the sound source localization at the far field.[6] This method does not require an array to be larger than the sound source. Furthermore, it can use an irregular array whereas the NAH usually requires a regular and rectangular sensor grid in the measurement. However, detection of coherent sources in the similar bearing angle is problematic and the precise estimation of the source parameters is somewhat tricky. A discrete Huygens model in rooms was suggested for source identification based on the time reversal process.[7] Pressures at receiver locations go back to a source point where sound is emanated, by multiplying inversed transmission line matrices.

For a reconstruction of the source information at medium frequencies, the phased beam tracing method (PBTM)[8,9] is adequate to calculate acoustic transfer func-

[a)]Author to whom correspondence should be addressed. Electronic mail: j.g.Ih@kaist.ac.kr

 0001-4966/2009/126(1)/158/9/$25.00

tions between source segments and receiver points. Once transfer functions are calculated by the PBTM, it is conceptually possible to recover a pressure distribution over the source surfaces by multiplying the measured field pressure data with the inverse transfer function. Similar problems with the NAH technique would arise in the matrix inversion process and the positioning of sensors. Numerical techniques for improving the singularity of the system matrix can be adopted to enhance the resolution of the reconstructed field.

This study aims at reconstructing the acoustic property of source surfaces in a room without moving it to an anechoic chamber. Theoretically, various holographic methods can be used for the reconstruction of the acoustic parameters on the source surface, but the modeling of all boundary surfaces of the room with precision is very difficult, in particular, at medium to high frequencies for large rooms. In this regard, it is thought that for a noise source in an enclosed room the PBTM can be a good alternative to inversely calculate surface pressures on sound sources. This paper mainly addresses the fundamental formulation, the numerical stability of applications, and the effects of field point locations.

II. PHASE GEOMETRICAL ACOUSTICS

First of all, the phased geometrical acoustics method is a modified geometrical acoustic technique for simulating sound propagation at medium to high frequency range by the aid of phase information. If the frequency of interest is far beyond the Schroeder cutoff frequency f_c,[10] the phase information is not really necessary because of heavy modal overlaps, which implies that each modal character cannot be clearly distinguished. In this range, the statistical models or geometrical acoustics have been successfully adopted. Among many geometrical acoustics techniques, image source methods can account for arbitrary geometry robustly, but it is limited to specular reflection.[11,12] Conventional ray-tracing methods model specular or diffuse reflections and arbitrary room geometry efficiently, but a volume receiver should be employed.[13,14] Due to the use of a volume receiver, responses of ray tracing are smeared temporally and spatially. Also discrete sampling of rays leads to undersampling errors, so enormous rays are needed to avoid sampling errors.[15] Beam tracing methods can analyze room acoustics in a similar way as the ray tracing with spatially extended beams and a point receiver, but problems occurs when beams intersects more than one surface.[8,9,16–19] If an intersecting polygon is detected, there are two solutions: The original beam is followed by its central axis ray[8,9,16] or the original beam can be split.[17,18] Splitting algorithms are robust and safe, but they become computationally voracious. However, conventional geometrical acoustics methods, which normally ignore phase and consider only energy quantities, cannot be used in the source identification at all.

At low frequencies, wave-based methods are the most reliable and appropriate tools in calculating transfer functions; thus the NAH based on wave-based methods is the core technique for reconstructing source information accurately. However, at around the Schroeder cutoff frequency f_c, both wave-based methods and high frequency methods cannot tackle acoustic problems appropriately: Wave-based methods require a lot of computational expenditure due to huge number of elements, whereas lack of phase information and modal characteristics leads high frequency methods to inaccurate outcomes. Therefore phased geometrical acoustics methods have been suggested. Inclusion of phase is twofold: phase at reflections from surfaces and propagation phase. Consequently outcomes of phased methods are sound pressures at receiver locations in time or frequency domain. Consequently the methods have been mainly used for calculating an impulse response or an acoustic transfer function for a source-receiver pair in simple rooms at medium frequencies. Initially phase was introduced into a ray-tracing model so that it could be applied to lower frequencies.[20] Suh and Nelson[21] analyzed several rectangular rooms and obtained satisfactory results at early reflections using the phased image source method. Jeong *et al.*[9] applied the phased beam tracing to predict impulse responses and acoustic parameters in a room. An improvement was found at medium frequencies in comparison with the conventional methods.

Wareing and Hodgson[8] developed a transfer-matrix model integrated into a beam tracing method for multilayered surfaces. An adaptive beam tracing method was tested in Bell laboratories by Tsingos *et al.*,[22] yielding a remarkable agreement with measurements. In their simulation, they incorporated the uniform theory of diffraction[23] with the beam tracing for invisible source-receiver pairs.

In this study, the triangular beam tracing approach proposed by Lewers[16] was extended to include phase. The beam tracing algorithm consists of source generation, surface-geometry definition, traces of beams, and receiver detection. Source division is based on an icosahedron, which makes the beam cross section an equilateral triangle. One edge of an equilateral triangle can be divided into p equal lengths, resulting in a polygon with $20p^2$ faces. Room surfaces should be planar, which are mathematically modeled as $A_i x + B_i y + C_i z + D_i = 0$. A trajectory of a beam is scanned by combined processes of determining the nearest plane, finding the new image source, calculating the reflected vector. Consider a beam, which is defined by a central axis and three boundary planes, each plane forming a side of the beam. Beams do not fragment on reflection and the direction after reflection is determined entirely by its central axis. Once the trajectory of the beam is identified, the possibility that a point receiver is surrounded by the beam boundary planes is tested using the normal vectors of the boundary walls. Following a positive receiver point test, the complex pressure amplitude for the beam is calculated and finally the transfer function is constructed.

Most published works have been tested in quite simple rooms, because of the inherent limitation of the geometrical acoustics methods. The most challenging task is to deal with wave phenomena, especially diffraction and diffuse reflection. Therefore, a room shape should be simple and wall surfaces should be relatively smooth and large. Highly faceted surfaces should be avoided and coarse models are adequate for the present method. If roughness of surfaces or dimensions of obstacles in rooms is much smaller than the

wavelength of interest, these can be neglected in the modeling procedure. Scattering of smooth surfaces of the test room is ignored.

As mentioned earlier, diffraction is the most challenging topic in geometrical acoustics and it has been argued that phased geometrical acoustics methods without proper diffraction algorithms cannot correctly predict the room acoustics of enclosed rooms. According to Pierce,[24] "amplitudes of the diffracted field are usually much weaker than direct and even reflected contributions." Diffraction was only considered in shadow regions in several previous works, assuming its contribution is relatively small in illuminated regions where direct and reflected contributions from a source also reach a listener.[22,25] On the other hand, Torres *et al.*[26] claimed that diffraction can be perceived in illuminated regions. It is still debatable if the diffraction should be included for entire enclosed sound field. According to the authors' experience, the present beam tracing algorithm neglecting diffraction can satisfactorily predict acoustic parameters and transfer functions in rooms with simple geometry.[9,27] It is true that if receivers in rooms are distant from diffracting obstacles, edges, and corners, geometrical acoustics components are dominant compared to diffracted components. Actually finding diffraction paths is a very elaborate task and, moreover, diffraction of finite, non-rigid wedges still needs to be studied further. In this regard, diffraction has not been taken into account for observation points far from diffracting objects in simple rooms. However, the authors believe that incorporation of recently developed diffraction algorithms[26,28,29] into phased geometrical acoustics models will perform better and more advanced phased geometrical acoustics methods will be successfully applied to source identification and reconstruction in the future.

III. INVERSE ALGORITHM FOR THE PHASED BEAM TRACING METHOD

The most important step in the source identification is the effective characterization of multiple transfer functions between a noise source and field points. A source surface radiating sound is modeled by appropriate boundary conditions, i.e., Neumann, Dirichlet, and mixed-type boundaries. Then, exactly speaking, the transfer function between source and field can be called a vibro-acoustic or a purely acoustic transfer function depending on the boundary condition type; however, a vibro-acoustic source is used in this paper assuming that the Neumann-type boundary, e.g., a hole, with a medium fluctuation can be also modeled as an equivalent vibrating source element like a moving piston as far as the size is small. Consider an extended vibro-acoustic source radiating sound waves into a large enclosure filled with homogeneous air medium. For modeling purposes, let the source surface, either actual or simplified one, be divided into small segments, called source segments. The resultant sound field in an enclosure is observed at many field points (or observation points). In order to obtain precisely restored results, over-determined field data are usually used for a uniformly distributed sensor points; that is, the number of field points N is larger than the number of source segments M. However, this over-determined condition is not the necessary condition for the inverse reconstruction. If one can somehow select a very independent set of field points, usually unevenly distributed sensor positions, even an underdetermined condition of source and field data may be acceptable with a reasonable precision.

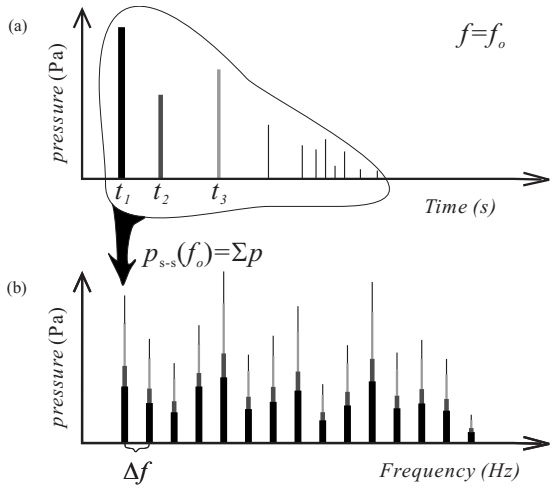


FIG. 1. Typical results by the phased beam tracing simulation. (a) Pressure reflectogram and (b) steady-state transfer function.

Using the PBTM, acoustic transfer functions between source segments and field points can be obtained. First, a pressure at a receiver by a beam is computed based on the information of the path length and surfaces that the beam hits, as

$$p\left(t=\frac{a_{\text{tot}}}{c_o};\theta_i,\omega\right)=\frac{p_{s,m}}{a_{\text{tot}}}e^{-j\tilde{k}a_{\text{tot}}}\prod_{i=1}^{q}r_i(\theta_i), \tag{1}$$

where $p_{s,m}$ is the pressure amplitude at the mth source segment, a_{tot} is the total travel distance of the beam, $\tilde{k}$ is the complex wave number (or $\tilde{k}=k-j0.5\text{AF}$), k is the wave number in a lossless free-field, AF is the attenuation factor of the air, $r_i(\theta_i)$ is the pressure reflection coefficient of the ith wall reflection, θ_i is the angle of incidence of the beam to the ith wall, q is the total number of wall reflections until the beam reaches the receiver, and c_o is the speed of sound in air. It is noted that the pressure in Eq. (1) is the complex pressure at the receiver point for a single frequency, and for the single beam departing from the mth source segment. Figure 1(a) shows a pressure reflectogram (or an echogram) for a single frequency. For simplicity, absolute magnitude plots are shown in Fig. 1. The total steady-state acoustic pressure is calculated by summing the total pressures of all beams detected at the receiver point. In Fig. 1(b), a steady-state transfer function for a unit input (at the source) is shown by collecting all frequency components. At this point in calculation procedure, frequency range and frequency resolution should be determined. The bottom parts of the pressure magnitude bars, which are the thickest, denote the contributions by the direct sound. The next bottom parts show the contributions from the reflected pressures at time t_2, and the further next parts show the contributions from the reflected pressures at time t_3, and so on.

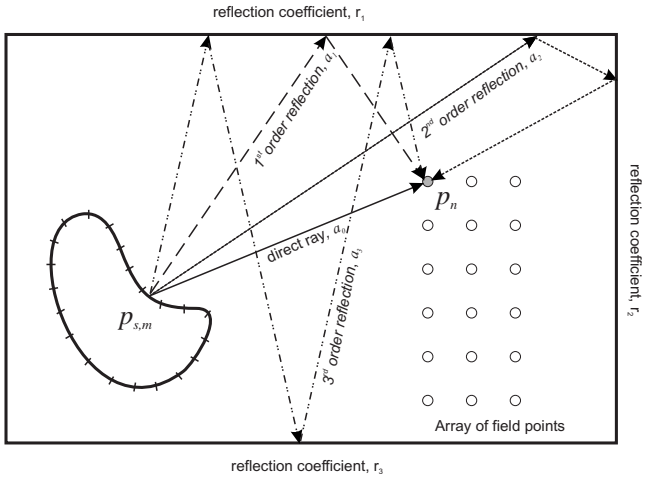


FIG. 2. A schematic two-dimensional presentation of the sound propagation within an enclosure. In the left, an extended source radiates the sound, of which the surface is comprised of M segmental areas. The surface pressure at the mth segment is expressed as $p_{s,m}$. Then, an array of N sensors receives the direct and the reflected sounds. Here, a_i denotes the travel distance of the ith reflection and r_j represents the reflection coefficient at the jth wall.

Figure 2 illustrates a conceptual presentation of the sound radiation and propagation within a two-dimensional enclosure. An extended source, which is comprised of M segmental surface areas, radiates sound. The surface pressure at the mth segment is expressed as $p_{s,m}$. Then, an array of N sensors receives the direct and the reflected sounds and the waves are superposed in time. In this figure, a_i denotes the travel distance of the ith reflection and r_j represents the reflection coefficient at the jth wall. Therefore, in notation, the zeroth reflection means the direct sound. In the actual situation, of course, three-dimensional geometrical modeling and scanning should be carried out to simulate the sound propagation in an enclosure. The field pressure at the nth field point caused by the mth source segment, p_n, can be expressed as a sum of direct component and successive reflected components (up to qth order) as follows:

$$p_n(\omega) = p_{s,m}\frac{e^{-j(k+j0.5\mathrm{AF})a_o}}{a_o} + p_{s,m}\frac{e^{-j(k+j0.5\mathrm{AF})a_1}}{a_1}r_1(\theta) + \cdots + p_{s,m}\frac{e^{-j(k+j0.5\mathrm{AF})a_q}}{a_q}\prod_{i=1}^{q} r_i(\theta)$$
$$= \left[\frac{e^{-j(k+j0.5\mathrm{AF})a_o}}{a_o} + \frac{e^{-j(k+j0.5\mathrm{AF})a_1}}{a_1}r_1(\theta) + \cdots + \frac{e^{-j(k+j0.5\mathrm{AF})a_q}}{a_q}\prod_{i=1}^{q} r_i(\theta)\right]p_{s,m} = H_{nm}p_{s,m}. \quad (2)$$

When a sufficient number of reflections is counted, the steady-state transfer function, H_{nm}, between the mth source segment and the nth receiver position, can be computed with precision. Similar equations can be obtained for combining other source segments and receiver positions. Consequently, for all source segments and receiver positions, a transfer matrix can be written as follows:

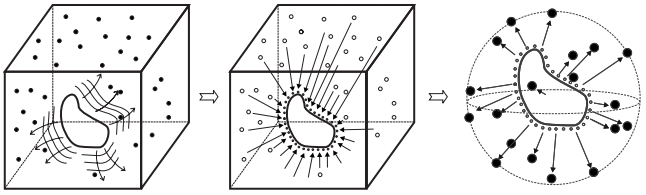


FIG. 3. An illustration for the sound power estimation. First, transfer functions are calculated and then backward reconstruction of the source pressure is carried out. With the estimated source pressure, the sound power of the source is calculated according to ISO 3745.

$$\begin{Bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{Bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1M} \\ H_{21} & H_{22} & \cdots & H_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NM} \end{bmatrix} \begin{Bmatrix} p_{s,1} \\ p_{s,2} \\ p_{s,3} \\ \vdots \\ p_{s,M} \end{Bmatrix}$$
$$\text{or} \quad \mathbf{P}_{f,N\times 1} = \mathbf{H}_{N\times M}\mathbf{P}_{s,M\times 1}. \quad (3)$$

From the measured field pressures of $\mathbf{P}_f$, the source pressures $\mathbf{P}_s$ can be inversely recovered from Eq. (3) as

$$\mathbf{P}_{f,N\times 1} = \mathbf{H}_{N\times M}\mathbf{P}_{s,M\times 1} \Rightarrow \mathbf{P}_{s,M\times 1} = \mathbf{H}^{-1}_{N\times M}\mathbf{P}_{f,N\times 1}. \quad (4)$$

Figure 3 briefly depicts the entire process of the study. First of all, an extended noise source with *a priori* known surface pressure distribution emits sound in a room. A transfer function matrix, $\boldsymbol{H}$, from source segments to receiver locations was calculated by the PBTM. As a result, noiseless field pressures, $\mathbf{P}_f$, are obtained by multiplying the transfer matrix with the known surface pressures according to Eq. (3). Measurement noise can be artificially added to the calculated $\mathbf{P}_f$. In a practical situation, field pressures are generally measured by a well-calibrated array technique. The second figure shows that surface pressures are reconstructed by the inversion of the calculated transfer matrix according to Eq. (4). This process is named *reconstruction of source data*. The reconstruction error must vanish in a noise-free condition, but it increases if the measurement noise is involved in. During this process, a regularization can be employed. For quantifying a reconstruction error, Monte Carlo simulations were conducted for pre-defined sets of field points. Once the surface pressures are reconstructed, the source data can be used to calculate the sound pressures at any receiver/field points. This process is called *regeneration of sound field*. The third figure shows a calculation of field pressures from the reconstructed source data as if it is situated in an anechoic chamber. The acoustic power of the source can be estimated by using the calculated field pressures.

IV. TEST EXAMPLE WITH OMNI-DIRECTIONAL SOURCES

For the simplest case, two omni-directional sources were chosen. The first one is an icosahedron source in Fig. 4(a), which consists of 20 equilateral triangles. The other is a cubic source having 24 isosceles triangles in Fig. 4(b). Figure 4(c) shows a rectangular room model and 96 evenly spaced

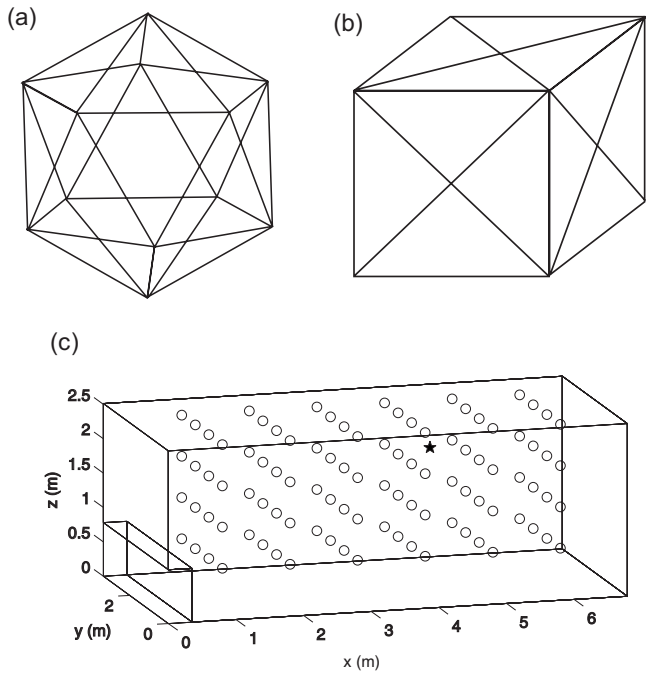


FIG. 4. Room and source model. (a) Icosahedron source, (b) cubic source, and (c) parallelepiped room. Locations of 96 candidates field points are shown as ○, whereas the symbol ★ denotes the source location.

field points in the target space. The size of the room is 6.8 $\times 3.4 \times 2.5$ m^3. The omni-directional source is located at 4.3, 1.5, and 2 m.

Wall boundary conditions were given by absorption coefficients shown in Table I, and the angle-dependent reflection coefficients[9,30] were calculated considering the size of the walls. The maximum number of reflection was limited to 50. Based on the measured reverberation time of 1.2 s at medium frequencies, the Schroeder cutoff frequency,[10] f_c, was about 300 Hz. The common definition of medium frequency is from f_c to $4f_c$.[31] This study also focuses only on the medium frequencies due to the basic premise of the PBTM, although the method can be extendedly applicable to the outside of the medium frequency range with small inevitable errors. The selected medium frequency for demonstration was 900 Hz, which corresponds to about $3f_c$.

A. Effect of field points and regularization with icosahedron source

Twenty beams emanate from the icosahedron source, which means basically 1 beam per source segment was emitted and traced. In order to investigate the effect of field point locations, three sets of field points were chosen. As the first set, only 20 points among 96 candidate field points were chosen by excluding the farthest 76 points from the source, in Fig. 5(a). On the contrary, the 76 closest sensor positions to the source are eliminated in the second set as can be seen in Fig. 5(b). The third set consists of 20 sensor positions determined by the effective independent (EfI) technique in Fig. 5(c).

TABLE I. Absorption coefficients of the room surfaces.

Surface (material)	Absorption coefficient
Floor (stone)	0.02
Ceiling (gypsum)	0.04
Wall (concrete)	0.02
Window (glass)	0.04
Ventilation grating (partially open)	0.60
Ventilation cover (thin steel)	0.05

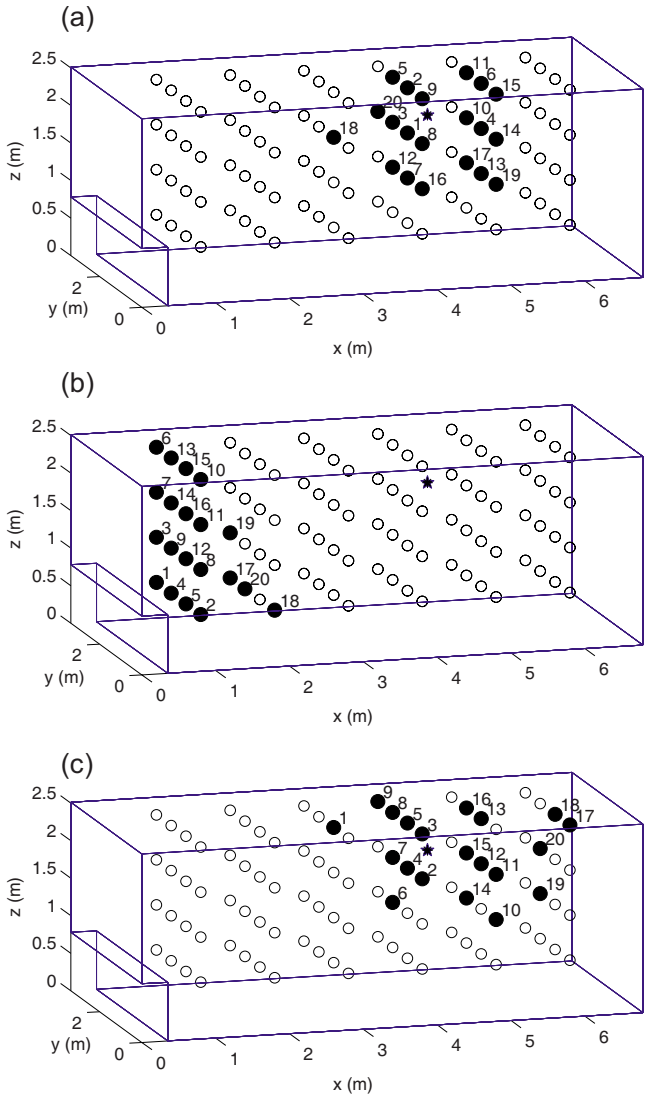


FIG. 5. (Color online) Locations of 20 field points. (a) Field point set 1, (b) field point set 2, and (c) field point set 3.

In the real situation, the backward reconstruction of the source field suffers from divergence phenomenon during the inverse of ill-conditioned matrix. The major cause of such an additional ill-conditioning, besides the effect of measurement noise, is originated from the redundancy of field points, which is reflected into the transfer matrix having dependency in between columns and rows. To assure the independence among field positions, a sensor positioning method, so called EfI method, can be employed in the initial setting of the field points. The method is based on the mathematical strategy of ranking the contribution of each candidate sensor location to the rank of the system matrix. This technique was already successfully applied to many inverse problems: for the wave-based identification of a large space structure[32,33] and the

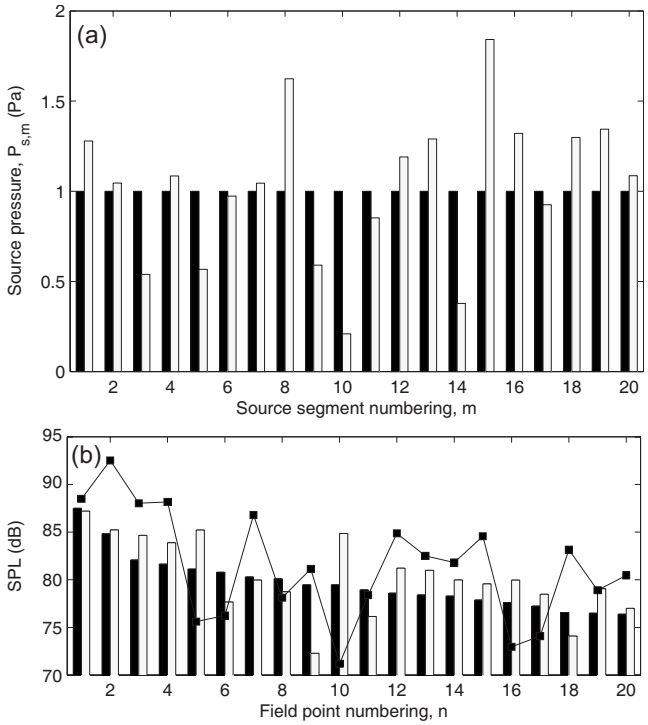


FIG. 6. Recovered source data and regenerated free-field data at set 1: ■, noise-free condition; □, condition with SNR=20 dB; and —■—, SPL in the presence of enclosure walls.

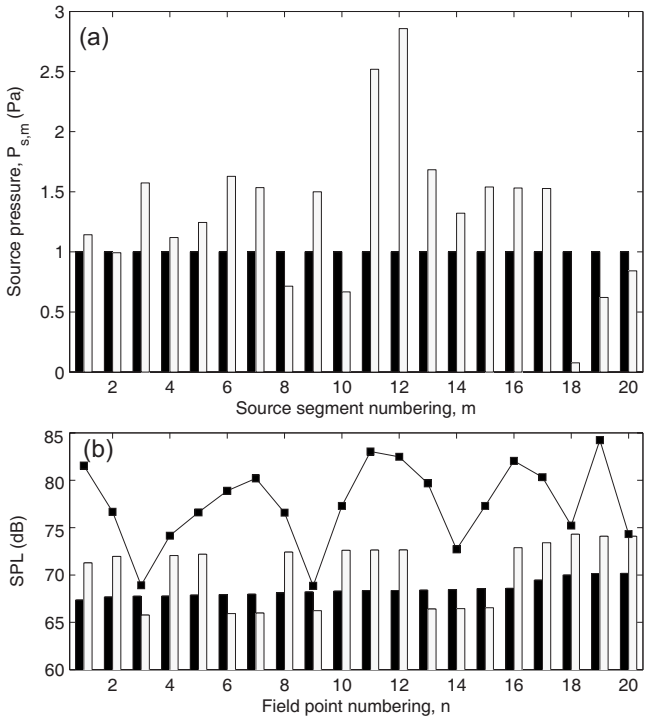


FIG. 7. Recovered source data and regenerated free-field data at set 2: ■, noise-free condition; □, condition with SNR=20 dB; and —■—, SPL in the presence of enclosure walls.

BEM-based NAH.[34,35] First, an initial candidate set of sensor location is selected, which covers the whole acoustic domain in a very fine manner. Then, the transfer matrix $\boldsymbol{H}$ in Eq. (3) is analyzed by the singular decomposition technique, which will express the matrix $\boldsymbol{H}$ as a multiplication of two unitary singular vectors, $\boldsymbol{U}$ and $\boldsymbol{V}$, and a diagonal singular matrix, $\boldsymbol{\Lambda}$. An EfI value is defined by a diagonal value of the multiplication of the left singular vector, $\boldsymbol{U}$, and its Hermitian form. Taking a point or several points as a group, one should calculate the EfI value of that corresponding point and the calculation of EfI value is carried out for all points. Because the EfI concept is to identify the contribution of each sensor position to the linear independence of the whole matrix, a small EfI value, in a relative sense, means that the corresponding sensor point depends a lot on the other points. Consequently, locations having smallest EfI value of all should be discarded from the initial population of the sensor positions. The same process continues with the remaining candidate set of sensor positions until a predetermined number of sensors is reached. In this way, the singularity factor of the transfer matrix can be reduced significantly before applying any regularization method. Hereafter, the three sets of field points are called "set 1," "set 2," and "set 3," respectively.

All the segmental source pressures were given by 1 Pa. The measurement noise, from environment and measurement system itself, will be always included in practice. Because a small noise in the field data will be amplified during the inverse process, the effect of noise on the reconstructed result cannot be overlooked. Field pressures were contaminated by random noises having the mean signal-to-noise ratio (SNR) of 20 dB, which follows a normal distribution. For set 1, the reconstructed source pressure distribution on the surface is shown in Fig. 6(a). One can find that the reconstructed data recovered from the noise-contaminated field pressure data show a large difference from the original source pressure distribution: An estimated maximum error was 90% at 15th source segment. To suppress the effect of measurement noise, thus enhancing the recovered source image, the regularization should be applied to the raw reconstructed data.

One can regenerate a sound field from the reconstructed source pressure. However, in regenerating pressures at the same field points in the same room from the "incorrect" recovered source data, a summation of the direct sound and successive reflections will end up with field pressures that are different from the original noise-contaminated field data. Therefore, in this back and forward calculation process, effect of room geometry is counted twice, viz., the reconstruction and regeneration. To avoid the double count of the wall reflection effect, the room surfaces are regarded invisible, in other words, as an anechoic condition in the regeneration. The calculated sound pressure levels at the initial field points are displayed in Fig. 6(b). In this figure, the field data without any measurement noise can be considered as a "true" data set. One can observe that the maximum difference in sound pressure levels is less than 5 dB. The result shows that the overall distribution of the regenerated field data is similar to the initial data set. The line with square symbol denotes the calculated sound pressure level (SPL) when the wall reflections are taken into account.

The reconstructed source pressures from the field data at set 2 are shown in Fig. 7(a), which differ much from the original values. The regenerated sound pressure levels at field points set 2 are also different from those of noise-free condition, as can be seen in Fig. 7(b). SPL differences are

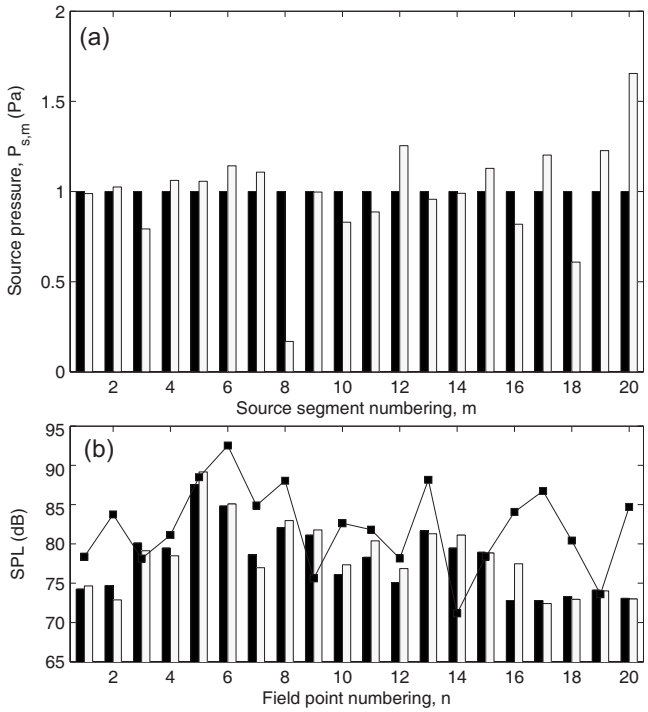


FIG. 8. Recovered source data and regenerated free-field data at set 3: ■, noise-free condition; □, condition with SNR=20 dB; and —■—, SPL in the presence of enclosure walls.

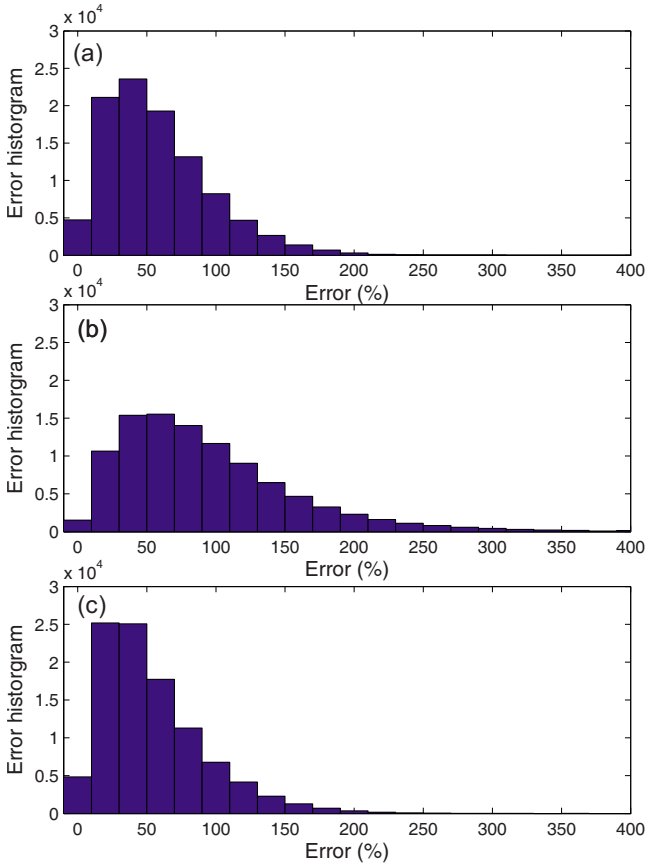


FIG. 9. (Color online) Error histograms of the reconstructed source pressures without regularization. (a) Field point set 1, (b) field point set 2, and (c) field point set 3.

magnified to 4 dB on average. This result emphasizes the importance of choosing proper locations of field sensor points. When the field points are far from the source, there is a tendency that the rank-deficient transfer matrix degrades the accuracy of the inverse problem, notwithstanding the fact that the far points from the source cannot always be regarded as bad points for the reconstruction.

Using the field pressure data measured at set 3, the reconstructed source pressure distribution and the regenerated sound pressure levels at the field positions are depicted in Fig. 8. Apparently, the errors are much smaller than the other two cases using different field data sets.

To compare the errors, the Monte Carlo simulation was conducted for three field point sets with 5000 trials. A percentage error is defined as the ratio of difference in the surface pressure to the original surface pressure as

$$\text{error} = \frac{1}{M}\sum_{m=1}^{M}\left|\frac{p_{s,m_{\text{noise}}} - p_{s,m_{\text{original}}}}{p_{s,m_{\text{original}}}}\right| \times 100(\%). \tag{5}$$

In Fig. 9, error histograms clearly show the importance of choosing appropriate field point locations. The average error for field points in set 3 amounts to 54%, while the average errors by set 1 and set 2 correspond to 58% and 92%, respectively.

In addition to the proper sensor placement technique, a regularization method was finally applied to overcome the instability of the inverse problem. The instability of the inverse reconstruction usually occurs due to the presence of small measurement noise and high order wave components, which are significantly amplified during the inversion. Due to this reason, the transfer matrix should be modified by adopting an appropriate wave number filter. Among various regularization methods, the modified Tikhonov regularization method[36,37] was applied to suppress the excessive effect of measurement noise. Regularization parameters were optimally determined by minimizing the generalized cross-validation function.[38]

After applying the regularization method to the inverse process, a significant enhancement of the source image could be obtained as can be seen in Fig. 10. For poorly chosen field points, set 2, the average error was reduced to 38% as shown in Fig. 10(b). This is a dramatic improvement compared with the original average error of 92%. The regularized results for field point set 1 and set 2 are similar in error distribution. The average error is of 37% when the field data set 1 is adopted, whereas the field data set 3 yield the lowest average error of 24%.

B. Cubic source

So far, a spherical source has been adopted in the PBTM simulation. Because most machinery is shaped in the parallelepiped, the simulation using such a parallelepiped source would be meaningful for the practical applications of the present method. The cubic source in Fig. 4(b) was located at the same source location in the same enclosure in Fig. 4(c). In the simulation, the emitted number of beams was 24, and the number of reflections was limited to 50. The absorption

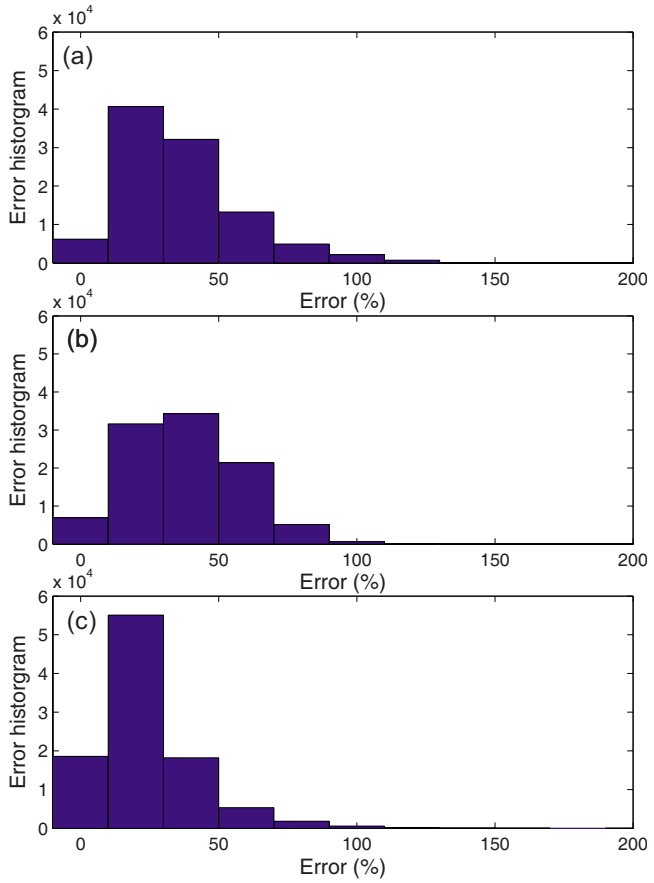


FIG. 10. (Color online) Error histograms of the reconstructed source pressures with the modified Tikhonov regularization method. (a) Field point set 1, (b) field point set 2, and (c) field point set 3.

coefficients in Table I were used in the simulation. Figure 11 shows the regenerated sound pressure data, which yield a maximum error of 3 dB.

C. Estimation of sound power level

Additional purpose of the present method was to estimate the sound power radiated from the source. When the field points were spherically distributed at a radius of 1 m from the source in an anechoic condition (see the rightmost part of Fig. 3), the sound power level could be estimated in accordance with ISO 3745.[39] Sound pressures measured at predetermined 20 positions (Annex C of ISO 3745) were regenerated and the sound power level of the source was estimated as follows:

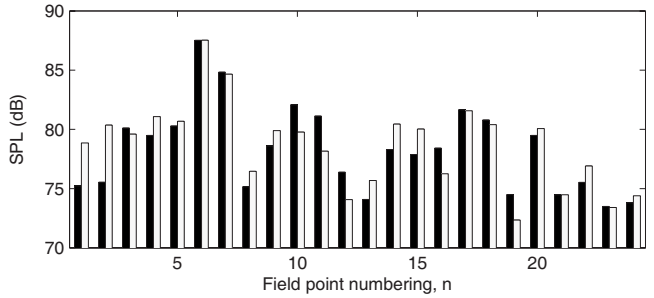


FIG. 11. Regenerated free-field data employing the cubic source: ■, noise-free condition; □, condition with SNR=20 dB.

$$L_w = \overline{L_p} + 10\log_{10}(4\pi). \quad (6)$$

Here, $\overline{L_p}$ is the average sound pressure level over the 20 microphone locations, which is equivalent to the sound intensity level in an anechoic condition. In practice, however, one cannot precisely measure the free-field pressures without moving the source into an anechoic chamber.

It has been observed that the field pressures in the enclosure are generally higher than those in the free-field in Figs. 6(b), 7(b), and 8(b), mainly due to the interference of reflected waves from the walls. Because of these amplified field pressures in the enclosure, the sound power level of the source would be overestimated. For example, the estimated sound power of the source using the amplified field pressures was 94.3 dB, whereas the actual sound power was 91.0 dB. Using the regenerated free-field sound pressures by the PBTM, the average sound power level over 500 random trials was obtained as 91.4 dB with SNR of 20 dB. This simulation shows a possibility to adopt the present method in estimating the sound power radiated from heavy and big noise sources located in factories or enclosures.

V. CONCLUSIONS

A method to inversely reconstruct sound source data within an enclosure was studied, which would be valid at medium frequencies. Using the phased beam tracing method, one could easily take the room effect into account in obtaining the transfer functions between the source segments and the field points. By solving the resultant inverse problem, it was shown that the source pressure could be recovered. EfI and a regularization technique were employed to overcome the inherent problem in the inversion process, thus enhancing the resultant source image. It is thought that the proposed technique for the source identification using the PBTM can be effectively applied to the estimation of the sound power level of noise source without moving a source to an ideal chamber. The present method will be experimentally validated and rather complicated room geometries and nonidealized pressure distributions of actual sources will be further investigated. Lastly, but not least, the controversial issue about adequacy of neglecting diffraction will be discussed in the future.

ACKNOWLEDGMENT

This work was partially supported by the BK21 Project and the NRL.

[1]J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis* (Wiley, New York, 1980).
[2]F. J. Fahy, *Sound Intensity* (E & FN Spon, London, 1976).
[3]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[4]B.-K. Kim and J.-G. Ih, "On the reconstruction of vibro-acoustic field over the surface enclosing an interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[5]M. R. Bai, "Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries," J. Acoust. Soc. Am. **92**, 533–548 (1992).
[6]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[7]Y. Kagawa, T. Tsuchiya, K. Fujioka, and M. Takeuchi, "Discrete Huygens' model approach to sound wave propagation—Reverberation in a room, sound source identification and tomography in time reversal," J. Sound Vib. **225**, 61–78 (1999).
[8]A. Wareing and M. Hodgson, "Beam-tracing model for predicting sound field in rooms with multilayer bounding surfaces," J. Acoust. Soc. Am. **118**, 2321–2331 (2005).
[9]C.-H. Jeong, J.-G. Ih, and J. H. Rindel, "An approximate treatment of reflection coefficient in the phased beam tracing method for the simulation of enclosed sound fields at medium frequencies," Appl. Acoust. **69**, 601–613 (2008).
[10]M. R. Schroeder and H. Kuttruff, "On frequency response curves in rooms: Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima," J. Acoust. Soc. Am. **34**, 76–80 (1962).
[11]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).
[12]H. Lee and B.-H. Lee, "An efficient algorithm for the image model technique," Appl. Acoust. **24**, 87–115 (1988).
[13]A. Krokstad, S. Stroem, and S. Soersdal, "Calculating the acoustical room response by the use of a ray tracing technique," J. Sound Vib. **8**, 118–125 (1968).
[14]A. Kulowski, "Algorithmic representation of the ray tracing technique," Appl. Acoust. **18**, 449–469 (1985).
[15]A. Kulowski, "Error investigation for the ray tracing technique," Appl. Acoust. **15**, 263–274 (1982).
[16]T. Lewers, "A combined beam tracing and radiant exchange computer-model of room acoustics," Appl. Acoust. **38**, 161–178 (1993).
[17]T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," J. Acoust. Soc. Am. **115**, 739–756 (2004).
[18]I. A. Drumm and Y. W. Lam, "The adaptive beam-tracing algorithm," J. Acoust. Soc. Am. **107**, 1405–1412 (2000).
[19]D. van Maercke and J. Martin, "The prediction of echograms and impulse responses within the Epidaure software," Appl. Acoust. **38**, 93–114 (1993).
[20]E. De Geest and H. Patzold, "Comparison between room transmission functions calculated with a boundary element method and a ray tracing method including phase," Proceedings of Inter-Noise (1996), Vol. **96**, pp. 3177–3180.
[21]J. S. Suh and P. A. Nelson, "Measurement of transient response of rooms and comparison with geometrical acoustic models," J. Acoust. Soc. Am. **105**, 2304–2317 (1999).
[22]N. Tsingos, I. Carlbom, G. Elko, R. Kubli, and T. Funkhouser, "Validating acoustical simulations in the Bell Labs box," IEEE Comput. Graphics Appl. **22**, 28–37 (2002).
[23]R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," Proc. IEEE **62**, 1448–1461 (1974).
[24]A. D. Pierce, *Acoustics. An Introduction to Its Physical Principles and Applications* (American Institute of Physics, New York, 1991).
[25]N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom, "Modeling acoustics in virtual environments using the uniform theory of diffraction," Proceedings of ACM SIGGRAPH (2001), CD-ROM.
[26]R. R. Torres, U. P. Svensson, and M. Kleiner, "Computation of edge diffraction for more accurate room acoustics auralization," J. Acoust. Soc. Am. **109**, 600–610 (2001).
[27]C.-H. Jeong and J.-G. Ih, "Introduction and applications of phased beam tracing method: Can we interpret low frequency response by the particle property?," Proceeding of the International Congress on Acoustics (2007), Paper No. RBA-05-016.
[28]U. P. Svensson and R. I. Fred, "An analytic secondary source model of edge diffraction impulse responses," J. Acoust. Soc. Am. **106**, 2331–2344 (1999).
[29]V. Pulkki, U. P. Svensson, and T. Paatero, "Efficient representation of edge diffraction impulse responses," Proceedings of the International Congress on Acoustics (2007), Paper No. RBA-05-014.
[30]J. H. Rindel, "Modeling the angle-dependent pressure reflection factor," Appl. Acoust. **38**, 223–234 (1993).
[31]F. A. Everest, *Master Handbook of Acoustics* (McGraw-Hill, New York, 2001).
[32]D. C. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," J. Guid. Control Dyn. **14**, 251–259 (1991).
[33]D. C. Kammer, ""Effects of noise on sensor placement for on-orbit modal identification of large space structures," ASME J. Dyn. Syst., Meas., Control **114**, 436–443 (1992).
[34]B.-K. Kim and J.-G. Ih, "Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by near-field acoustical holography," J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[35]I.-Y. Jeon and J.-G. Ih, "On the holographic reconstruction of vibroacoustic fields using equivalent sources and inverse boundary element method," J. Acoust. Soc. Am. **118**, 3473–3482 (2005).
[36]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[37]G. H. Golub and C. F. V. Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, MD, 1996).
[38]P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).
[39]ISO 3745:2003, "Acoustics—Determination of sound power levels of noise sources using sound pressure—precision methods for anechoic and hemi-anechoic rooms" (International Organization for Standardization, Geneva, 2003).

Use of the standard rubber ball as an impact source with heavyweight concrete floors

Jin Yong Jeon,[a] Pyoung Jik Lee, and Shin-ichi Sato
Department of Architectural Engineering, Hanyang University, Seoul 133-791, Korea

(Received 21 February 2008; revised 12 May 2009; accepted 12 May 2009)

To select an appropriate standard floor impact source to simulate real floor impacts, objective and subjective evaluations of the floor impact sounds were conducted in a box-frame-type structure with reinforced concrete slab floors. The sounds simulated in the test were those that would result from an adult walking barefoot, children running and jumping (represented by a heavy-weight impact source, such as a bang machine or an impact ball), as well as those of a person walking in high-heels or a lightweight object being dropped (represented by a tapping machine). Similarity tests between human-made impact sounds and standard heavy-weight impact sounds were performed. Sound quality (SQ) metrics were used to predict the results of the similarity tests. These results showed that the impact sound of an impact ball is more similar to a human-made impact sound than the sound of a bang machine. A multiple regression analysis showed that loudness and roughness are significant factors describing the results of similarity judgment among SQ metrics. Much of the data from the standard impact sources, measured in reinforced concrete floors with rigid floor coverings, have been collected. An empirical relationship was established to convert the impact pressure sound level from the bang machine or tapping machine to that from the impact ball. This study indicates that the use of an impact ball is reliable for simulating human impact sounds.
 [DOI: 10.1121/1.3148193]

PACS number(s): 43.50.Ba, 43.50.Jh, 43.50.Pn [NX] Pages: 167–178

I. INTRODUCTION

Floor impact sounds, which are mainly caused by walking, running, and jumping, are regarded as the most annoying type of noise in apartment buildings. To simulate such human running and jumping, standard heavy-weight floor impact sources, such as the bang machine [tire, Japanese Industrial Standard JIS A 1418-2:2000 (Ref. 1) since 1978; Korean Standard KS F 2810-2 (Ref. 2) since 1981] and the more recently developed impact ball (JIS A 1418-2:2000; ISO 140-11:2005; ISO CD 10140-3:2007), have been used. The physical characteristics of the floor impact sounds were investigated to show the similarity of impact ball sounds to human impact sounds.[3]

A light-weight impact sound was first made to mimic high-heel tapping, or the dropping of a light-weight object, which produced floor impact sounds with relatively high frequencies. The tapping machine was then adopted by ISO as a standard impact source in 1960, and it has been widely used as a standard impact source. However, some studies indicated that the floor impact noise evaluation using a tapping machine did not properly reflect the characteristics of human-made floor impact noise.[4–6] Watters[7] reported that the floor impact spectrum of light-weight impact noise, generated by the tapping machine, was different from the spectrum of the high-heeled foot traffic of women. Warnock[8] emphasized the usefulness of the tapping machine, but pointed out that it did not produce enough low-frequency sounds to be compared with the bang machine data.[9] In case of the tapping machine, due to the round shape of the hammer, inconsistent impact results, which causes nonlinearity, when dropped on carpets or spongy surfaces.[10]

The impedance levels of modified tapping machine[10,11] were reported and its investigation led to the proposal of a modified tapping machine in the ISO.[12] Although this modified tapping machine reasonably simulates a walking human (300–500 N), it still cannot simulate the problematic impact sources such as running children (600–1000 N) and jumping children (2000–3000 N). In reinforced concrete buildings, the low-frequency floor impact sound pressure level generated by the modified tapping machine is lower than that of real running and jumping. Therefore, the use of spectrum adaptation term C_I to the floor impact noise level $L'_{n,AW}$,[13] which was suggested to compensate the low-frequency components in heavy-weight impact sounds, has been at issue. There are large variations in heavy-weight impact sound levels with different floor structures, as heavy-weight floor impact is impulsive. Hence, an impact sound should be evaluated by the maximum sound pressure levels, $L_{\max}$, not by the 6 s equivalent sound pressure levels, L_{eq}.

In Japan, heavy-weight floor impact sounds mainly caused by children's running and jumping became obvious as well as light-weight floor impact sounds. So in 1973, an experimental method for measuring floor impact noise (JIS A 1418) was established. Henceforth the measuring method for the application of heavy-weight impact sound was generated by a tire for Japanese residential buildings. For the past 30 years, in Japan and Korea, floor impact sounds have been evaluated using a bang machine; however, the impact force of the bang machine exceeds the range of real impact forces,

[a]Author to whom correspondence should be addressed. Electronic mail: jyjeon@hanyang.ac.kr

especially at low frequencies, and can damage structural components of wooden-framed houses. Therefore, a new standard impactor with a lower impact force was needed, which emerged as the impact ball.

In addition to lowering the impact force and reducing damage to structural components, the impact ball also better approximates real impacts. Tachibana *et al.*[14] reported that the characteristics of the impact ball were most similar to the frequency characteristics of real impact noise in different Japanese residential buildings. Jeon *et al.*[3] compared impact sounds made by humans, an impact ball, and a bang machine. The impedance, impact force, and the impact sound level of the impact ball sound were most similar to real human impact sound. The impact force exposure level of the bang machine was higher than that of the impact ball, below 63 Hz. However, above 125 Hz, the spectrum level of the bang machine was lower than that of the impact ball. Subsequently these physical characteristics of the floor impact sound need to be evaluated.

In most cases, the floor impact sound is evaluated according to sound pressure levels and frequency characteristics. Jeon *et al.*[15] investigated the loudness and annoyance of floor impact sounds for different sound insulation treatments in rooms of an apartment. Although a given floor impact sound level and its frequency characteristics are mainly related to loudness, psychoacoustical parameters other than floor impact sound levels may also affect the perception of the noise source. Jeon *et al.*[16] examined the relationship between sound quality (SQ) metrics and loudness of floor impact sounds through auditory experiments, with subjects exposed to low-frequency sound. According to Tachibana *et al.*[17] loudness is an exact measure of noise. Jeon and Sato[18] examined the relationship between SQ metrics and annoyance of heavy-weight floor impact sounds. Autocorrelation function (ACF) parameters have also been used to characterize and describe the relationship between objective and subjective evaluations of floor impact sounds.[16,18]

SQ metrics, originally proposed by Zwicker,[19] were defined in consideration of the listener's perception and evaluation of SQ. SQ metrics also reflect both frequency and temporal masking through the application of equal loudness contours. Loudness can be calculated from the loudness chart measured by the 1/3-octave band levels of a noise, which indicates a single index. High-frequency components of a sound make it sharp, and this sharpness increases annoyance. Roughness and fluctuation strength describe the fluctuation of a low-frequency signal, and the target modulation frequencies of roughness and fluctuation strength are typically around 70 and 4 Hz, respectively. Currently, only methods for calculating Zwicker loudness for stationary sound have been standardized,[20] thus Zwicker's model[19] has also been applied to non-stationary and impulsive sounds such as vehicle noises, golf impact sounds, and heavy-weight impact sounds.[16,18,21,22] Recent investigations by ISO/TC 43/SC 1/WG 9 are aimed at determining the parameters necessary for evaluating the psychoacoustical aspects of sound, especially the loudness of non-stationary sounds and the detection of pitch. With regard to other SQ metrics (roughness, sharpness, and fluctuation strength), the calculation procedure for stationary sounds was also used for non-stationary sounds with the original definitions proposed by Zwicker.[16,18,21,22] By using these metrics, the parameters that describe human perception and explain the similarity of the standard impact sources to the human impact sounds were investigated.

The purpose of the present study is to objectively and subjectively evaluate the similarity between human impact sounds and standard floor impact sound sources (impact ball, bang machine, and tapping machine). First, floor impact sounds generated by real impact sounds and the standard impact sources were measured in a test building. Then, the effects of SQ metrics on the similarity judgment of the heavy-weight floor impact sound were investigated in order to analyze the noise source in terms of human perception. Similarity tests between human-made impact sound and heavy-weight impact sound were conducted in a room of an apartment. The relationships between the SQ metrics and the similarity of the standard impact sources to the human impact sounds were examined by multiple regression analysis.

As a new standard impact source, studies have been conducted to compare the floor impact sound levels of an impact ball with those of the bang machine and the tapping machine. These studies aim to develop a method to convert the floor impact sound levels of each standard impact source into that of the other standard impact source. In our previous study,[3] the relationship between the bang machine and impact ball using an inverse *A*-weighted sound pressure level was investigated. Tanaka and Murakami[23] also investigated the relationship between the bang machine and impact ball using the single-number index (*L*-number according to JIS A 1419-2). However, conversion of a single-number rating of the bang machine into that of the impact ball has not been available. Therefore, a method for converting the floor impact sound level of the bang machine and the tapping machine to that of the impact ball was investigated, in order to use existing bang machine and tapping machine data for evaluating floor impact sound.

II. OBJECTIVE AND SUBJECTIVE EVALUATIONS OF FLOOR IMPACT SOUNDS

A. Objective evaluation in a test building

Floor impact sounds were measured in a test building, a model of a living room similar to an apartment unit in Korea. The test building used in this study employed a box-frame-type structural system to simulate a real apartment for practical testing purposes. The receiving room of the test building was rectangular in shape with a volume of 62.2 m^3. Figure 1 shows the details of floor treatments for two units. Figure 1(a) shows a floor type inserted with a resilient isolator, which is widely used in Korea, to mainly reduce light-weight floor impact sounds, and Fig. 1(b) shows the other floor without a resilient isolator. Wooden flooring was installed as a finishing material to reflect the usual condition of the floors in normal residential buildings.

The standard impact sources (impact ball, bang machine, and tapping machine), and the real impact sources were generated. Dropping of a light-weight object (0.5 ℓ plastic bottle

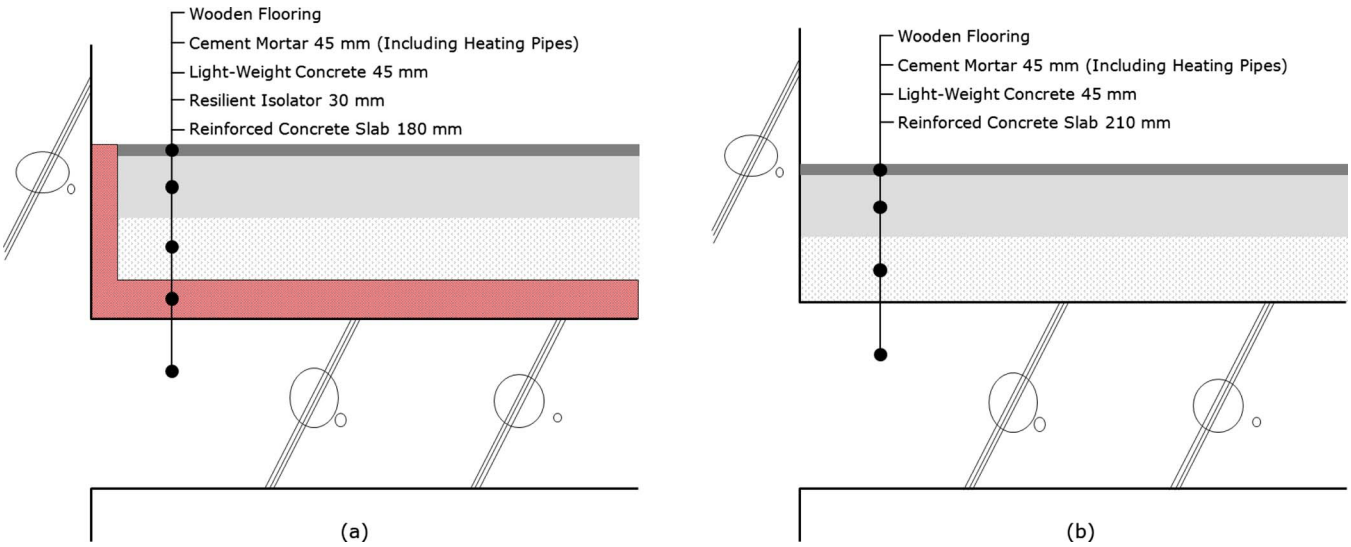


FIG. 1. (Color online) Sectional details of layer treatments for two unit floors: (a) 180-mm-thick floor and (b) 210-mm-thick floor.

full of water, 0.56 kg), high-heel walking (52 kg), and adult barefoot walking (72 kg) were applied as real impact sources with a child's (33 kg) running, jumping on the floor, and jumping from a chair, which were considered in previous study.[3] Among them, a child's running and jumping were generated only in the 210-mm-thick floor. The dropping height of the bang machine was 0.85 m, and that of the impact ball and a light-weight object was 1 m. Because interior surfaces of the test building were bare concrete, sound absorption materials were installed to adjust the reverberation time of the units (0.7 s) to that of a usual living room. Each impactor was used at the center of the driving room. Each impact sound was recorded binaurally through a dummy head (Brüel & Kjær 4100) at around the center of the receiving room, representing the typical listening location of a tenant below. The ears of the dummy head were located at the longitudinal position of 0.3 m away from the exact center of the room facing the entrance door, which indicates the modal point of the slab. In this study, floor impact sound levels were also obtained from the recorded impact sounds through a dummy head to investigate the comparative differences between impact sources.

Figures 2(a) and 2(b) compare the impact sound pressure level from the standard impactors with those of the real impact sources generated in the two units of the test building. These sounds are categorized into heavy-weight (adult walking, jumping, and running of child) and light-weight (dropping a plastic bottle and high-heel walking) impact sources. The correlation coefficients between the frequency characteristics in the range 31.5–2000 Hz of the standard impact sources and each human impact source are listed in Table I. The correlation coefficients for the impact ball are higher than those of the bang machine and tapping machine. This result extends the findings of the earlier study,[3] which reported that frequency characteristics of impact ball sound were more similar to real impact sounds such as child jumping and running. The average of the correlation coefficients for the impact ball is the highest (0.93), and those for the bang machine and tapping machine are almost the same (0.82). Thus, subjective similarity judgments were conducted to clarify which heavy-weight standard impact source was more similar to the real impact sounds.

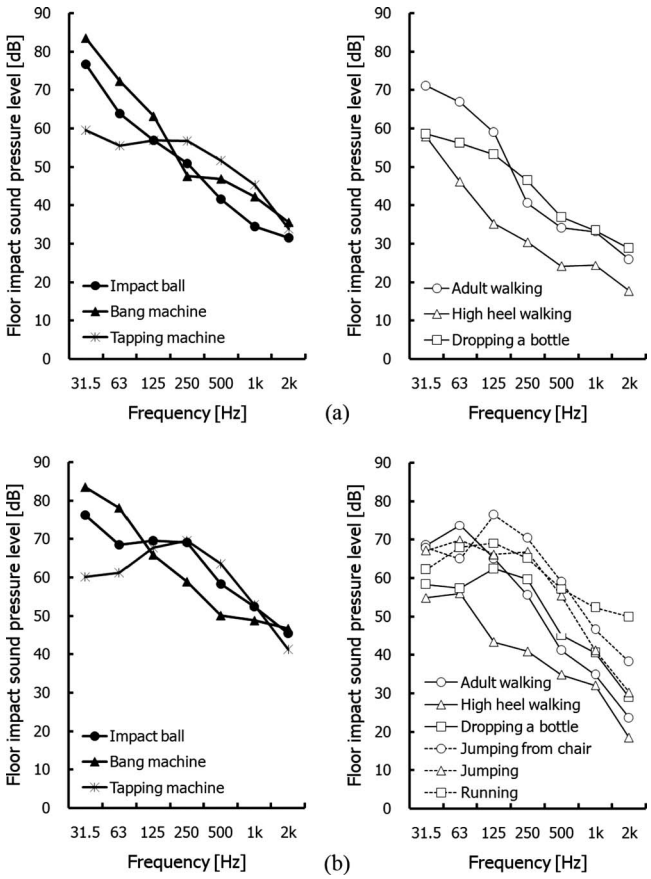


FIG. 2. Frequency characteristics of floor impact sounds generated in the two units of the test building: (a) 180-mm-thick floor and (b) 210-mm-thick floor.

TABLE I. Correlation coefficients between the frequency characteristics (31.5–2000 Hz) of each real impact source and standard impactors (average of two units).

	Weight (kg)	Impact ball	Bang machine	Tapping machine
Adult walking	72	0.95	0.95	0.74
Running	33	0.87	0.73	0.83
Jumping from a chair	33	0.92	0.66	0.92
Jumping on the floor	33	0.95	0.78	0.88
High-heel walking	52	0.95	0.96	0.70
Dropping a bottle	0.56	0.96	0.84	0.88

B. Subjective evaluation in a laboratory (similarity judgment)

In the subjective tests, the floor impact sounds recorded in a living room of an apartment unit were used. Because the test building did not reflect the spatial characteristics of real apartment unit, thus it may affect the subjective responses. The plan and floor section of the apartment are shown in Fig. 3. The floor layer of the apartment consisted of a 135-mm-thick concrete slab, a 70-mm-thick layer of lightweight concrete, and a 40-mm-thick layer of finishing mortar. The room was equipped with a set of furniture, and wooden flooring was installed as a finishing material. The reverberation time in the room was 0.54 s at 500 Hz. Previous study[3] adopted a 65 kg adult jumping as a real impact source, whereas in this study, human-made impact sounds were generated by two children jumping on the floor and jumping down from a chair to the floor. This time, the children were 5 and 10 years old and weighed 15 and 25 kg, respectively. The heavy-weight impact sound was generated with both the bang machine and the impact ball at the central position of the upstairs source room. The heavy-weight impact sounds were recorded binaurally through a dummy head (Brüel & Kjær 4100) positioned in the center of the receiving room on the floor below representing the typical listening location of a tenant.

Recorded floor impact sounds were presented through headphones (Sennheiser HD-600) as the same manner of previous study[18] in a soundproof chamber. It was been reported that the correct reproduction of binaural signals can be obtained by a proper artificial head equalization with the flat frequency response of the headphone.[24] The headphone used in the present study showed the flat frequency response within 2 dB in the frequency range of 32–2000 Hz. But the artificial head equalization to the signal presented was not applied. If the purpose of the experiment is to evaluate the floor impact sounds including the spatial information, it would be better that the actual sound field is represented as if the subject were sitting in the original recording place by a correct artificial head equalization, because the subjective responses to heavy-weight impact sound are related to the spatial impression as well as to the frequency characteristics.[25] However, the purpose of this study is to evaluate the similarity between human impact and standard floor impact sounds in terms of the frequency contents and SQ metrics, not the spatial impression such as interaural cross-correlation function, localization, and directional information. The average differences in SQ metrics between reproduced sounds through headphone and actual floor impact sources were 0.013 sone (loudness), 0.006 acum (sharpness), 0.304 asper (roughness), and 0.026 vacil (fluctuation strength).[18]

Fifteen human-made impact sounds were compared with heavy-weight impact sounds by the impact ball and the bang machine. The sound signals were presented to the subjects in groups. The human impact sound was always presented first, but the order for impact ball and bang machine sound was randomized. Before the experiment, the purpose and procedure of the experiment were explained to the subjects.

Fifteen subjects, each of whom had normal hearing, participated in the study. The subjects were asked to judge whether the impact ball impact sound or the bang machine impact sound was more similar to the human-made impact sound. The percentages of the impact ball selection for each human impact sound are shown in Fig. 4. "10yo" and "5yo"

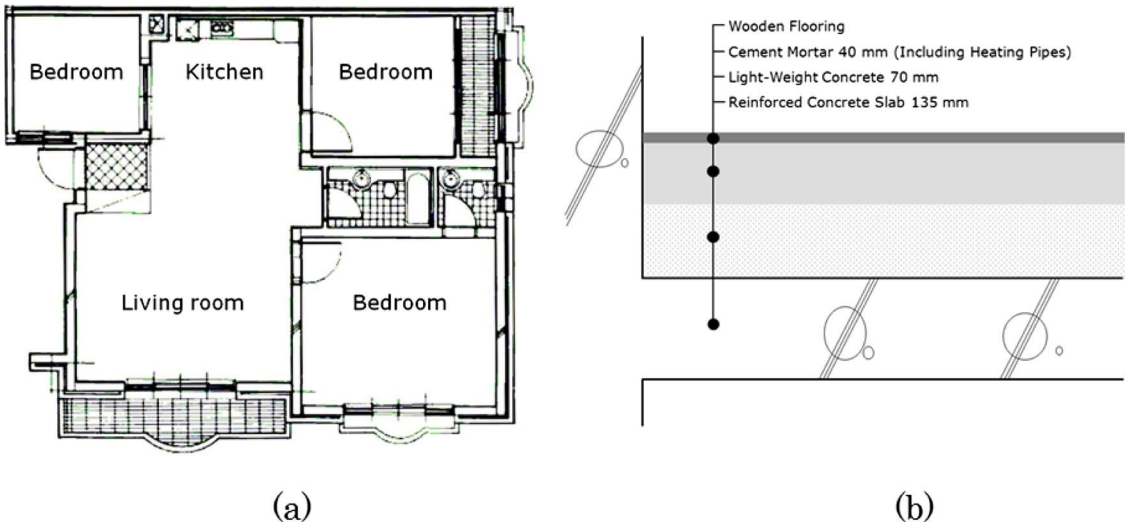


FIG. 3. (Color online) Plan of the apartment unit where the measurements of the floor impact sounds for the subjective evaluation were conducted (a) and section details (b).

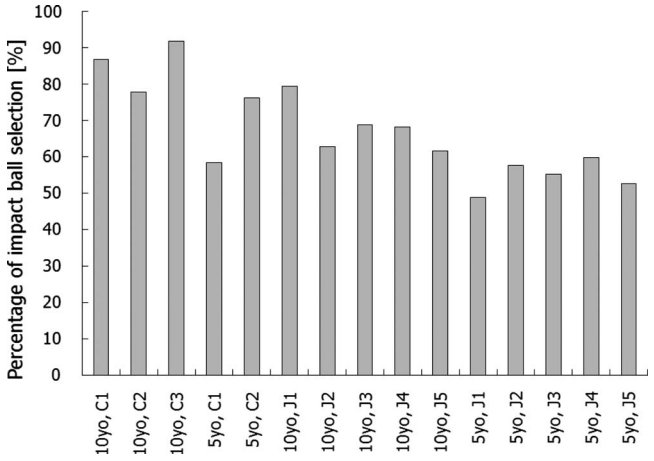


FIG. 4. Percentage of the impact ball selection of similarity judgments (10yo: 10-year-old child, 5yo: 5-year-old child, C: jumping from a chair to the floor, and J: jumping on the floor).

in the horizontal axis of Fig. 4 indicate the 10-year-old and 5-year-old children, respectively. In addition, "C" and "J" indicate jumping from a chair to the floor and jumping on the floor, respectively. The percentage of the impact ball selection as being closest to the real sound, for 14 of 15 stimuli, was more than 50%. The sound made by children jumping from a chair indicated a higher percentage of the impact ball selection than the sound of children jumping on the floor. The correlation between the subjective similarity (percentage of the impact ball selection) and the objective similarity (correlation coefficient between the frequency characteristics of the standard impact sources and each human impact sources) was not significant ($r=0.43$ for the impact ball and $r=-0.22$ for the bang machine). These results suggested that subjective similarity cannot be explained by only the frequency characteristics; therefore other parameters than sound pressure level should be introduced.

C. SQ evaluation of floor impact sounds

SQ metrics (loudness, roughness, and fluctuation strength) were used to characterize the subjective aspects of floor impact noise. Roughness and fluctuation strength describe the amount of mid-to-low frequency (from 20 to 300 Hz with maximum at 70 Hz) and the low-frequency (20 Hz) variations in the signal amplitude, respectively. The values depend on modulation frequency, modulation depth, and sound pressure level. In general, roughness and fluctuation strength are defined for stationary signals. But the attenuation pattern of floor impact sound varies due to absorption/reverberation of the receiving room and floor/ceiling structure. Such attenuation pattern may be described by roughness and fluctuation strength. The previous study[18] showed that roughness and fluctuation strength, as well as loudness, were related to the perception of heavy-weight impact sounds. Sharpness was not considered in this study because the frequency components in heavy-weight impact sounds below 1 kHz significantly affect the perceived loudness. The procedures for calculating loudness of a stationary sound are set forth in ISO 532;[20] however, a procedure for calculating loudness and the other SQ metrics of an impulsive sound has not yet been standardized. Thus, the results calculated by one software package may differ from those attained by others. In the present study, SQ metrics were calculated using pulse software (Brüel & Kjær) as a same manner of the previous study.[18] The time interval between spectra was set at 5 ms for the calculation of specific roughness and fluctuation strength.

Figure 5 shows examples of human impact sound with higher and lower percentages of the impact ball selection in terms of SQ metrics. The left and right sides of Fig. 5 indicate the higher and lower percentages of the impact ball selection, respectively. Three human impact sounds ("10yo, C3," "5yo, C2," and "10yo, J1") were selected as a higher percentage category and three impact sounds ("5yo, J3," "5yo, J1," and "5yo, J5") were chosen as a lower percentage category. The thick black line indicates the impact ball sound. The gray thick line indicates the bang machine sound. The other lines with circles, triangles, and squares indicate the human impact sounds. As shown in Fig. 5(a), the human impact sounds with a higher percentage of the impact ball selection indicate that loudness has a peak at 1.5–3 barks. The frequency characteristics of the loudness of the impact ball is similar to the above-mentioned, six human sounds. The main frequency range of floor impact sounds is up to 6–7 barks in terms of loudness. As shown in Fig. 5(b), the human impact sounds with a lower percentage of the impact ball selection indicate that loudness does not have a peak at 1.5–3 barks, except for "5yo, J1." This frequency characteristic was similar to that of the bang machine sound.

The roughness of the floor impact sounds was less than 0.02 asper, up to 2 barks. The roughness of the impact ball and the bang machine was about 0.05 larger than those of the human impact sounds above 3 barks [Fig. 5(c)]. The roughness of the human impact sounds was less than 0.08 asper at the frequencies investigated. There was little difference of roughness between the higher and lower percentages of the impact ball selection [Fig. 5(d)].

The fluctuation strength of the human impact sounds was 0 vacil up to 3 barks, while the fluctuation strength of the impact ball and the bang machine was about 0.1 vacil up to 7 barks [Fig. 5(e)]. The fluctuation strength of the human impact sounds with the lower percentage of the impact selection was higher than that of the higher percentage of the impact ball selection [Fig. 5(f)].

To quantify the objective similarity between the children's jumping and the standard impact sounds, a correlation coefficient of the SQ metrics for the human and standard floor impact sounds was calculated. Then the objective and subjective similarities were investigated using multiple regression analysis. The iteration regarding the frequency range was conducted to obtain a regression equation that produced a maximum regression coefficient. When the correlation coefficient between the human and the standard impact sounds was calculated by using SQ metrics in the range 0.5–6.0 barks, the regression coefficient indicated was maximum. This result may be reasonable because the loudness indicated that the main frequency component of the floor impact sounds were up to 5–7 barks.

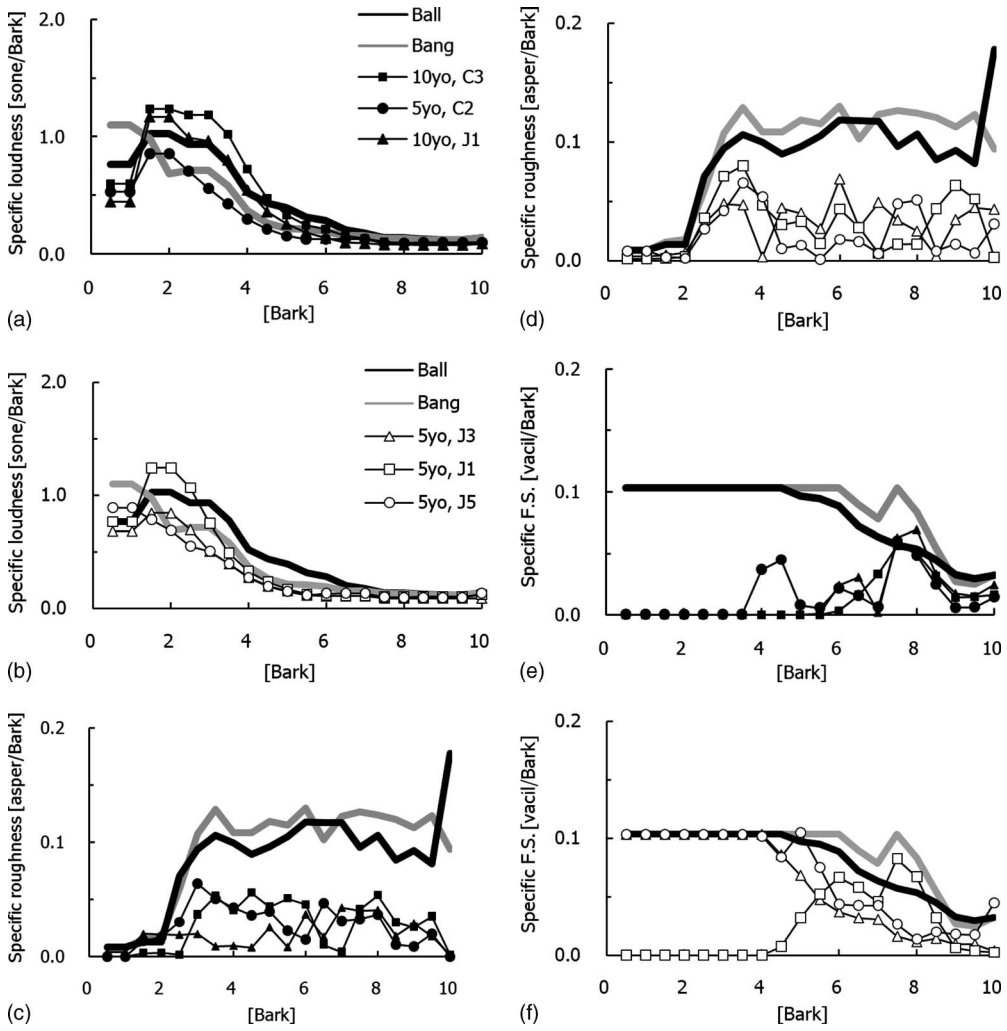


FIG. 5. Example of the human impact sound with higher and lower percentages of the impact ball selection in terms of [(a) and (b)] loudness, [(c) and (d)] roughness, and [(e) and (f)] fluctuation strength.

$$P_{\text{ball}} \approx a_1 \text{Corr\_Loud}_{\text{ball}} + a_2 \text{Corr\_Loud}_{\text{bang}} + a_3 \text{Corr\_Rough}_{\text{ball}} + a_4 \text{Corr\_Rough}_{\text{bang}} + c. \quad (1)$$

P_{ball} indicates the subjective similarity (the percentage of the impact ball selection) and the independent variables indicate the objective similarity (correlation coefficient of the SQ metrics for the human and standard floor impact sounds) in Eq. (1). For example, $\text{Corr\_Loud}_{\text{ball}}$ indicates the correlation coefficient of loudness for human impact sounds and the impact ball sound. The standardized partial regression coefficients a_1, a_2, a_3, and a_4 in Eq. (1) are 0.17, −0.53, 0.13, and −0.12, respectively. Using these values, the total coefficient 0.62 was obtained. The correlation between independent variables was not significant. The coefficient for the impact ball ($\text{Corr\_Loud}_{\text{ball}}$ and $\text{Corr\_Rough}_{\text{ball}}$) and the bang machine ($\text{Corr\_Loud}_{\text{bang}}$ and $\text{Corr\_Rough}_{\text{bang}}$) had positive and negative values, respectively. Thus, a higher objective similarity with the impact ball resulted in a higher percentage of the impact ball selection, while a lower objective similarity with the bang machine resulted in a higher percentage of the impact ball selection.

III. CONVERSION OF FLOOR IMPACT SOUND LEVELS FOR STANDARD SOURCES

In Sec. II, it was shown that the impact ball sound was judged to be more similar to children's jumping than the bang machine. Thus, the impact ball is more appropriate to the standard heavy-weight impact source in subjective and objective evaluations. Considering the fact that the heavy-weight impact sound has been evaluated with the bang machine for the past 30 years, a conversion method from the floor impact sound level of the bang machine into that of the impact ball should be proposed to use existing bang machine data. This section discusses the method of conversion of the floor impact sound level generated by the bang machine to that of the impact ball. In addition, the conversion of the sound levels of the tapping machine into those of the impact ball is also discussed.

To obtain the floor impact sounds with variations in structures and sound insulation treatments, more floor impact sound measurements were conducted. Floor impact sounds were recorded in 39 living rooms and 62 bedrooms of 35 reinforced concrete apartments. Most tested units had an area of about 100–120 m^2, which is the most common apartment

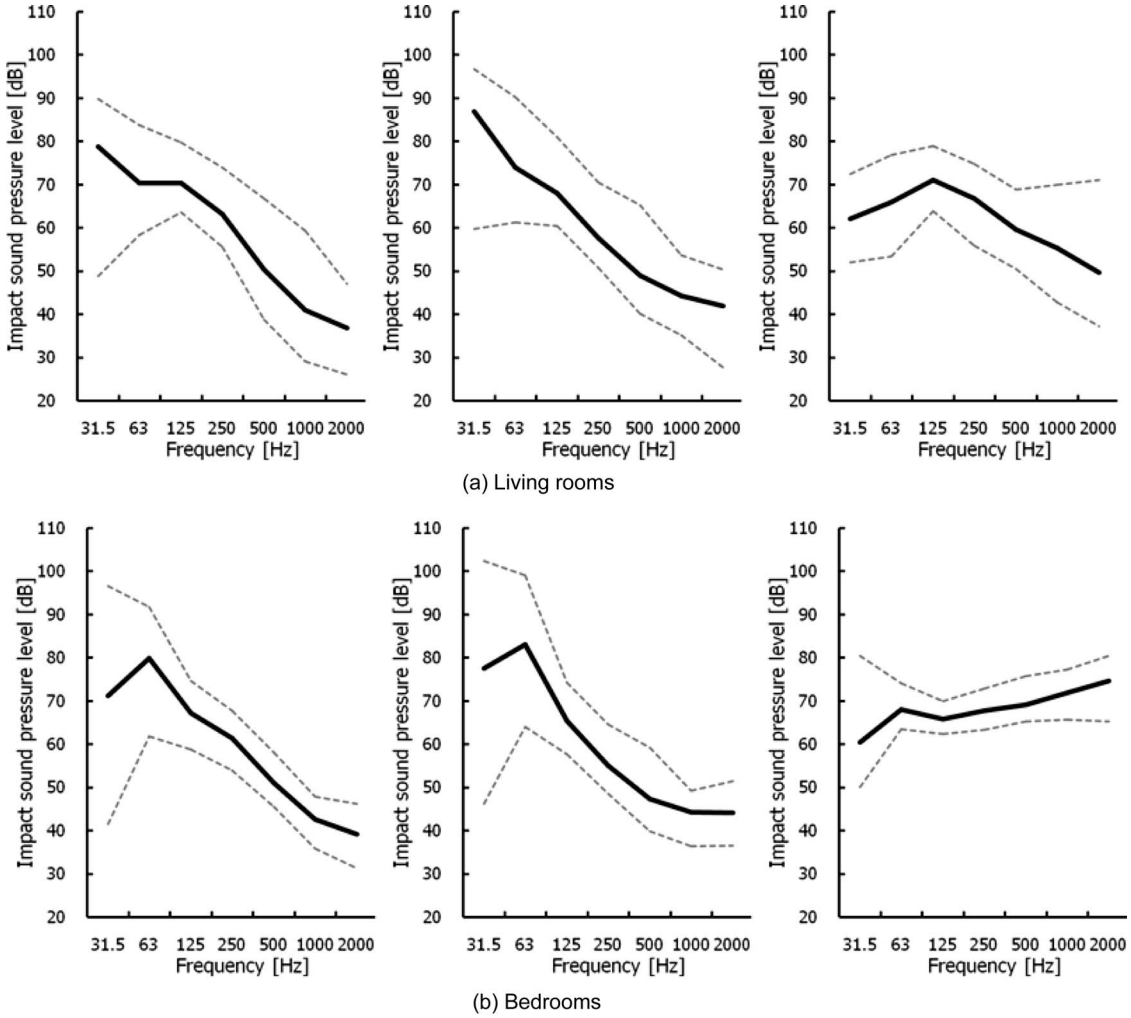


FIG. 6. Frequency characteristics of the impact ball (left), the bang machine (middle), and the tapping machine (right).

size in Korea. The concrete slab in each apartment was 150–180-mm-thick. The floor structure of the apartments consists of a concrete slab, lightweight concrete, and finishing mortar. And most floor structures have a resilient isolator between the concrete slab and the lightweight concrete for reducing floor impact sound. Nowadays in Korea, most living rooms are finished with wooden flooring, and bedrooms are covered with either wooden flooring or laminated paper flooring. These rigid floor coverings contribute to constant impact by heavy-weight impact sources, while minimizing a low possibility of conversion. In every apartment unit, floor impact sounds were generated at the center of the room by a bang machine, an impact ball, and a tapping machine. The floor impact sounds were recorded binaurally through a dummy head positioned at the center of the room below. The measured floor impact sound level data show the variations of frequency characteristics due to the different measurement conditions such as floor plan, interior finishing, floor structure, and sound insulation treatment. The floor impact sounds were classified into two groups, "living rooms" and "bedrooms," because the room sizes and boundary conditions affect the frequency characteristics. The frequency characteristics of the floor impact sounds for the bang machine, the impact ball, and the tapping machine are shown in Fig. 6. The solid line corresponds to the average sound pressure levels in each group, and the dashed lines correspond to the range of sound pressure levels.

In Korea and Japan, floor impact sound levels for the impact ball and the bang machine are evaluated in terms of $L_{i,F\max,AW}$ (inverse A-weighted impact sound pressure level).[26,27] To determine the $L_{i,F\max,AW}$, the measured maximum impact sound pressure levels ($L_{\max}$) were plotted against four-octave band frequencies from 63 to 500 Hz. The fitting procedure allowed for a total deviation of 8 dB above the inverse A-weighted reference curve in each of the four-octave bands. The $L_{i,F\max,AW}$ of the floor was read as the impact sound pressure level at 500 Hz on the inverse A-weighted reference curve. The $L_{i,F\max,AW}$ of the bang machine ranged from 41 to 63 dB, whereas that of the impact ball ranged from 45 to 66 dB. In the case of the tapping machine, single-number rating value, $L'_{n,AW}$, is determined at the levels of 125, 250, and 500 Hz or higher octave bands. ISO 140-11 describes the measurement results of the impact ball as impact sound pressure levels in the frequency range of 63–500 Hz and does not introduce any single-number rating value. Thus, the conversion methods of floor impact sounds were proposed to predict the sound pressure levels at each octave band frequency from 63 to 500 Hz.

TABLE II. Level differences between the impact ball and other standard sources (dB). (a) Impact ball-bang machine and (b) impact ball-tapping machine.

Groups	Frequency (Hz) 31.5	63	125	250	500	1000	2000
			(a)				
Living room	−7.9	−3.6	2.2	5.2	1.2	−3.3	−5.0
Bedroom	−6.1	−3.6	1.5	6.1	3.8	−2.1	−5.6
			(b)				
Living room	16.9	4.2	−1.0	−3.8	−9.6	−14.6	−13.1
Bedroom	9.3	11.3	1.5	−5.5	−15.5	−26.0	−31.1

A. Conversion method 1

The difference between average sound pressure level of the impact ball sounds and floor impact sounds generated by other standard impact sources was calculated at each octave band frequency from 63 to 500 Hz, and listed in Table II. The sound pressure levels of the bang machine and tapping machine were converted to those of the impact ball using

$$\mathrm{SPL}_{(\text{ball: 63–500 Hz})} = \mathrm{SPL}_{(\text{bang: 63–500 Hz})} + \Delta\mathrm{SPL}_{(\text{ball-bang: 63–500 Hz})}, \quad (2)$$

$$\mathrm{SPL}_{(\text{ball: 63–500 Hz})} = \mathrm{SPL}_{(\text{tapping: 63–500 Hz})} + \Delta\mathrm{SPL}_{(\text{ball-tapping: 63–500 Hz})}. \quad (3)$$

In Eqs. (2) and (3), ΔSPL indicates the sound pressure level difference between the impact ball and other standard impact sources at each octave band. Figure 7(a) shows the relationship between the measured sound pressure levels for the impact ball and the converted sound pressure levels from the bang machine. The correlation coefficients between the measured and the converted sound pressure levels are listed in Table III. The correlation coefficients at 63 Hz are higher than those at other frequencies, and bedrooms show slightly higher correlation than living rooms.

In 45 of 101 rooms, measured data of the tapping machine are available; thus, the conversion of the sound level from the tapping machine to the impact ball was also investigated. Figure 7(b) shows the relationship between the measured sound pressure levels for the impact ball and the converted sound pressure levels from the floor impact pressure levels of the tapping machine. In this case, the correlation coefficients are slightly higher than those from the bang machine even though the sound pressure levels were calculated in terms of L_{eq} and not L_{max}.

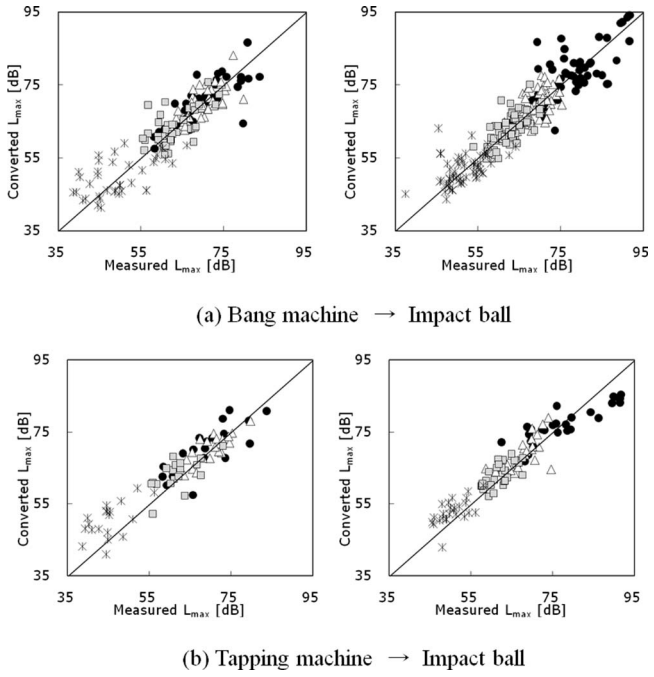


FIG. 7. Relationship between the measured and calculated sound pressure levels (method 1): (●) 63 Hz, (△) 125 Hz, (■) 250 Hz, and (∗) 500 Hz (living rooms: left, bedrooms: right).

B. Conversion method 2

Another conversion method using a regression analysis was investigated. A linear regression analysis was performed on the sound pressure levels for the impact ball at each octave band (dependent variable), and also on the sound pressure levels for the bang machine and tapping machine as a single predictor. Equations (4) and (5) were derived from regression analysis, and unstandardized regression coefficients of each equation are listed in Table IV. The sound pressure levels for the bang machine and tapping machine at

TABLE III. The correlation coefficients between the measured and converted sound pressure levels. (a) Bang machine → impact ball and (b) tapping machine → impact ball.

Groups	Frequency (Hz) 63	125	250	500
	(a)			
Living room	0.72[a]	0.63[a]	0.55[b]	0.37[b]
Bedroom	0.80[a]	0.69[a]	0.59[a]	0.52[a]
	(b)			
Living room	0.78[a]	0.80[a]	0.65[a]	0.59[b]
Bedroom	0.88[b]	0.66[a]	0.64[b]	0.79[a]

[a] $p < 0.01$.
[b] $p < 0.05$.

TABLE IV. Unstandardized coefficients and correlation coefficients of Eqs. (4) and (5).

Frequency (Hz)	Room	Bang machine→impact ball			Tapping machine→impact ball		
		a_1	c	r	a_1	c	r
63	Living room	0.78	12.6	0.72[a]	0.85	12.9	0.78[a]
	Bedroom	0.75	16.9	0.80[a]	1.53	−23.1	0.88[a]
125	Living room	0.60	29.2	0.63[a]	0.84	10.6	0.80[a]
	Bedroom	0.53	32.9	0.69[a]	0.59	27.7	0.66[a]
250	Living room	0.54	31.7	0.55[b]	0.71	14.5	0.65[a]
	Bedroom	0.58	30.2	0.69[a]	0.59	22.0	0.64[a]
500	Living room	0.85	8.8	0.63[a]	0.55	12.0	0.59[b]
	Bedroom	0.59	23.7	0.52[b]	0.64	7.1	0.79[b]

[a] $p<0.01$.
[b] $p<0.05$.

each octave band from 63 to 500 Hz were converted to those for the impact ball by Eqs. (4) and (5). The correlation coefficients between the measured and converted sound pressure levels at each octave band are also listed in Table IV. And the correlation coefficients are similar to those from conversion method 1. Figure 8 shows the relationship between the measured sound pressure levels for the impact ball and the calculated sound pressure levels by the floor impact levels of the bang machine and tapping machine.

$$\text{Sound pressure level(ball:63,125,250, or 500 Hz)} = a_1 L_{\max}(\text{bang:63,125,250, or 500 Hz}) + c, \quad (4)$$

$$\text{Sound pressure level(ball:63,125,250, or 500 Hz)} = a_1 L_{\text{eq}}(\text{tapping:63,125,250, or 500 Hz}) + c. \quad (5)$$

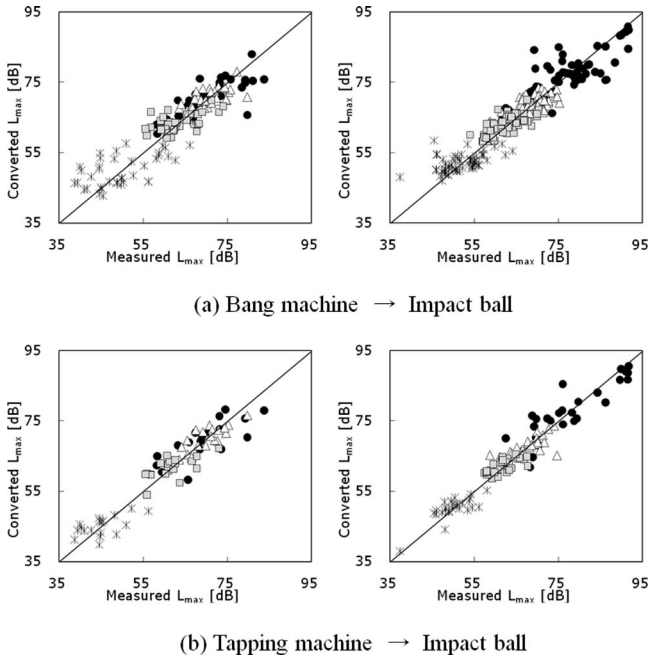


FIG. 8. Relationship between the measured and calculated sound pressure levels (method 2): (●) 63 Hz, (△) 125 Hz, (■) 250 Hz, and (∗) 500 Hz (living rooms: left, bedrooms: right).

The single-number rating values ($L_{i,F\max,AW}$) were calculated using the predicted sound pressure levels from conversion methods 1 and 2. In the previous study, 5 dB in terms of $L_{i,F\max,AW}$ was proposed as the level difference for the rating of floor impact noise.[3] In conversion method 1, 96 of 101 cases for the bang machine, and 43 of 45 cases for the tapping machine showed the difference between measured and predicted $L_{i,F\max,AW}$ to be less than 5 dB. And 97 of 101 cases for the bang machine, and 44 of 45 cases for the tapping machine showed the level difference to be less than 5 dB using conversion method 2.

IV. DISCUSSION

It was found that the frequency characteristics of the floor impact sounds generated by standard heavy-weight impact sources are similar to those of human-made impact sounds in the objective evaluation. The results of the subjective tests also showed that the floor impact sound generated by the impact ball is more similar to human-made impact sounds than to those of the bang machine. Thus, subjective and objective evaluations of the impact ball are more accurate compared to the standard heavy-weight impact source. Previous study[3] described the similarity to the impact sounds made by children's jumping by the frequency characteristics, but the result of this study showed that the similarity could not be explained by the floor impact level only. To further explain these sounds, similarity tests were conducted and SQ metrics were used to evaluate these floor impact sounds. The important factors for evaluating the similarity between real human impacts and standard impact sources are loudness and roughness. Listeners judged that the real impact sources are similar to the impact ball because (1) the sound characteristics of the impact ball are similar to the real impact sources (similarity) and (2) the sound characteristics of the bang machine are not similar to the real impact sources (dissimilarity). The frequency ranges in 1–6 barks were important for evaluation of similarity in terms of loudness. Our previous study[18] showed that loudness and fluctuation strength were significant in describing the subjective loudness of the heavy-weight floor impact sounds. Fluctuation strength,

which reflects the level of fluctuation in the time domain, can describe the difference of sound insulation treatment of the apartment floors; subjective evaluation is based on this factor when heavy-weight floor impact sources are used. However, the difference of fluctuation strength of the impact ball and the bang machine is small with respect to the difference between the human impact sound and the standard impact sources. Therefore, fluctuation strength in this study cannot describe the similarity of the standard impact sources to human impact sounds. The target modulation frequencies of roughness and fluctuation strength were around 70 and 4 Hz, respectively. Relatively fast fluctuation, which is related more to roughness rather than fluctuation strength, was more suitable for describing the subjective similarity to human floor impact sounds.

In another study[28] impact ball sounds were classified into groups A-C, which indicated that impact sound level deviation above the inverse *A*-weighted reference curve was the highest at 63, 125, 250, and 500 Hz in octave bands. Groups A-C had loudness peaks at 0.5–1, 1.5–2, and 2–3 barks, respectively, according to frequency characteristics. The human impact sounds with higher percentage of the impact ball selection showed loudness characteristics of groups B and C. Impact sounds generated by an adult walking, and a 33 kg child's jumping and jumping from a chair, shown in Fig. 2, can also be classified into groups B and C; whereas bang machine sound can be classified into group A. In addition, impact sounds generated by a 25 kg child jumping and running, and the impact ball show the frequency characteristics of group B, and the bang machine sound shows that of group A in a previous study.[3] This indicates that even though the subjective evaluation in a laboratory has the limited sample of an apartment unit, impact ball sounds can be perceived more similar to human impact sounds than to bang machine sounds.

In a previous study,[3] an on-site auditory experiment was conducted with 98 subjects to acquire a rating for the impact sound level of the impact ball according to annoyance. The impact ball, bang machine, tapping machine, and an adult jumping were used as noise sources. The annoyance was evaluated according to three aspects of noise perception (noisiness, disturbance, and amenity) using a nine-point scale. The perceived floor impact sound level of the impact ball noise was more similar to the noise of a jumping adult than to the bang machine noise. Jeong[29] investigated the percentage of subjects' satisfaction on floor impact noises using the annoyance on-site auditory experiment results of the previous study.[3] The satisfaction percentage was calculated by the Probit analysis[30,31] and 1–4 for the nine-point scale was assigned as "satisfaction." As shown in Table V, 50% satisfaction levels for the bang machine, impact ball, and an adult jumping were at 48, 57, and 56 dB ($L_{i,F\max,AW}$), respectively. Average slopes of the regressions for the bang machine, the impact ball, and the adult jumping were 10, 6, and 4%/dB, respectively. These results also show that the subjective responses to floor impact noises and impact ball noises were the most similar of the three standard impact sources to actual human impact.

TABLE V. Floor impact noise levels ($L_{i,F\max,AW}$) at 50% satisfaction.

Evaluation scale		Bang machine	Impact ball	Adult jumping
Noisiness	(dB)	46	54	53
	(%/dB)	12	7.4	4.4
Disturbance	(dB)	51	60	57
	(%/dB)	8.7	4.8	3.6
Amenity	(dB)	48	58	58
	(%/dB)	9.8	5.5	3.7
Average	(dB)	48	57	56
	(%/dB)	10	6	4

Regarding the sound level, measured by using the tapping machine, the C_I term[15] can be applied to the single-number index to consider the floor structure itself. However, as shown in Fig. 9, the $L_{i,F\max,AW}$ of floor impact noises generated by the impact ball showed less correlation with $L'_{n,AW}$ of the tapping machine ($r=0.33$). It also shows the floor impact level for the tapping machine after applying the spectrum adaptation term C_I to the $L'_{n,AW}$. Even though C_I is introduced, there is a large discrepancy between the heavy- and light-weight impact sounds ($r=0.44$). The tapping machine produced the continuous sounds, which were analyzed by L_{eq}. However, our previous study[18] showed that the level difference between the peak and the tail parts affects annoyance of the floor impact sounds. The temporal characteristics of the decay energy of the floor impact sound were described by the SQ metrics and ACF parameters. Thus, the tapping machine is not the appropriate floor impact source to simulate real human floor impacts.

It is noted that the conversion method of the bang machine and tapping machine sounds to the impact ball sounds proposed in this study cannot be directly applied to other floors such as wooden structures. The reason is that both the impact source and the floor have mechanical impedance, and the floor impact sounds are affected by the ratio of the impedances of the impact sources and floors.[11] The floor impact sounds used in this study are measured in box-frame-type

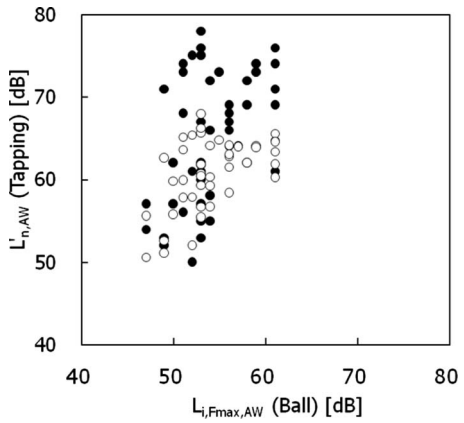


FIG. 9. Relationship between the $L_{i,F\max,AW}$ of the impact ball and the L'_{nAW} of the tapping machine (●: $r=0.42$, $p<0.01$) as well as the floor impact level for the tapping machine after applying the spectrum adaptation term CI to the $L'_{n,AW}$ (○: $r=0.52$, $p<0.01$).

apartments of reinforced concrete structures. Thus, the interaction between the impact source and the floor in the reinforced concrete floor is different from that in other floor structures.

Nonlinearity problems caused by floor coverings applied in this study were investigated using standard impact sources with different drop-heights, because if some floor construction and finishing with nonlinearity were included, the simple conversion would be impossible. For linear system, each successive increase in drop-height would correspond to increase in impact sound pressure level.[32] In this study, vibration acceleration level of the floor structures was only considered because impact sound levels can be affected by other factors such as room mode of receiving room. The experiment was conducted in a room with 180-mm-thick of concrete slab in a test building. When standard impact sources (bang machine, impact ball, and tapping machine) were generated in the center of the source room, the vibration acceleration was measured ten times using an accelerometer located on the ceiling of the receiving room. In case of heavy-weight impact sources, drop-heights were varied from 60 to 140 cm in steps of 20 cm. And tapping machine had three different drop-heights: 2, 4, and 6 cm. The vibration acceleration levels from the floor with different floor coverings (wooden flooring, laminated paper, and linoleum) were analyzed. As shown in Fig. 10, the higher drop-height causes the larger vibration acceleration level linearly. This indicates that the contact stiffness of floor covering used in this study can be considered as a linear spring. The slopes of the regression lines between drop-heights and vibration acceleration levels for bang machine and impact ball were almost same, whereas the slopes of the regression lines for tapping machine were different to respective floor finishing coverings; because impact force of the tapping machine transmitted to the floor structure, which is much less than that of heavy-weight impact sources, was reduced in the cases for the linoleum and wooden flooring.

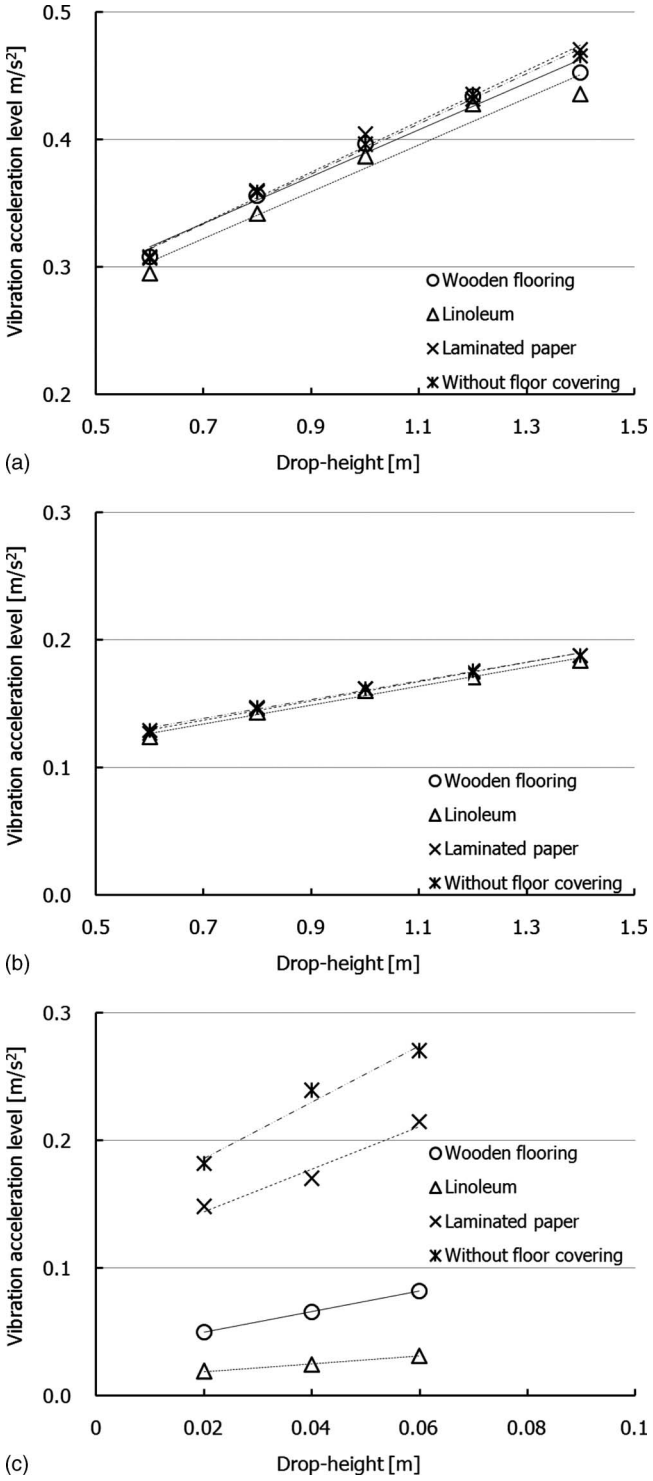


FIG. 10. Vibration acceleration levels measured using different drop-heights on standard impact sources: (a) bang machine, (b) impact ball, and (c) tapping machine.

V. CONCLUSIONS

Objective and subjective evaluations of the similarity between human impact sounds and standard floor impact sources were conducted. Measurements of floor impact sounds generated by the standard impact sources and real impact sources showed that the frequency characteristics of real human floor impacts were similar to those of the heavy-weight standard impact sources. The similarity tests between the human-made impact sounds and the standard heavy-weight impact sounds showed that the impact sound level of an impact ball is more similar to a human-made impact sound than the bang machine. The results for the jumping children and the impact ball, in contrast to the bang machine, indicated that the use of an impact ball is suitable to evaluate heavy-weight impact sounds. The investigation of the SQ metrics of the heavy-weight floor impact sounds showed that loudness and roughness are important for evaluating the similarity of the standard and the human-made floor impact sounds.

To use the bang machine and tapping machine data collected in earlier experiments, a conversion method was proposed to convert floor impact sound level data of the bang machine and tapping machine into that of impact ball data. Conversion method 2, using a correlation analysis, was slightly better than the method that applied the sound pressure level difference between the impact ball and other standard impact sources. In order to prove the reliability of the

conversion methods in light-weight floors, such as wooden floors, much of the measurement data in light-weight floors should be investigated, and then the difference of conversion methods between heavy-weight and light-weight floors should be compared.

The impact ball is included as a heavy/soft impact source in ISO 140-11, which is concerned with laboratory measurement of the impact sound improvement of floor coverings on light-weight floors. Recently, ISO 10140-3 draft, which is applicable to all types of floors (heavy-weight or light-weight) with all types of floor covering (single or multilayered), includes the impact ball as a standard impact source in (informative) Annex A, but it does not provide *in-situ* assessment of impact sound insulation. This study confirms previous studies showing that the impact ball correlates well with subjective impressions of real heavy impacts in buildings. The empirical relationships resulting from this work indicate that it is possible to convert between the sound pressure levels from different impact sources in heavyweight buildings such as those made from reinforced concrete.

ACKNOWLEDGMENTS

This work was supported by Grant No. 10023489 from the Standardized Technology Development Project of the Ministry of Commerce, Industry and Energy. P.J.L. was supported by a Seoul Fellowship from the Seoul Metropolitan Government.

[1]JIS A 1418: Acoustics: Measurement of floor impact sound insulation of buildings. Part 2: Method using standard heavy impact sources (Japanese Standards Assn., Tokyo, Japan, 2000).
[2]KS F 2810: Method for field measurement of floor impact sound insulation. Part 2: Method using standard heavy impact sources (Korean Standards Assn., Seoul, Korea, 2001).
[3]J. Y. Jeon, J. K. Ryu, J. H. Jeong, and H. Tachibana, "Review of the impact ball in evaluating floor impact sound," Acta Acust. Acust., **92**, 777–786 (2006).
[4]K. Bodlund, "Alternative reference curves for evaluation of the impact sound insulation between dwellings," J. Sound Vib. **102**, 381–402 (1985).
[5]D. Olynyk and T. D. Northwood, "Subjective judgments of footstep-noise transmission through floors," J. Acoust. Soc. Am. **38**, 1035–1039 (1965).
[6]W. Shi, C. Johansson, and U. Sundback, "An investigation of the characteristics of impact sound sources for impact insulation measurement," Appl. Acoust. **51**, 85–108 (1997).
[7]B. G. Watters, "Impact-noise characteristics of female hard-heeled foot traffic," J. Acoust. Soc. Am. **37**, 619–630 (1965).
[8]A. C. C. Warnock, "Low frequency impact sound transmission through floor systems," Proceedings of Inter-noise 2000 (2000).
[9]A. C. C. Warnock, "Prospects for a test method for rating floor-toppings for use on joist floors," J. Acoust. Soc. Am. **112**, 2227 (2000).
[10]W. Scholl and W. Maysenhoelder, "Impact sound insulation of timber floors: Interaction between source floor coverings and load bearing floor," Build. Acoust. **6**, 43–61 (1999).
[11]W. Scholl, "Impact sound insulation: The standard tapping machine shall learn to walk!" Build. Acoust. **8**, 245–256 (2001).
[12]ISO 140: Acoustics—Measurement of sound insulation in buildings and of building elements. Part 11: Laboratory measurements of the reduction of transmitted impact sound by floor coverings on lightweight reference floors (International Organization for Standardization, Geneva, 2005).
[13]ISO 717: Acoustics—Rating of sound insulation in buildings and of building elements. Part 2: Impact sound insulation (International Organization for Standardization, Geneva, 1996).
[14]H. Tachibana, H. Tanaka, M. Yasuoka, and S. Kimura, "Development of new heavy and soft impact source for the assessment of floor impact sound insulation of buildings," Proceedings of Inter-noise 98 (1998).
[15]J. Y. Jeon, J. H. Jeong, M. Vorländer, and R. Thaden, "Evaluation of floor impact sound insulation in reinforced concrete buildings," Acta Acust. Acust. **90**, 313–318 (2004).
[16]J. Y. Jeon, J. H. Jeong, and Y. Ando, "Objective and subjective evaluation of floor impact noise," Journal of Temporal Design in Architectural and the Environment **2**, 20–28 (2002); www.jtdweb.org (Last viewed 8/13/2008).
[17]H. Tachibana, H. Yano, and Y. Sonoda, "Subjective assessment of indoor noises-basic experiments with artificial sounds," Appl. Acoust. **31**, 173–184 (1990).
[18]J. Y. Jeon and S. Sato, "Annoyance caused by heavyweight floor impact sounds in relation to the autocorrelation function and sound quality metrics," J. Sound Vib. **311**, 767–785 (2008).
[19]E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin, 1999).
[20]ISO 532: Acoustics—Method for calculating loudness level (International Organization for Standardization, Geneva, 1975).
[21]J. R. Roberts, R. Jones, N. J. Mansfield, and S. J. Rothberg, "Evaluation of impact sound on the 'feel' of a golf shot," J. Sound Vib. **287**, 651–666 (2005).
[22]Y. S. Wang, C.-M. Lee, D.-G. Kim, and Y. Xu, "Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model," J. Sound Vib. **299**, 933–947 (2007).
[23]M. Tanaka and T. Murakami, "A study on the standardized heavy and soft impact source for the measurement of floor impact sound pressure level," Proceedings of the Symposium on Floor Impact Sound, Tokyo (2005), pp. 3–8.
[24]D. Hammershøi and H. Møller, "Methods for binaural recording and reproduction," Acta Acust. Acust. **88**, 303–311 (2002).
[25]J. Y. Jeon, P. J. Lee, J. H. Kim, and S. Y. Yoo, "Subjective evaluation of heavy-weight floor impact sounds in relation to spatial characteristics," J. Acoust. Soc. Am. **125**, 2987–2994 (2009).
[26]JIS A 1419: Acoustics: Rating of sound insulation in buildings and of building elements. Part 2: Floor impact sound insulation (Japanese Standards Assn., Tokyo, Japan, 2000).
[27]KS F 2863: Rating of floor impact sound insulation for impact source in buildings and of building elements. Part 1: Floor impact sound insulation against standard light impact source (Korean Standards Assn., Seoul, Korea, 2002).
[28]P. J. Lee, J. H. Kim, and J. Y. Jeon, "Psychoacoustical characteristics of impact ball sounds in concrete floors," Acta Acust. **95**, 707–717 (2009).
[29]J. H. Jeong, "Floor impact noise classification based on subjective evaluations and comparisons of standard impact sources," Ph.D. thesis, Hanyang University, Korea.
[30]J. H. Rindel, "Acoustic quality and sound insulation between dwellings," Build. Acoust. **5**, 291–301 (1999).
[31]G. Clausen, L. Carrick, P. O. Fanger, S. W. Kim, T. Poulsen, and J. H. Rindel, "A comparative study of discomfort caused by indoor air pollution thermal load and noise," Indoor Air **3**, 255–262 (1993).
[32]C. Hopkins, *Sound Insulation* (Butterworth Heinemann, Oxford, 2007), Chap. 3.6.

Multiple acoustic diffraction around rigid parallel wide barriers

Hequn Min[a] and Xiaojun Qiu
State Key Laboratory of Modern Acoustics and Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

(Received 18 September 2008; revised 30 March 2009; accepted 8 May 2009)

A ray-based method is presented for evaluating multiple acoustic diffraction by separate rigid and parallel wide barriers, where two or more neighboring ones are of equal height. Based on the geometrical theory of diffraction and extended from the exact boundary solution for a rigid wedge, the proposed method is able to determine the multiple diffraction along arbitrary directions or at arbitrary receiver locations around the diffracting edges, including the positions along the shadow or reflection boundaries or very close to the edges. Comparisons between the results of the numerical simulations and the boundary element method show validity of the proposed method.
 [DOI: 10.1121/1.3147491]

PACS number(s): 43.50.Gf, 43.20.El, 43.50.Lj [AJMD] Pages: 179–186

I. INTRODUCTION

To assess the impacts of noise on multiple residential buildings along highways, a quantitative description of the multiple sound diffraction over these buildings is required. Sometimes these buildings are similar with same height and the receiving points are close to the top edges of these buildings compared to the wavelength, such as the top floor windows. Separate parallel wide barriers with some neighboring ones of equal height can be considered as the simplified model of such buildings.

Much research has been undertaken on the multiple acoustic diffraction around similar obstacles. Fujiwara *et al.*[1] introduced a technique to estimate the double sound diffraction over one wide barrier by replacing the obstacle with an equivalent thin screen of a certain height, and then the thin screen's noise reduction due to the single diffraction can be evaluated with an empirical formula derived by Kurze and Anderson.[2] Though this technique has been applied to roughly estimate the sound attenuation due to diffraction for engineering purposes,[3,4] it leads to highly erroneous results for diffraction over several obstacles.[5] About 50 years ago, Keller[6–8] presented the geometrical theory of diffraction (GTD) to describe the diffraction. Although the GTD is a geometrical acoustics method, it is accurate for most practical cases when the sound wavelength is smaller than obstacle dimensions.[7] Pierce[5] presented an asymptotic solution and later an exact one together with Hadden[9] to solve the single diffraction around a wedge. He extended that asymptotic solution to evaluate the double-edge diffraction around a single wide barrier[5] based on the concepts of Keller's GTD.[6–8] Although the second-order diffraction term in this double-edge model has been afterwards extended by Chu *et al.*[10] to higher orders for evaluating the diffraction around a wide barrier with finite thickness, Pierce's methods[5,9] mentioned above have not been extended for the separate wide barriers yet.

Kawai[11] developed a method for diffraction around a rigid multi-sided barrier, which was later modified by Kim *et al.*[12] for many extended cases such as multiple wedges or knife edges and polygonal-like shapes. However, for diffracted waves traveling along the shadow or reflection boundaries from the edges, some terms in their methods become infinite, which needs additional complex asymptotic approach to approximate.[13] Additionally these methods require confirming the total field continuity close to each shadow or reflection boundary, which leads to quite complicated computation for the diffraction by several obstacles with more than four edges.

Based on the concepts of Pierce's double-edge model,[5] Salomons[14] presented a model for sound propagation over several wedges in three-dimensional field. In his method, however, both source and receiver are required being far from edges and there are singularities similar to the methods of Kawai[11] and Kim *et al.*[12] for diffraction along the reverse direction of the incident wave on edges.[9,14] Such diffraction occurs commonly around the coplanar edges on top of barriers with same height.

Wadsworth and Chambers[15] modified the Biot–Tolstory–Medwin model[16] for diffraction around single wide barrier or double knife edges in time domain, with both source and receiver far away from the edges also. But the computational load of this model is much greater than that of the frequency domain based solutions such as that of Salomons[14] in most practical cases.[15] Based on the GTD and statistical energy analysis, Reboul *et al.*[17] recently proposed equations able to evaluate the multiple diffraction around several diffracting edges. Nonetheless this method has acceptable accuracy only at relatively high frequencies because it considers the field as the energetic summation of different waves and loses the interference between the waves. Bougdah *et al.*[18] experimentally investigated the acoustic performance of a rib-like structure used for traffic noise control lately, which includes periodically spaced edges or walls, and no theoretical or numerical model on the multiple acoustic diffraction over such structure has been proposed yet.

[a]Author to whom correspondence should be addressed. Electronic mail: hqmin@nju.edu.cn

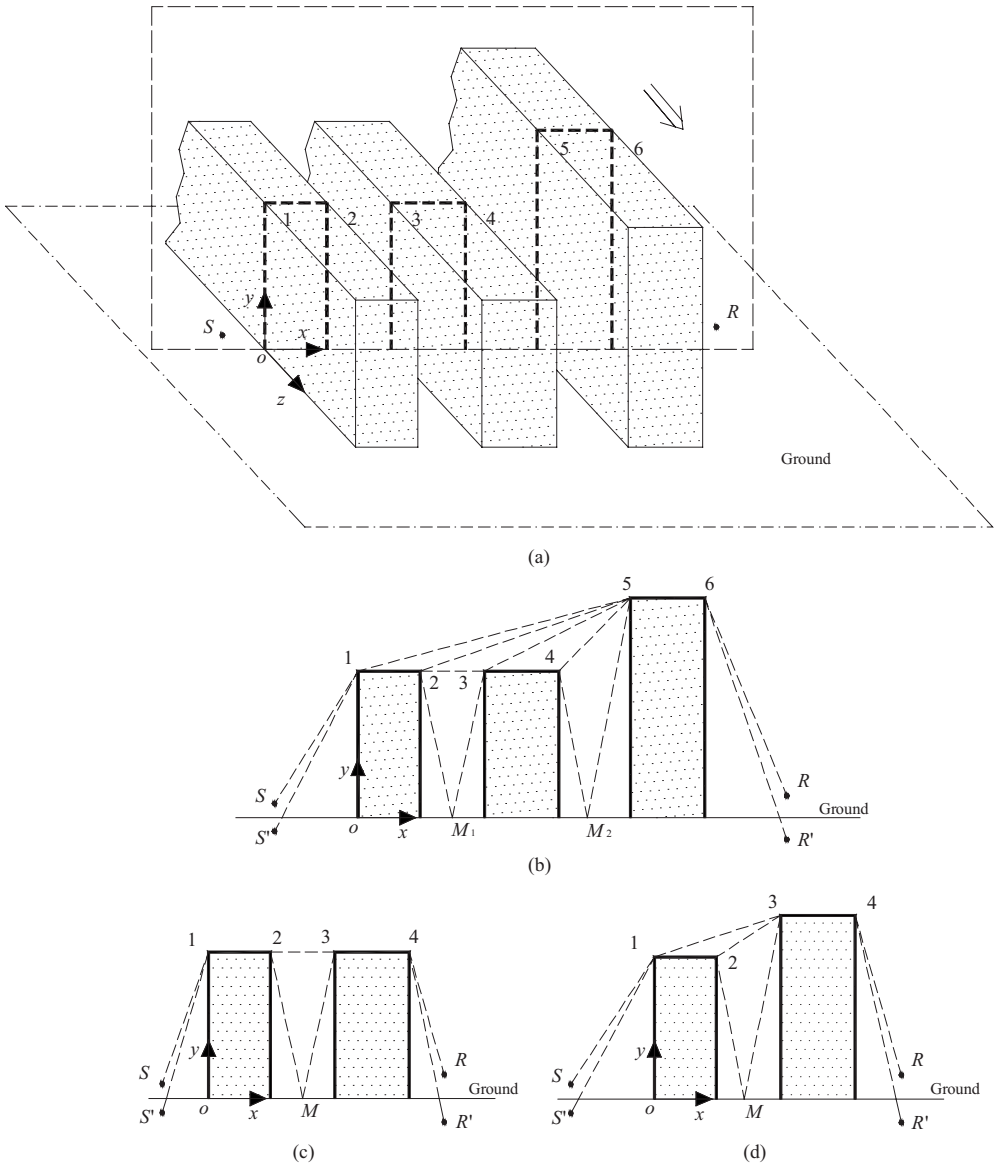


FIG. 1. Typical scenarios of parallel infinitely long wide barriers with source S and receiver R. The dashed lines in (b)–(d) represent the propagation paths of the diffracted waves over barriers. Barrier top edges 1,2,…,6 are diffracting edges. S' and R' are images of S and R to the infinite rigid ground, respectively. M, M_1, and M_2 are ground reflection points of the rays between two barriers. (a) Three-dimensional scenario with three barriers, where two neighboring ones are of equal height. A point source and receiver located in a same plane perpendicular to lengthwise axis of barriers. (b) Cross-section geometry of the scenario in (a). (c) Cross-section geometry with two barriers of equal height. (d) Cross-section geometry with two barriers of different heights.

Despite the previous studies reviewed above, currently there is no appropriate analytical solution for the multiple sound diffraction over a few rigid and parallel wide barriers yet, where some neighboring ones are of equal height. Based on Keller's GTD,[6–8] this paper proposes a method to evaluate the multiple diffraction at arbitrary receiver locations around such obstacles.

II. THEORETICAL METHOD

A typical scenario with three barriers is shown in Fig. 1(a), where two neighboring ones have the same height. Here infinitely long and rigid parallel wide barriers are assumed on the infinite and rigid ground. Right-handed Cartesian coordinates are defined and the origin is located on the intersection line between ground and the leftmost vertical side of the barriers. Only the incident wave normal to axis z (the lengthwise axis of barriers) is considered here. Accordingly the geometry in Fig. 1(a) can be simplified to a plane that is perpendicular to axis z and contains the receiver and source locations as shown in Fig. 1(b). The solution for oblique incidence can be easily obtained from the one for normal incidence with the method mentioned in Ref. 9. When the heights of all barriers are identical, only two barriers of equal height as shown in Fig. 1(c) are analyzed for succinctness.

Based on the GTD method,[6–8] the sound rays that are able to reach a certain receiving point come only from the sources and diffracting edges that can be "observed" from that point. And the multiple sound diffraction is described as individual multiply diffracted waves. Therefore the total sound field at receiver R around the wide barriers comprises the direct rays, the reflected rays, and all the diffracted rays. The direct or reflected rays are determined in the common

way. And the total diffracted field at receiver R with source S, $\phi_{d,\mathrm{tot}}(S,R)$, is the summation of overall diffracted rays coming along all possible diffraction paths[6,8] and is evaluated as

$$\phi_{d,\mathrm{tot}}(S,R) = \sum_{n=1}^{N} \phi_{d,n}(S,R|E_1,E_2,\ldots,E_n), \tag{1}$$

where $\phi_{d,n}(S,R|E_1,E_2,\ldots,E_n)$ represents the field of an individual ray $S \to E_1 \to E_2 \to \cdots \to E_n \to R$, which has been diffracted for n times (orders) and is called an n-order diffracted ray in this paper. E_1, E_2, …, and E_n are, respectively, the edge positions that the n-order diffracted ray propagates along in turn. Accordingly, $\phi_{d,1}(S,R|E_1)$ is the field of a singly diffracted ray and $\phi_{d,2}(S,R|E_1,E_2)$ is the field of a doubly diffracted ray as referred to in previous studies.[11,12] The value N is the considered maximum diffraction orders in the field. In the work presented, it is assumed that every two edges are spaced apart with a sufficiently large distance so that the rays, which are diffracted for two or more times by a same edge, can be neglected,[10,11] for example, the rays $S \to 1 \to 4 \to 1 \to 4 \to R$, $S \to 1 \to 4 \to 1 \to 4 \to 1 \to 4 \to R$, etc., in Fig. 1(c). The numerical results presented in Fig. 5 serve to validate this assumption, where these rays diffracted more than once by a same edge are found to be much weaker than the rays diffracted only once by each edge. Thus an n-order diffracted ray propagates along n different edges and N in Eq. (1) equals the number of all the edges.

A similar case with two wide barriers of different heights shown in Fig. 1(d) is taken as an example to illustrate the search scheme for all the possible diffracted rays reaching R in Figs. 1(b) and 1(c). When there is no ground and each vertical side of the barriers becomes semi-infinite in Fig. 1(d), the diffracted rays reaching R are $S \to 1 \to 2 \to 3 \to 4 \to R$ and $S \to 1 \to 3 \to 4 \to R$ only. After taking the ground reflection into account, ten additional rays appear coming from the images of source or edge 2 over the barriers to the image of receiver. Then the overall rays reaching R in Fig. 1(d) are $S(S') \to 1 \to 2 \to 3 \to 4 \to R(R')$, $S(S') \to 1 \to 2 \to M \to 3 \to 4 \to R(R')$, and $S(S') \to 1 \to 3 \to 4 \to R(R')$, where the letters in the brackets mean optional. S' and R' represent the respective images of the source and receiver to the ground and M is the ground reflection point of the rays from edge 2 to edge 3.

When the heights of these two barriers become identical as shown in Fig. 1(c), it is assumed that every edge can observe all the others and receive rays from all the others at its location. Then in Fig. 1(c) the total number of diffracted rays reaching R is counted as 28, whose details are not presented for concision. Based on cases in Figs. 1(c) and 1(d), the total number of diffracted rays reaching R in Fig. 1(b) is up to 116.

A. Diffraction coefficient

Before the presence of a diffracting edge E_l, the initial sound field at this edge location delivered by the ray $S \to E_1 \to E_2 \to \cdots \to E_l$ is denoted by ϕ_{ini}, where subscript l is an integer. Once the edge E_l is encountered, the initial ray is diffracted and the sound field of the corresponding diffracted

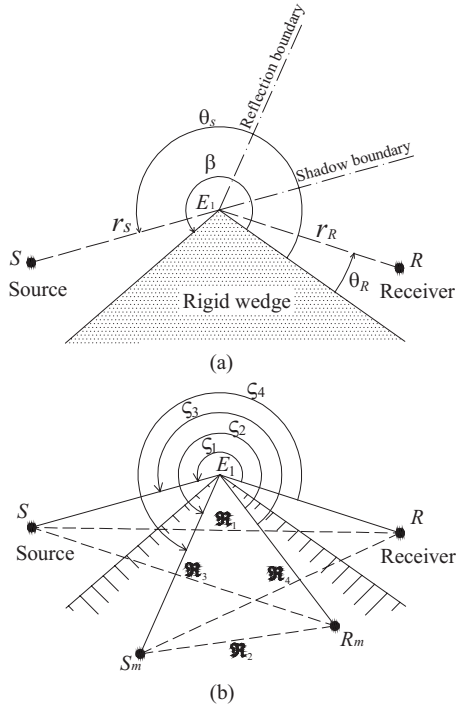


FIG. 2. Cross-section geometry of single diffraction over a rigid wedge whose vertex is edge E_1. (a) Illustration of the locations of source and receiver. θ_S and θ_R are turn angles from right side of wedge to source S and receiver R separately. r_S and r_R are distances in cross-section plane from edge E_1 to source S and receiver R, respectively. Shadow boundary is the extending line of the incident direction from S to E_1, and reflection boundary is the reflected line of the incident direction from S to E_1 due to the source-face of the wedge. (b) Illustration of the parameters ς_i in Eq. (3) and $\Re_i$ in Eq. (5). S_m and R_m are the images of S and R, respectively, to the nearest wedge face. $\varsigma_1 = \angle RE_1S$, $\varsigma_2 = \angle R_mE_1S_m$, $\varsigma_3 = \angle R_mE_1S$, and $\varsigma_4 = \angle RE_1S_m$, where each angle is determined by anticlockwise turning its initial side to its terminate side. Meanwhile $\Re_1 = |RS|$, $\Re_2 = |R_mS_m|$, $\Re_3 = |R_mS|$, and $\Re_4 = |RS_m|$.

ray at the receiving location, ϕ_d, is assumed to be proportional to ϕ_{ini} and can be expressed from the GTD (Refs. 6 and 8) as

$$\phi_d = \phi_{\mathrm{ini}} \cdot D(S \to E_1 \to E_2 \to \cdots \to E_l, R|E_l), \tag{2}$$

where $D(S \to E_1 \to E_2 \to \cdots \to E_l, R|E_l)$ is the complex diffraction coefficient correlated with the diffracting edge E_l, the initial ray to E_l, $S \to E_1 \to E_2 \to \cdots \to E_l$, and the receiver location R. In particular, $D(S \to E_1, R|E_1)$ denotes the diffraction coefficient for a single diffraction and can be simplified as the expression of $D(S,R|E_1)$. It is called the single diffraction coefficient below.

The detailed form of $D(S,R|E_1)$ can be obtained by dividing the singly diffracted field ϕ_d with ϕ_{ini} in Eq. (2), where ϕ_{ini} becomes the direct sound from source and ϕ_d is solved with the Hadden–Pierce solution,[9] which is an exact boundary solution for single diffraction with a point source incidence as shown in Fig. 2(a). Although the Hadden–Pierce solution[9] is only presented for the three-dimensional field, it is used in this paper for both three-dimensional and two-dimensional fields by extracting the free field Green function out of its presented formulas.[9] The deduced single diffraction coefficient is expressed as

$$D(S,R|E_1) = -\frac{\sum_{i=1}^{4} A(\varsigma_i)\cdot F_v(\omega, r_S, r_R, \varsigma_i, \beta)}{\pi \cdot G_f(S|E_1)}, \tag{3}$$

where $G_f(S|E_1)$ denotes the free field Green function in the two-dimensional or the three-dimensional field between two locations S and E_1, indicating the directly incident field at edge E_1, and $F_v(\omega, r_S, r_R, \varsigma_i, \beta)$ is a derived integral

$$F_v(\omega, r_S, r_R, \varsigma_i, \beta) = \int_0^1 I(q)dq, \tag{4}$$

where ω is the angular frequency of the wave, and the parameters r_S and r_R are distances in cross-section plane from edge E_1 to source S and receiver R, respectively. β is the exterior angle of the wedge corresponding to diffracting edge E_1. ς_i are the diffracting turn angles and defined individually as[9] $\varsigma_1 = |\theta_R - \theta_S|$, $\varsigma_2 = 2\beta - |\theta_R - \theta_S|$, $\varsigma_3 = \theta_R + \theta_S$, and $\varsigma_4 = 2\beta - (\theta_R + \theta_S)$, whose constructions are illustrated in Fig. 2(b). The integrand function $I(q)$ in Eq. (4) is[9]

$$I(q) = \begin{cases} e^{jk\Re_i}/\Re_i & \text{for the three-dimensional field} \\ (-j/4)H_0^{\,2}(k\Re_i) & \text{for the two-dimensional field,} \end{cases} \tag{5}$$

where k is the wave number and $j = \sqrt{-1}$.

The parameter $\Re_i$ is defined as distance between two points where the turn angle anticlockwise encircling the diffracting edge from one point to another is ς_i, which is illustrated in Fig. 2(b) and depends on the integrant q in Eq. (4) by[9]

$$\Re_i = [L^2 + r_R r_S (Y - Y^{-1})^2]^{1/2}, \tag{6}$$

in which

$$Y = \left[\frac{\tan(A(\varsigma_i)) + \tan(qA(\varsigma_i))}{\tan(A(\varsigma_i)) - \tan(qA(\varsigma_i))}\right]^{\beta/(2\pi)}. \tag{7}$$

$A(\varsigma_i)$ is an angular function and can be expressed as

$$A(\varsigma_i) = \frac{\pi}{2\beta}(-\beta - \pi + \varsigma_i) + \pi U(\pi - \varsigma_i) \tag{8}$$

and

$$U(\theta) = \begin{cases} 1 & \text{if } \theta \geq 0 \\ 0 & \text{if } \theta < 0. \end{cases} \tag{9}$$

The quantity L in Eq. (6) is defined as the total distance along the path of diffracted ray from S to edge E_1 and then to R, which equals $r_S + r_R$ in Fig. 2(a).

In particular, when receiver R is located on the shadow boundary or the reflection boundary of E_1 shown in Fig. 2(a) with ς_1 or ς_4, respectively equaling π and then the corresponding $A(\,)$ becoming $\pi/2$, Eq. (4) leads to singularities and cannot be used to calculate $F_v(\,)$ due to the singular values of Y with Eq. (7) and then those of $\Re_1$ or $\Re_4$, respectively, with Eq. (6). Under these situations, $\Re_i$ for Eq. (5) can be calculated by using its geometrical definition as

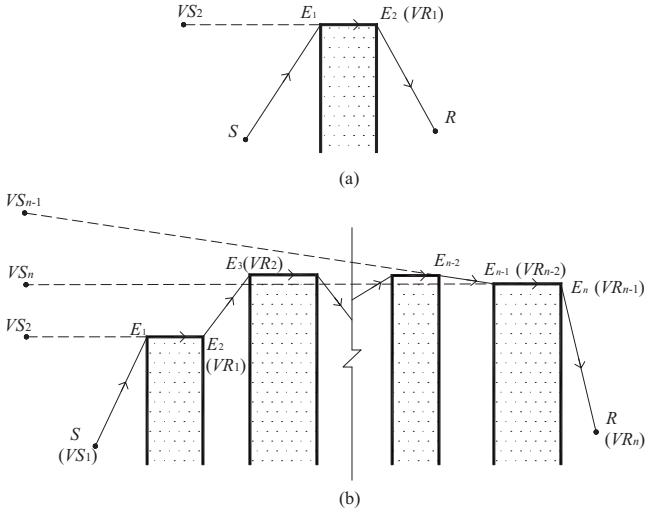


FIG. 3. Illustration of one individual diffracted ray over wide barriers. (a) One generic doubly diffracted ray by edges E_1 and E_2 over a single wide barrier whose width is not less than one wavelength. (b) One generic n-order diffracted ray over several wide barriers from source S to receiver R by edges E_1, E_2, $E_3, \ldots, E_{n-1}$, and E_n in turn. VS_n is a virtual source for edge E_n, VR_{n-1} is a virtual receiver for edge E_{n-1}, and so on.

$$\Re_i = (r_S^{\,2} + r_R^{\,2} - 2r_S \cdot r_R \cdot \cos\varsigma_i)^{1/2}, \tag{10}$$

which results $r_S + r_R$ for $\Re_i$ when $\varsigma_i = \pi$. The function $I(q)$ from Eq. (5) now becomes independent of integrant q in Eq. (4) and accordingly

$$F_v(\omega, r_S, r_R, \pi, \beta) = G_f(r_S + r_R). \tag{11}$$

B. Doubly diffracted ray

In Fig. 3(a), a general doubly diffracted ray is investigated, which is diffracted by E_2 with the initial ray $S \to E_1 \to E_2$. The latter ray can be treated as a singly diffracted ray by edge E_1 with location E_2 being a virtual receiver defined as VR_1 as if the edge E_2 does not exist. Then the field of the ray $S \to E_1 \to E_2$ at edge E_2 can be denoted by $\phi_{d,1}(S, VR_1|E_1)$. Following Eq. (2), the field at receiver R by doubly diffracted ray $S \to E_1 \to E_2 \to R$, $\phi_{d,2}(S, R|E_1, E_2)$, can be determined by

$$\phi_{d,2}(S,R|E_1,E_2) = \phi_{d,1}(S,VR_1|E_1)\cdot D(S \to E_1 \to E_2, R|E_2), \tag{12}$$

where $D(S \to E_1 \to E_2, R|E_2)$ is the diffraction coefficient correlated with the diffracting edge E_2, the initial ray $S \to E_1 \to E_2$, and receiver R. Meanwhile the singly diffracted field $\phi_{d,1}(S, VR_1|E_1)$ can also be obtained with Eq. (2) as

$$\phi_{d,1}(S,VR_1|E_1) = G_f(S|E_1)\cdot D(S,VR_1|E_1), \tag{13}$$

where the initial field at edge E_1 is the direct sound field, $G_f(S|E_1)$.

From the GTD (Refs. 6–8) and Pierce's ray-based approach,[5] the generic diffraction coefficient in Eq. (2) can be approximately evaluated with the specific one for a single diffraction in Eq. (3). For the ray $S \to E_1 \to E_2 \to R$, the diffraction by edge E_2 can be viewed as a single diffraction by assuming that the sides of wedge E_2 become semi-infinite

and the initial ray comes from a virtual source VS_2 illustrated in Fig. 3(a). Here VS_2 locates on the reverse extension line of $E_1 \rightarrow E_2$ and is apart from E_2 for a distance equaling the total length of the initial ray $S \rightarrow E_1 \rightarrow E_2$. Then $D(S \rightarrow E_1 \rightarrow E_2, R|E_2)$ in Eq. (12) can be determined as

$$D(S \rightarrow E_1 \rightarrow E_2, R|E_2) = D(VS_2, R|E_2) \cdot \alpha(E_1, E_2), \quad (14)$$

where $D(VS_2, R|E_2)$ is a single diffraction coefficient and can be calculated with Eq. (3). $\alpha(E_1, E_2)$ is a weighting factor introduced to avoid the redundant counting of the reflection on the connecting side between two edges E_1 and E_2 and is unit if E_2 is separated from E_1. In Fig. 3(a) the weighting factor equals 1/2,[5,10] where E_2 and E_1 are connected with a side whose width is greater than one wavelength. Further details for determining the weighting factor can be found in Ref. 10, which developed an interpolation method to determine the weighting factor for two successive arbitrarily spaced and connected edges.

Thus the field at R delivered by the given doubly diffracted ray, $\phi_{d,2}(S, R|E_1, E_2)$, can be rewritten by substituting Eqs. (13) and (14) into Eq. (12) as

$$\phi_{d,2}(S, R|E_1, E_2) = G_f(S|E_1) \cdot D(S, VR_1|E_1) \cdot D(VS_2, R|E_2) \cdot \alpha(E_1, E_2), \quad (15)$$

which is a product of the direct sound field at the first diffracting edge E_1, the diffraction coefficients, and the weighting factor correlated with the two edges.

C. Generic equations for the individual *n*-order diffracted ray

Similarly, $\phi_{d,n}(S, R|E_1, E_2, E_3, \ldots, E_{n-1}, E_n)$, the sound field of a generic n-order diffracted ray over several wide barriers shown in Fig. 3(b), can be evaluated by multiplying the direct sound field at edge E_1 with the diffraction coefficients and weighting factors at the n diffracting edges,

$$\phi_{d,n}(S, R|E_1, E_2, E_3, \ldots, E_{n-1}, E_n) = G_f(S|E_1) \cdot \prod_{l=1}^{n} D(VS_l, VR_l|E_l) \cdot \alpha(E_{l-1}, E_l), \quad (16)$$

where

$$\alpha(E_{l-1}, E_l) = \begin{cases} 1 & \text{if the edges } E_{l-1} \text{ and } E_l \text{ are separated} \\ 1/2 & \text{if the edges } E_{l-1} \text{ and } E_l \text{ are connected.} \end{cases} \quad (17)$$

In Eq. (16), VR_l denotes the virtual receiver from edge E_l as illustrated in Fig. 3(b), which is the location of E_{l+1}, the next edge along the diffraction ray path. VS_l denotes the virtual source to edge E_l and locates on the reverse extension line of $E_{l-1} \rightarrow E_l$, apart from E_l for a distance equaling the total length of the ray $S \rightarrow E_1 \rightarrow E_2 \rightarrow \cdots \rightarrow E_l$. In fact VS_l, VR_l, and edge E_l construct a complete geometry for a single diffraction at wedge E_l. Particularly, VS_1 represents the location of source S, VR_n represents the location of receiver R, and $\alpha(E_0, E_1) \equiv 1$.

When two or more neighboring barriers have same height, the barriers' configuration in Fig. 3(b) becomes equivalent to that in Fig. 1(b) and 1(c). This causes that the virtual receiver VR_l correlated with edge E_l locates on the shadow boundary or reflection boundary of VS_l for some diffraction rays. For example, in the propagation of ray $S \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow R$ in Fig. 1(c), the virtual receiver VR_2 (edge 3) locates on the shadow boundary of initial ray $S \rightarrow 1 \rightarrow 2$ to edge 2. Under such a situation, the integral term in $D(VS_l, VR_l|E_l)$ can be evaluated with Eq. (11) to avoid singularities. Additionally, receiver R may be located on edge E_n in Fig. 3(b), which causes that $r_R = 0$ and the parameter θ_R fails to be assigned for evaluating $D(VS_n, VR_n|E_n)$ with Eq. (3). In such case, the n-order diffracted ray actually degrades to one with $(n-1)$ orders, $S \rightarrow E_1 \rightarrow E_2 \rightarrow \cdots \rightarrow E_{n-1} \rightarrow E_n$. Accordingly the field $\phi_{d,n}(S, R|E_1, E_2, E_3, \cdots, E_{n-1}, E_n)$ can be replaced with $\phi_{d,n-1}(S, E_n|E_1, E_2, E_3, \cdots, E_{n-1})$, which avoids the evaluation difficulty from $r_R = 0$. There is no other limit of r_R and θ_R for diffraction coefficient evaluation with Eq. (3) and the field at arbitrary receiver locations can be explicitly calculated with Eq. (13), even when receivers are quite close to the diffracting edge compared with the wavelength.

It is worth noting that the method proposed in Eq. (16) depends on the assumption from Eqs. (14) and (17) that the edge-edge distances, $|E_l E_{l+1}|$, are greater than one wavelength. That is, in principle, the proposed method works as well as the GTD with the edge-edge distances larger than the wavelength.

The proposed method is validated with numerical simulations to investigate its accuracy and applicability. The presentation of results is facilitated with insertion loss (IL), which is defined as

$$\text{IL} = 20 \log_{10}(|P_{\text{tot},0}|/|P_{\text{tot},t}|), \quad (18)$$

where $P_{\text{tot},0}$ is sound pressure in the total field at receiver without the barriers while $P_{\text{tot},t}$ is the one with the barriers.

III. RESULTS AND DISCUSSIONS

When there is only a single wide barrier or double parallel knife edges, which are discussed abundantly in the previous studies,[5,10–12,14,15] the method of Eq. (16) reduces to a double-edge form of Eq. (15). Preliminary numerical comparisons in such cases between the proposed method and the previous models, such as the models of Pierce,[5] Chu *et al.*,[10] Kawai or Kim *et al.*,[11,12] and Wadsworth *et al.*,[15] have been

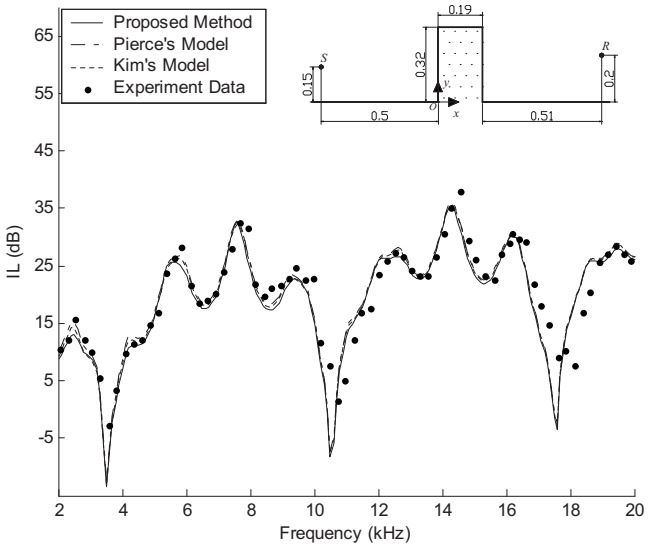


FIG. 4. The IL spectra of a rigid single barrier with width of 0.19 m and height of 0.32 m in the three-dimensional field, where a point source is located at S (−0.5 m, 0.15 m, 0 m) and receiver is R (0.7 m, 0.2 m, 0 m). The unit of the illustrated geometry in the inset figure is meter. The solid line represents the predicted results with the proposed method (proposed method). The dashed-dotted line and dotted line represent the predictions with Pierce's model (Ref. 5) (Pierce's model) and those with the models of Kim *et al.* (Ref. 12) (Kim's model), respectively. The solid points are experimental data from study by Wadsworth and Chambers (Ref. 15) (experimental data).

carried out. Only the results from a three-dimensional case with a rigid single wide barrier are presented in Fig. 4 for succinctness. The corresponding inset figure shows the cross-section geometry. In Fig. 4, good agreements are observed among the predicted results and the experimental data, which serve to validate the proposed method for predicting the double-edge diffraction on the other hand. Additionally, computations have been carried out in advance to investigate how weak the rays diffracted more than once by a same edge in comparison with the rays diffracted only once at each edges. In this case, energy amplitudes of rays $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$, $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$, and $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$ are compared to that of the ray $S\rightarrow1\rightarrow2\rightarrow R$. Figure 5 presents the corresponding results of the energy magnitude ratio, where the wavelength at frequency 1820 Hz equals the barrier width. From Fig. 5, the rays diffracted for two, three, and four times at a same edge are, respectively, weaker than the ray diffracted only once at each edge by 30, 60, and 90 dB at least. Although the magnitude ratios increase a little when the barrier width is smaller than one wavelength, the rays diffracted twice or more by a same edge are sufficiently weak to be neglected in the proposed method.

For the current problem of several wide barriers with some neighboring ones of equal height, since it is hard to use the previous analytical models to calculate the sound field, the boundary element method (BEM) is employed for numerical validation.[12,14,17] To ensure high numerical accuracy, discretization in the BEM is executed by quintic boundary elements with the largest length smaller than one fifteenth of the considered smallest wavelength.

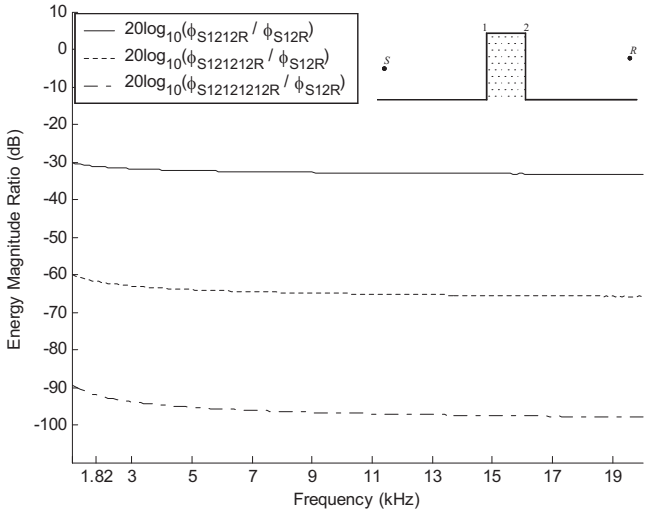


FIG. 5. The spectra of energy magnitude radio of the rays $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$, $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$, and $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$ that are diffracted twice or more by a same edge, compared to the ray $S\rightarrow1\rightarrow2\rightarrow R$ that is diffracted only once at each edge. The location of source, receiver, and two edges are illustrated in the inset figure. The parameter ϕ_{S1212R} represents the sound pressure amplitude of ray $S\rightarrow1\rightarrow2\rightarrow1\rightarrow2\rightarrow R$, ϕ_{S12R} represents that of ray $S\rightarrow1\rightarrow2\rightarrow R$, and so on.

Two numerical cases whose geometries correspond to those in Figs. 1(c) and 1(b), respectively, are investigated with typical dimensions of barriers in the two-dimensional field. In the first case shown in the inset figures of Figs. 6 and 7, two parallel rigid barriers with identical width of 0.6 m and the same height of 2.4 m are spaced from 1 m on the infinite rigid ground. A coherent line source parallel to lengthwise axis is defined and located at S (−3.2 m, 0.4 m), while receivers are chosen at R (2.37 m, 2.3 m) in Fig. 6 apart from the nearest edge for 0.2 m (equaling wavelength at frequency 1720 Hz) and at (5.14 m, 0.5 m) in Fig. 7 apart

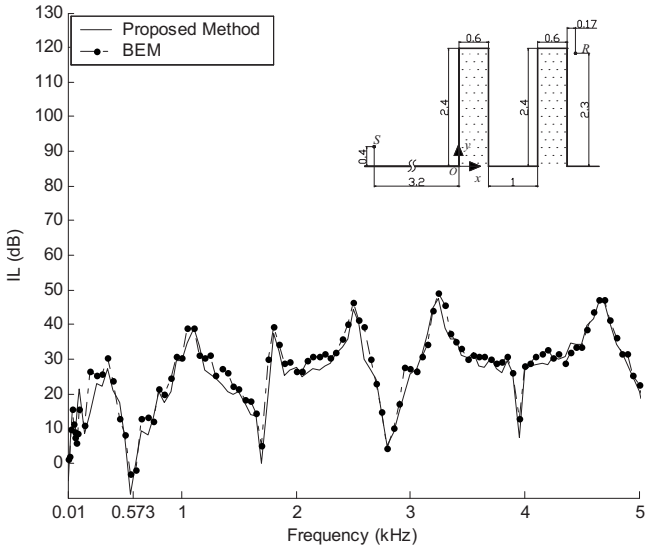


FIG. 6. The spectra of IL at receiver R (2.37 m, 2.3 m), which is in the shadow zone of source S (−3.2 m, 0.4 m) due to two rigid barriers with a same height of 2.4 m blocking the incidence. The barriers have identical width of 0.6 m and are spaced for 1 m. The unit of the illustrated geometry in the inset figure is meter. The solid line represents the predicted results with the proposed method (proposed method) and the dashed-dotted line represents the numerical results from the BEM.

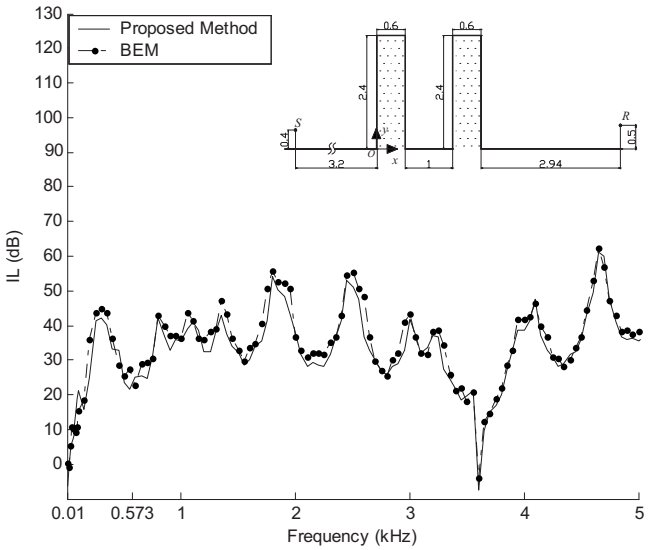


FIG. 7. Same caption as Fig. 6 except that location of R changes to (5.14 m, 0.5 m).

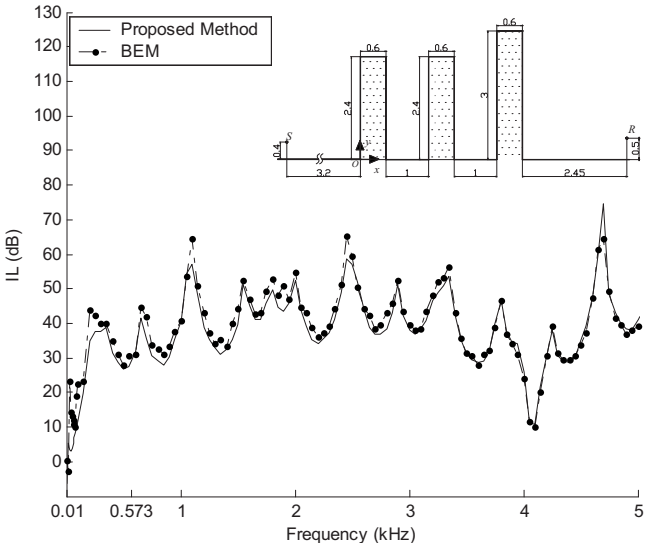


FIG. 9. Same caption as Fig. 8 except that location of R changes to (6.25 m, 0.5 m).

from the nearest edge for 3.5 m (equaling wavelength at frequency 98 Hz). Such choice of receiver locations allows the investigation on sound field in different areas of interest, being close or far from the diffracting edges compared to the wavelength.

The second case is shown in the inset figures of Figs. 8 and 9 with three barriers, where barrier widths remain 0.6 m and the barrier-barrier spaces remain 1 m. In this case, two neighboring barriers have same height of 2.4 m while the other is 3 m high. The definition of source is same as the first case but the receiver locations change to R (3.97 m, 2.9 m) in Fig. 8 and to (6.25 m, 0.5 m) in Fig. 9 based on the same relevant consideration in the first case. In both cases, receivers are in the shadow zone of source due to barriers blocking and either source or receivers can only observe the nearest edge, respectively.

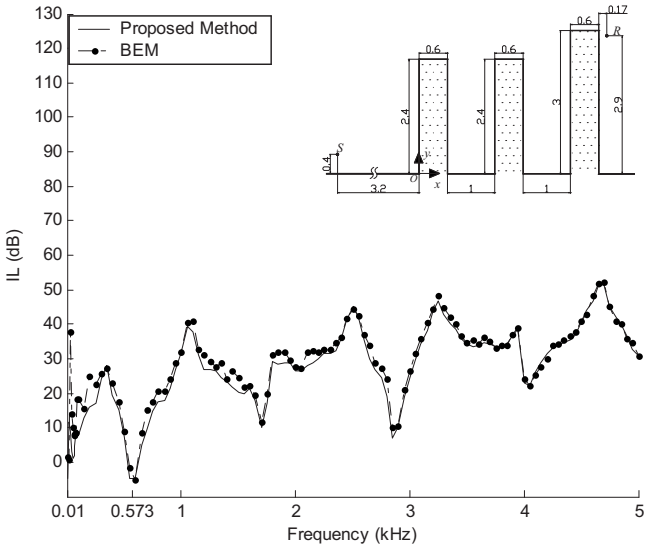


FIG. 8. The spectra of IL at receiver R (3.97 m, 2.9 m) in the shadow zone of source S (−3.2 m, 0.4 m) due to three rigid barriers blocking the incidence, where heights of two barriers are 2.4 m and the other is 3 m high. The barriers have identical width of 0.6 m and are spaced for 1 m one by one. The unit of the illustrated geometry in the inset figure is meter. The solid line represents the predicted results with the proposed method (proposed method) and the dashed-dotted line represents the numerical results from the BEM.

In the case with two barriers, the maximum diffraction order is 4 and the total number of diffracted rays reaching R is 28 considered with the proposed method. The corresponding evaluated IL spectra are shown in Figs. 6 and 7 with receiver locations (2.37 m, 2.3 m) and (5.14 m, 0.5 m), respectively. Here the minimum edge-edge distance is 0.6 m equaling the wavelength at frequency 573 Hz. From Figs. 6 and 7, over the frequencies range from one to eight times larger than 573 Hz, the predictions from the proposed method agree well with those from the BEM, except small discrepancies at some frequencies, whose reasons are not completely clear yet. Moreover, it is found in Figs. 6 and 7 that at frequencies around 573 Hz the agreement between the results with these two methods remains good. The IL curves in Figs. 6 and 7 are quite complex with large fluctuations over the broad frequency range, because of interference between the different waves diffracted by the barriers and reflected from the ground.[19]

Figures 8 and 9 show the corresponding IL spectra for the case of three barriers with receiver locations being (3.97 m, 2.9 m) and (6.25 m, 0.5 m), respectively, where the maximum diffraction orders become 6 and the total number of the diffracted rays reaching R increases up to 116. In this case, the minimum edge-edge distance remains 0.6 m, which is the wavelength at 573 Hz. In Figs. 8 and 9, over the frequency range higher than 573 Hz, the results with the proposed method and those with the BEM are in good agreement again, except small discrepancies at some frequencies. Additionally, at the frequencies around 573 Hz in Figs. 8 and 9, the agreement between the predictions with these two methods is found to be good also.

The computational times with these two methods in both cases are compared. A total of 9800 quintic elements are considered at the highest frequency of 5 kHz in the BEM for both numerical cases. And it takes over 1 h by a personal

computer with a 2.4 GHz Intel Q6600 processor and 4 Gbytes of random access memory to execute the BEM evaluation at such single frequency. The evaluation with the proposed method takes only 1.4 min in the first case and 7.9 min in the second case on the same computer for a single frequency and the corresponding computational time is frequency independent. This indicates that for an equivalent accuracy degree, the proposed method is much faster than the BEM.

From the above results, the proposed method can evaluate the multiple acoustic diffraction over wide barriers more efficiently than the BEM. As a ray-based method, the accuracy of the method for evaluating the individual diffracted rays depends on the geometry dimensions compared to the wavelength, which are edge-edge distances in the current problem. The results of the numerical simulations show that the method is accurate when the edge-edge distances are larger than one wavelength. It is also found that the method remains accurate even when the edge-edge distances become a little less than the wavelength. Furthermore, though the method in Eq. (16) is proposed for wide barriers, it can be used in principle to evaluate the individual multiply diffracted rays around parallel knife edges also, for example, around the rib-like structure studied by Bougdah *et al.*[18]

IV. CONCLUSION

In this paper, a ray-based method is developed to solve the multiple acoustic diffraction around parallel wide barriers with some neighboring ones of equal height. The method is based on Keller's GTD (Refs. 6–8) and extended from Pierce's exact boundary solution[5] for a rigid wedge. The proposed method can avoid singularities while solving multiple diffraction along the shadow boundaries or the reflection boundaries. Numerical simulations show that the method can predict the field at arbitrary receiver locations.

The accuracy and applicability of the method are validated numerically with the BEM in the two-dimensional field where the model was shown to be considerably accurate when the edge-edge distances are larger than one wavelength. The proposed method has more computational efficiency than the BEM and is useful for predicting acoustic diffraction along arbitrary directions or at arbitrary receiver locations around parallel barriers with various configurations.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for their constructive comments on improving the original manuscript. Projects 10674068 and 10604030 supported by NSFC.

[1]K. Fujiwara, Y. Ando, and Z. Maekawa, "Attenuation of a spherical sound wave diffracted by a thick plate," Acustica **28**, 341–347 (1973).
[2]U. J. Kurze and G. S. Anderson, "Sound attenuation by barriers," Appl. Acoust. **4**, 35–53 (1971).
[3]D. A. Bies and C. H. Hansen, *Engineering Noise Control: Theory and Practice*, 2nd ed. (E & FN SPON, London, 1996).
[4]RAYNOISE user's manual, version 3.1 LMS, Numerical Technologies (2002).
[5]A. D. Pierce, "Diffraction of sound around corners and over wide barriers," J. Acoust. Soc. Am. **55**, 941–955 (1974).
[6]J. B. Keller, "Diffraction by an aperture," J. Appl. Phys. **28**, 426–444 (1957).
[7]J. B. Keller, "Diffraction by an aperture. II," J. Appl. Phys. **28**, 570–579 (1957).
[8]J. B. Keller, "Geometrical theory of diffraction," J. Opt. Soc. Am. **52**, 116–130 (1962).
[9]W. J. Hadden and A. D. Pierce, "Sound diffraction around screens and wedges for arbitrary point source location," J. Acoust. Soc. Am. **69**, 1266–1276 (1981).
[10]D. Chu, T. K. Stanton, and A. D. Pierce, "Higher-order acoustic diffraction by edges of finite thickness," J. Acoust. Soc. Am. **122**, 3177–3193 (2007).
[11]T. Kawai, "Sound diffraction by a many-sided barrier or pillar," J. Sound Vib. **79**, 229–242 (1981).
[12]H. S. Kim, J. S. Kim, H. J. Kang, B. K. Kim, and S. R. Kim, "Sound diffraction by multiple wedges and thin screens," Appl. Acoust. **66**, 1102–1119 (2005).
[13]R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," Proc. IEEE **62**, 1448–1461 (1974).
[14]E. M. Salomons, "Sound propagation in complex outdoor situations with a non-reflecting atmosphere: Model based on analytical solutions for diffraction and reflection," Acust. Acta Acust. **83**, 436–454 (1997).
[15]G. J. Wadsworth and J. P. Chambers, "Scale model experiments on the insertion loss of wide and double barriers," J. Acoust. Soc. Am. **107**, 2344–2350 (2000).
[16]H. Medwin, E. Childs, and G. M. Jebsen, "Impulse studies of double diffraction: A discrete Huygens interpretation," J. Acoust. Soc. Am. **72**, 1005–1013 (1982).
[17]E. Reboul, A. L. Bot, and J. P. Liaudet, "Radiative transfer equation for multiple diffraction," J. Acoust. Soc. Am. **118**, 1326–1334 (2005).
[18]H. Bougdah, I. Ekici, and J. Kang, "A laboratory investigation of noise reduction by riblike structures on the ground," J. Acoust. Soc. Am. **120**, 3714–3722 (2006).
[19]D. Duhamel, "Efficient calculation of the three-dimensional sound pressure field around a noise barrier," J. Sound Vib. **197**, 547–571 (1996).

Annoyance from environmental noise across the lifespan

Pascal W. M. Van Gerven[a)]
Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Henk Vos
Department of Environment and Health, Netherlands Organization for Applied Scientific Research (TNO), P.O. Box 49, 2600 AA Delft, The Netherlands

Martin P. J. Van Boxtel
Department of Psychiatry and Neuropsychology, Faculty of Health, Medicine, and Life Sciences, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Sabine A. Janssen and Henk M. E. Miedema
Department of Environment and Health, Netherlands Organization for Applied Scientific Research (TNO), P.O. Box 49, 2600 AA Delft, The Netherlands

(Received 21 October 2008; revised 23 April 2009; accepted 12 May 2009)

Curvilinear effects of age on self-reported annoyance from environmental noise were investigated in a pooled international and a Dutch sample of in total 62,983 individuals aged between 15 and 102 years. All respondents were frequently exposed to varying levels of transportation noise (i.e., aircraft, road traffic, and railway noise). Results reveal an inverted U-shaped pattern, where the largest number of highly annoyed individuals was found in the middle-aged segment of the sample (peaking around 45 years) while the lowest number was found in the youngest and oldest age segments. This pattern was independent of noise exposure level and self-reported noise sensitivity. The inverted U-shape explains the absence of linear age effects in previous studies. The results are discussed in light of theories predicting an age-related vulnerability to noise.
 [DOI: 10.1121/1.3147510]

PACS number(s): 43.50.Qp, 43.50.Sr, 43.50.Lj, 43.50.Rq [BSF] Pages: 187–194

I. INTRODUCTION

Auditory noise is ubiquitous in densely populated and industrialized societies. Research has shown that environmental noise may have serious adverse effects on cognition and health (e.g., Evans *et al.*, 1995; Stansfeld *et al.*, 2005). An important index of these effects is annoyance (e.g., Guski *et al.*, 1999; Ouis, 2001). The central question of this cross-sectional study is whether annoyance from environmental noise varies as a function of age.

Age is often considered as a vulnerability factor with regard to the harmful effects of noise. The World Health Organization (WHO) considers children and older adults to be at particular risk (e.g., Hygge, 2003). Empirical findings, however, are not necessarily consistent with this viewpoint. Enmarker and Boman (2004), for example, found school pupils to report less annoyance from aircraft noise than their teachers. Further results that are inconsistent with the WHO's viewpoint can be found at the other end of the age spectrum. Beaman (2005), for example, found older people to report less annoyance from irrelevant background speech than their younger counterparts. Similar results are reported by Michaud *et al.* (2005, 2008), who found people aged 65 and over to be more likely to report no annoyance at all from road traffic noise, whereas people aged 25–44 were more likely to report high levels of annoyance. The reason for this inconsistency might be that, apart from age-related vulnerability, the relation between age and noise annoyance is mediated by multiple psychological, social, and environmental factors, such as trust in the authorities' policy toward noise, residential characteristics, noise sensitivity, and available opportunities to cope with noise (e.g., Fields, 1993; Miedema and Vos, 1999). The relative contribution of these factors might vary across samples and thus lead to mixed, and sometimes unanticipated, results.

Although the role of age as a vulnerability factor has been extensively studied in children and both younger and older adults (e.g., Bluhm *et al.*, 2004; Enmarker and Boman, 2004; Hygge *et al.*, 2002; Nivison and Endresen, 1993), only few studies have considered the full lifespan. Moreover, most previous studies were aimed at linear effects of age on noise annoyance, not curvilinear effects. Bluhm *et al.* (2004), for example, examined annoyance from road traffic noise among 1000 individuals ranging in age between 19 and 80 years, but did not find a linear relationship between age and level of annoyance. In a smaller but comparable study, Nivison and Endresen (1993) investigated determinants of noise-induced annoyance among 94 individuals aged between 19 and 78 years, but they also did not find a linear relationship between age and annoyance. These studies do not exclude an effect of age on noise annoyance, however. In fact, the absence of a linear effect might be fully consistent with the existence of a curvilinear effect.

[a)]Author to whom correspondence should be addressed. Electronic mail: p.vangerven@maastrichtuniversity.nl

To detect a possible curvilinear effect, we analyzed a large pooled international and a smaller Dutch sample of—in total—over 60 000 individuals, ranging in age between 15 and 102 years, who were frequently exposed to aircraft, road traffic, and railway noise. For that purpose, we statistically modeled the effects of both age and average noise exposure level [in dB(A)] on self-reported annoyance. Furthermore, we investigated whether possible effects of age on annoyance are mediated by age-related differences in self-reported noise sensitivity.

II. DATA

Since 1990, researchers at the Netherlands Organization for Applied Scientific Research (TNO) in Delft have compiled an archive of original datasets from studies on annoyance caused by transportation noise. These studies concerned different modes of transportation—aircraft, road traffic, and railway—and were carried out in Europe, North America, Japan, and Australia. As far as possible, a common set of variables was derived from these datasets, which includes noise exposure, noise annoyance, and noise sensitivity measures. In addition, demographic variables, such as gender, education, and age, were available. Table I provides an overview of the datasets analyzed in the current study (for further details, see Miedema and Vos, 1998). Shown are the minimum, maximum, and mean age of the respondents in the different studies. As can be seen, age ranges between 15 and 102 years.

In addition to the pooled international dataset (total $N=51\,987$), we analyzed a large aircraft noise dataset—coded NET-371—from a study conducted around Amsterdam Airport Schiphol ($N=10\,996$). A description of this dataset is given by TNO-PG (TNO Prevention and Health) and RIVM (National Institute for Public Health and the Environment) (TNO-PG and RIVM, 1998). The percentage of highly annoyed persons found in NET-371 at a given noise exposure level was about four times higher than the percentage found in earlier aircraft noise studies. This may be due to local circumstances, in particular, the intense public debate about the future expansion of Amsterdam Airport Schiphol, possibly enhancing the role of non-acoustic (e.g., personal and social) influences on annoyance. In addition, there appears to be a general trend toward a higher level of annoyance at the same level of noise exposure (Guski, 2004; Van Kempen and Van Kamp, 2005). Considering the large size of the Schiphol dataset, it was expected to have a substantial impact on the results. Therefore, NET-371 was analyzed separately. In this way, we were also able to show whether the effect of age on noise annoyance is sensitive to circumstances such as public debate on noise issues.

A. Noise exposure measures

Previous synthesis studies (e.g., Miedema and Vos, 1999) used L_{dn} or L_{den} as the descriptor of noise exposure. L_{dn} is defined in terms of L_{Aeq} (average equivalent noise level) during daytime (d) and nighttime (n), and applies a 10 dB(A) penalty to noise in the night:

$$L_{dn} = 10\log((15/24)\cdot 10^{LD/10} + (9/24)\cdot 10^{(LN+10)/10}).$$

In this formula, LD and LN are long-term L_{Aeq}'s as defined by the ISO (2002) for the day (7:00 a.m.–10:00 p.m.) and the night (10:00 p.m.–7:00 a.m.), respectively. The new metric for noise mapping in Europe is L_{den}, which includes an evening level (e, 7:00 p.m.–11:00 p.m.). Evening noise level is not available for all datasets in the TNO archive, however. Confining the analysis to those data that do contain L_{den} would reduce the total number of cases to 47 626, which is 76% of the total sample. To maximize the number of cases in the analysis, we used L_{dn} as the exposure measure. The results were not expected to be influenced by this choice, because L_{dn} and L_{den} are highly correlated.

People in areas with low exposure levels [<45 dB(A)] were excluded from the analyses, because the assessment of these low levels is relatively inaccurate. Moreover, in situations with low exposure levels, sources of noise other than transportation, which were not available in this study, may be more important. Also, people exposed to very high noise levels [>75 dB(A)] were excluded, because in areas with very high levels of noise there may be self-selection of people that are not affected by noise. Although there is no direct evidence for such self-selection, noise-sensitive people seem to be more prepared to move to a different area (Nijland *et al.*, 2007).

B. Annoyance measures

Annoyance questions in the different datasets did not contain the same number of response categories. For example, some questions had only four response categories, whereas others had as many as 11 categories. In this study, all sets of response categories were converted into a scale ranging from 0 to 100. This conversion is based on the assumption that a set of categories divides the range of 0 to 100 in equally spaced intervals. The general rule that gives the position of an inner category boundary on the scale of 0 to 100 is $\text{score}_{\text{boundary } i} = 100i/m$, where i is the rank number of the category boundary, starting from 1 for the upper boundary of the lowest category, and m is the number of categories. The distribution of the annoyance scores at a given noise exposure level can be summarized in different ways. Often, the percentage of responses exceeding a certain cutoff point on the scale is reported. If the cutoff is 72 on a 0–100 scale, then the result is called the percentage of "highly annoyed" persons (%HA). Likewise, a cutoff of 50 indicates the percentage of "annoyed" (%A) and a cutoff of 28 the percentage of "at least a little annoyed" persons (%LA). Although annoyance was treated as a continuous dependent variable in our exposure-effect model, we selected %HA as its primary indicator.

C. Noise sensitivity measures

Operational definitions of noise sensitivity are based on different types of questions. Respondents in social surveys (e.g., Langdon, 1976; McKennell, 1963) or participants in experiments (e.g., Ellermeier and Zimmer, 1997; Stansfeld,

TABLE I. Datasets included in the analyses.

			N per noise source				Age (years)		
	Code[a]	Country	Air	Road	Rail	Total	Min.	Max.	Mean
Pooled international dataset									
	USA-022	United States	2 208			2 208	15	75	44
	UKD-024	United Kingdom	3 843			3 843	23	80	48
	USA-032	United States	1 535			1 535	25	75	47
	USA-044	United States	1 607			1 607	25	75	46
	UKD-071	United Kingdom		2 058		2 058	20	69	48
	UKD-072	United Kingdom		901		901	18	88	46
	USA-082	United States	362			362	18	88	47
	FRA-092	France		876		876	15	65	39
	NET-106	The Netherlands		418		418	19	85	47
	UKD-116	United Kingdom			1,073	1 073	18	91	48
	CAN-120	Canada		1 108		1 108	15	65	42
	CAN-121	Canada		1 070		1 070	15	93	38
	NET-153	The Netherlands			602	602	17	61	41
	UKD-157	United Kingdom		301		301	20	69	50
	CAN-168	Canada	617	556		1 173	15	88	38
	SWI-173	Switzerland		1 215		1 215	24	73	50
	GER-192	Germany		1 570	1,559	3 129	17	80	44
	USA-203	United States	564			564	25	60	40
	AUL-210	Australia	3 202			3 202	23	74	43
	UKD-238	United Kingdom	588	520		1 108	18	66	45
	FRA-239	France	564	523		1 087	18	66	41
	NET-240	The Netherlands	573	473		1 046	18	65	44
	UKD-242	United Kingdom	1 979	408		2 387	17	94	46
	NET-258	The Netherlands		302		302	21	97	48
	NET-276	The Netherlands		697	265	962	18	89	50
	NOR-311	Norway	1 369			1 396	15	98	45
	NOR-328	Norway	673			673	16	88	48
	AUS-329	Australia		784		784	24	70	42
	ITL-350	Italy		876		876	19	90	46
	NET-361	The Netherlands		788	69	857	16	99	48
	NET-362	The Netherlands		293		293	19	67	50
	FRA-364	France		824		824	15	93	44
	SWE-365	Sweden			2,802	2 802	18	76	47
	NOR-366	Norway	321			321	15	87	44
	TRK-367	Turkey		123		123	20	70	43
	SWE-368	Sweden		1 234		1 234	18	75	49
	JPN-369	Japan		799		799	20	80	46
	JNP-370	Japan			435	435	21	80	50
	GER-372	Germany		539		539	15	65	47
	GER-373	Germany		428		428	15	65	47
	GER-374	Germany		572		572	15	87	44
	JNP-382	Japan		729		729	20	78	49
	NET-522	The Netherlands	741			741	18	89	48
	GER-523	Germany		841	707	1 548	17	82	44
	GER-526	Germany		941		941	17	91	46
	SWE-529	Sweden		917		917	17	102	42
Total/mean			20 773	23 702	7,512	51 987			45
Schiphiol dataset	NET-371	The Netherlands	10 996			10 996	17	98	48
Grand total/mean			31 769	23,702	7,512	62 983			45

[a]Codes refer to Fields's (2001) updated catalog.

1992) are classified with respect to noise sensitivity on the basis of (1) self-reported characterizations of their noise sensitivity, (2) self-reported general attitudes toward noise, or (3) self-reported reactions to noise in specific situations. Noise sensitivity (*Sensi*) of the first type was available for part of the datasets in the present study, although questions in the different datasets did not contain the same number of response categories. For example, some questions had only three response categories, whereas others had as many as eight categories. In this study, we converted all sets of re-

sponse categories into a scale ranging from 0 to 100 in the same way as we did for annoyance (see Sec. II B).

III. EXPOSURE-EFFECT MODEL

We used an existing statistical model to analyze the relationship between L_{dn} and annoyance (Groothuis-Oudshoorn and Miedema, 2006; Miedema and Oudshoorn, 2001). In this model, the percentage of highly annoyed individuals (%HA) is calculated as follows on the basis of a given L_{dn}:

$$P_C(L_{dn}) = 100 \cdot (1 - \Phi((C - [\beta_0 + \beta_1 L_{dn} + \Sigma_i \beta_i X_i])/\sigma)),$$

where $P_C(L_{dn})$ is the percentage of persons exposed to L_{dn} with an annoyance score (0–100) above cutoff point C (72 in the present model), Φ is the cumulative standard normal distribution, X_i indicates additional predictors apart from noise exposure level (i.e., age and noise source) and their interactions, and β_0, β_1, β_i, and σ are model parameters.

The model was applied to both the pooled international and the Schiphol dataset. We used a multilevel model to take methodological differences between the datasets of the different studies into account. This "study effect" was treated as random error. The individual error and the study effect were assumed to be independent and randomly distributed. The distributions of the error components were assumed to be normal with zero mean. In our model, the standard deviation was defined as $\sigma = \sqrt{\sigma_0^2 + \sigma_1^2}$, where σ_0^2 and σ_1^2 represent the variance of the study effect and the individual error, respectively. The parameter β_0 represents the intercept and β_1 represents the slope, which describes the change of self-reported annoyance as a function of L_{dn}. Additional variables are represented in the model by X_i and the parameters β_i indicate how annoyance increases as a function of these variables.

On the basis of earlier findings (Groothuis-Oudshoorn and Miedema, 2006; Miedema and Oudshoorn, 2001; Miedema and Vos, 1998), it was anticipated that the relationships between L_{dn} and annoyance were different for the different types of transportation noise in the international dataset. To allow for these differences, dummy variables were included in the model for each type of noise (i.e., aircraft, road traffic, and railway noise). Furthermore, we allowed not only for a different intercept but also for a different slope by also including the product of the dummies and L_{dn}. Taking road traffic as the reference source, dummy variables Air and Rail and their products with L_{dn} (i.e., Air $\times L_{dn}$ and Rail $\times L_{dn}$) were entered as additional interaction terms. Consequently, the intercept (β_0) and slope (coefficient of L_{dn}) pertain to road traffic noise, the coefficients of Air and Air $\times L_{dn}$ are the "adjustments" that must be applied to obtain the intercept and the slope for aircraft noise, and the coefficients of Rail and Rail $\times L_{dn}$ are the adjustments that must be applied to obtain the intercept and the slope for railway noise.

The other major variables that were entered in the model were Age and Age^2. In the case of a U-shaped relationship between age and annoyance, the age at which the effect is minimal (or maximal) was computed by equating the derivative of the curvilinear effect of age to zero and then solving for Age. Thus, the minimum (or maximum) annoyance occurs at the following age: $-100 \cdot (b_{\text{Age}}/2 \cdot b_{\text{Age}^2})$, where b_{Age} is the parameter estimate (coefficient) for Age and b_{Age} is the parameter estimate for Age^2. The original model also included demographic variables, such as gender and education, but since these variables yielded little or no effects and did not alter the effect of age, they were left out of the current model. Moreover, because information about education was missing in a considerable number of studies, inclusion of this variable would have reduced the number of valid cases by more than 50%.

TABLE II. Parameter estimates.

Variable/ parameter	Parameter estimate (b)	Standard error (SE)
Pooled international dataset		
Measurement variance s_0=492		
Individual variable s_1=1 218		
Intercept	−122.54[a]	4.37
L_{dn}	2.22[a]	0.04
Air	23.74[a]	3.85
Rail	−7.99[a]	5.14
Air $\times L_{dn}$	0.02	0.06
Rail $\times L_{dn}$	−0.05	0.08
Age/100	83.57[a]	6.17
(Age/100)^2	−95.36[a]	6.48
Max. annoyance at Age=44		
Schiphol dataset		
Individual variance s_1=1772		
Intercept	−2.36.50[a]	8.03
L_{dn}	4.36[a]	0.13
Age/100	254.87[a]	15.44
(Age/100)^2	−265.30[a]	15.21
Max. annoyance at Age=48		

[a] $p < 0.001$.

The parameters β_0, β_1, σ_0, σ_1, and the additional parameters β_i and σ_i corresponding to predictors X_i were estimated on the basis of the data. The dependent variable %HA at a given L_{dn} was determined by the parameter estimates b_0, b_1, b_i, s_0, s_1, and s_i, where cutoff C=72.

A separate model was constructed with Sensi as the dependent variable, using the same predictors as in the model of annoyance. In case of an effect of Age or Age^2 on Sensi, the variable Sensi was included in the model of annoyance in order to investigate possible mediation of the age effect on annoyance. Because noise sensitivity was only measured in part of the studies, the number of cases in the pooled dataset was reduced by almost 50%.

IV. RESULTS

The results of the analysis for air traffic, road traffic, and railway noise in the pooled international sample are presented in the upper part of Table II. We concentrate on the coefficients for Age and Age^2 only, because this is our main interest here (note that for technical reasons the unit of age is years/100). From the analysis and Fig. 1, it appears that %HA as a function of age follows a curvilinear trajectory. At a given noise exposure level, our model predicts self-

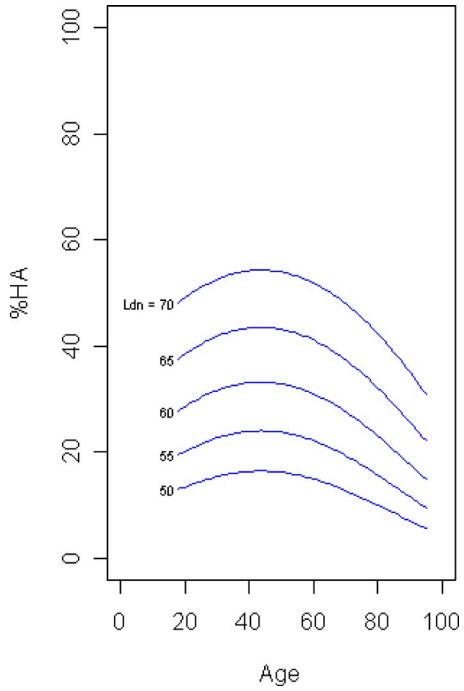


FIG. 1. (Color online) Predicted percentage of highly annoyed persons (%HA) as a function of age and noise level [L_{dn} in dB(A)] for the pooled dataset.

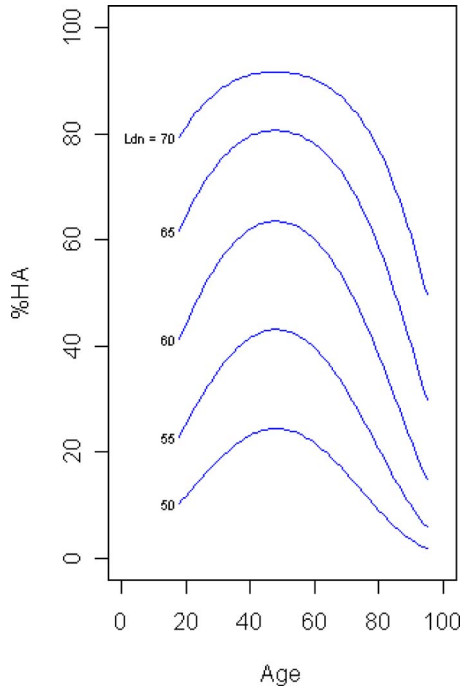


FIG. 3. (Color online) Predicted percentage of highly annoyed persons (%HA) as a function of age and noise level [L_{dn} in dB(A)] for the Schiphol dataset.

reported annoyance to be highest in people in their mid-40s, peaking at the age of 44 years. Figure 2 shows the predicted %HA as a function of noise exposure level. Curves are presented for three selected ages (20, 45, and 80 years). Regardless of the noise exposure level, the 45-year-olds show the highest %HA, the 80-year-olds show the lowest %HA, and the 20-year-olds show an intermediate %HA.

The effect of age on self-reported noise sensitivity (Sensi) shows a similar inverted U-shaped curve (estimate for Age=117.84, SE=7.29, $p<0.001$; estimate for Age^2= −119.54, SE=7.70, $p<0.001$). Nevertheless, even after including Sensi in the model, which leads to a reduced dataset, the effect of age on annoyance is still significant (estimate for Age=69.17, SE=8.07, $p<0.001$; estimate for Age^2= −85.11, SE=8.52, $p<0.001$).

In the lower part of Table II the results of the Schiphol sample are given, which show much larger coefficients for Age and Age^2 than the results for the pooled international dataset. This suggests a much stronger effect of age on annoyance than in the international sample. The pattern of results shown in Figs. 3 and 4 is generally the same, however. That is, similar to the pooled dataset, peak annoyance occurs at middle age, more precisely at the age of 48 years. Furthermore, as can be seen from Fig. 4, it appears that, again similar to the pooled data, the 45-year-olds reveal the highest

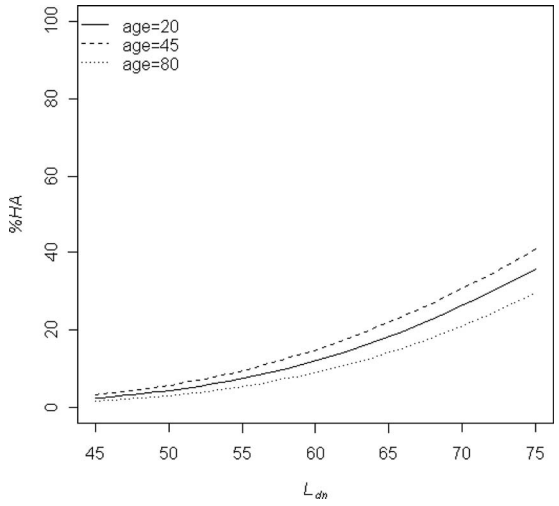


FIG. 2. Predicted dose-response curves [%HA as a function of noise level in dB(A)] of the pooled dataset for three different ages (20, 45, and 80 years). The curves are based on road traffic noise only.

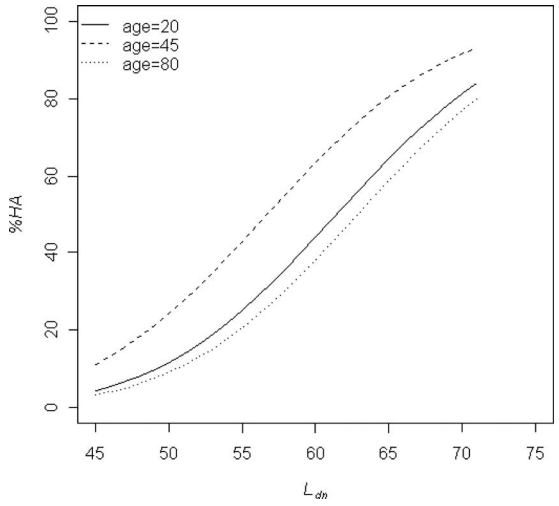


FIG. 4. Predicted dose-response curves [%HA as a function of noise level in dB(A)] of the Schiphol dataset for three different ages (20, 45, and 80 years).

%HA and the 80-year-olds show the lowest %HA. The %HA within the 20-year-olds lies in between, but closer to the 80-year-olds than is the case for the pooled dataset. For a given exposure level, age differences are larger than in the pooled dataset, but the general pattern is the same.

Again, the effect of age on self-reported noise sensitivity (Sensi) shows a similar inverted U-shaped curve as the effect of age on annoyance (estimate for Age=84.73, SE=9.24, $p<0.001$; estimate for Age^2=−79.25, SE=9.11, $p<0.001$). Furthermore, the effect of age on annoyance is still significant after including Sensi in the model (estimate for Age =193.88, SE=14.18, $p<0.001$; estimate for Age^2=−206.19, SE=14.01, $p<0.001$).

In sum, the effect of age on annoyance is described by inverted U-shaped curves. Relatively young as well as relatively old individuals report less annoyance than people of intermediate ages do.

V. DISCUSSION

The curvilinear, inverted U-shaped effect of age on noise annoyance found in the current study is consistent with previous studies, which did not yield significant linear effects of age (e.g., Bluhm *et al.*, 2004; Nivison and Endresen, 1993). Furthermore, the effect is in line with several smaller-scale studies (e.g., Beaman, 2005; Enmarker and Boman, 2004), which found adolescents and older adults to report relatively low levels of annoyance from different sources of noise. Finally, the effect is roughly consistent with a recent study by Michaud *et al.* (2008), who found people aged between 25 and 44 to be more likely to report high levels of annoyance from road traffic noise.

From a theoretical, cognitive developmental perspective, our results are quite remarkable, however. A substantial body of research has shown that the prefrontal cortex and its associated cognitive control functions, such as inhibition, mature relatively late and decline relatively early in life (Blakemore and Choudhury, 2006; Buckner *et al.*, 2006; Dempster, 1992; Hasher *et al.*, 1991). These prefrontal networks also play a crucial role in dealing with background noise while being engaged in a cognitive task (e.g., Gisselgård *et al.*, 2004; Tomasi *et al.*, 2005). Hence, there is ample reason to believe that the susceptibility to noise depends on the development of these prefrontal networks, and thus on age. This implies, however, that the youngest and oldest rather than the middle-aged respondents should report the highest levels of annoyance, which is exactly opposite to the current pattern of results.

There are alternative views, however. One potential determinant of age-related noise annoyance is the average level of mental workload or cognitive challenge a person experiences in daily life. High average workload may limit the ability of a person to adapt to uncontrollable environmental noise. Support for this idea is provided by Wallenius (2004), who investigated the impact of environmental noise and stress from daily cognitive demands on self-perceived health. One of the main findings from this study was that daily cognitive stress and noise annoyance interact in their impact on perceived health. Although annoyance was not a dependent variable in Wallenius's study, there was a moderate correlation (r=0.24) between cognitive stress and annoyance from road traffic noise. This correlation might be in line with theories of noise perception in which it is contended that the level of annoyance from noise depends on the availability of cognitive resources required to adapt to this noise (Fields, 1993; Job, 1988). Availability of these adaptive resources is constrained by the presence of other stressors, such as work and social demands, which all draw on the same cognitive capacity: the fewer resources available, the higher the level of annoyance from noise. In the same vein, Tafalla and Evans (1997) found that although cognitive performance can be maintained in noise by investing more mental resources, this leads to an increased physiological stress response (and possibly a higher level of annoyance, although this was not measured). Cohen (1980), finally, showed that the disturbing effects of a stressor may continue even after the stressor has disappeared. These "after effects" might lead to a continuous lack of resources required to cope with noise and other stressors, and thus to continuously elevated levels of annoyance.

There is reason to believe that the daily level of mental workload varies with age. People obviously face an increasing number of responsibilities from adolescence to early and middle adulthood (e.g., study, work, financial matters, child care, care for older parents, etc.), which decreases again from middle to old age. Among the few studies delineating this everyday workload as a function of age, Lundberg *et al.*, (1994) found different indices of labor-, family-, and household-related workload to peak between the ages of 35 and 39 years in a sample of over one thousand white collar workers (aged 32–58 years). The current results might indicate that people in early to middle adulthood are most susceptible to environmental noise, because they experience the highest levels of daily workload and thus have relatively few resources left to adapt to it (see Tafalla and Evans, 1997). Indeed, Miedema and Vos (1999) showed that noise annoyance is slightly higher in people who have a more demanding job and who have more people to care for in their household. Although the effects of these factors were generally small, their combined effect in middle adulthood and their absence in early and late lives may have shaped the age by annoyance curve as it was found in the present study. Further research is needed to substantiate this adaptive resources account, for instance, by comparing noise effects between groups of people experiencing different levels of daily workload.

Another important factor that may have an impact on noise-induced annoyance is hearing acuity. It is well known that aging is accompanied by a general loss of sensory acuity (e.g., Fozard and Gordon-Salant, 2005), which may make older people less susceptible to noise. Although previous research has shown that effects of verbal noise on cognitive performance in young and older adults are not influenced by an age-related decline of auditory function (Beaman, 2005; Bell and Buchner, 2007; Murphy *et al.*, 1999), presbycusis might to some extent explain why especially the oldest respondents report the lowest levels of annoyance from environmental noise. On the other hand, the signal-to-noise ratio required for speech comprehension increases with declining auditory function, which may contribute to an increased vul-

nerability to noise with aging (Dubno *et al.*, 1984; Plomp and Mimpen, 1979; Tun, 1998). However, this effect may be subordinate to the alleviating effect of sensory decline.

The present study shows a curvilinear effect of age on self-reported noise sensitivity that is similar to the effect of age on annoyance. This suggests that noise sensitivity may change depending on the individual's age or circumstances. Although self-reported noise sensitivity is an important determinant of noise annoyance (e.g., Miedema and Vos, 1999, 2003), it does not fully explain the effect of age on noise annoyance in the present study. This means that the age-related changes in susceptibility to noise cannot be entirely attributed to lifespan changes in self-reported noise sensitivity.

VI. CONCLUSION

Our analyses consistently show that annoyance from noise follows an inverted U-shaped pattern as a function of age, where the youngest and oldest respondents report the lowest, and people in their mid-40s report the highest levels of annoyance. For example, among 45-year-olds, our model predicts up to 10% more highly annoyed individuals than among 20-year-olds, and up to 20% more highly annoyed individuals than among 80-year-olds. This pattern of results was found in a very large, pooled international dataset as well as in a separate large aircraft noise dataset (Amsterdam Airport Schiphol). The latter appears to have been affected by the social unrest that was evoked by plans to expand Amsterdam Airport Schiphol. However, this was only the case for the model coefficients, not for the nature of the effect of age on annoyance (e.g., the age at which annoyance peaks).

The samples analyzed in the present study are highly representative, not only because of their size, but also because they include individuals from culturally and socially diverse countries, including two non-Western countries, Turkey and Japan. This adds to the potential importance of the current results. First, our results may indicate that cognitive development, if of any influence at all, is not a major determinant of noise-induced annoyance. Second and more importantly, our results suggest that those age groups that supposedly are most resilient to environmental noise are, on balance, most vulnerable to it. This age-related vulnerability may also apply to effects of noise on health, given that De Kluizenaar *et al.* (2007), who studied the role of age in the relationship between traffic noise and hypertension, only found an effect on health in middle-aged individuals. A possible explanation for this phenomenon is that, because of a relatively high level of daily mental workload, the adaptive resources of middle-aged people are pushed to the limit by the presence of noise. Further exploration of this idea, with a special focus on the joint effects of daily mental workload and ambient noise exposure on annoyance, is highly warranted.

Beaman, C. P. (**2005**). "Irrelevant sound effects amongst younger and older adults: Objective findings and subjective insights," European J. Cognitive Psychology **17**, 241–265.

Bell, R., and Buchner, A. (**2007**). "Equivalent irrelevant-sound effects for old and young adults," Mem. Cognit. **35**, 352–364.

Blakemore, S.-J., and Choudhury, S. (**2006**). "Development of the adolescent brain: Implications for executive function and social cognition," J. Child Psychol. Psychiatry **47**, 296–312.

Buckner, R. L., Head, D., and Lustig, C. (**2006**). "Brain changes in aging: A lifespan perspective," in *Lifespan Cognition: Mechanisms of Change*, edited by F. I. M. Craik and E. Bialystok (Oxford University Press, New York), Chap. 3, pp. 27–42.

Bluhm, G., Nordling, E., and Berglind, N. (**2004**). "Road traffic noise and annoyance: An increasing environmental health problem," Noise Health **6**, 43–49.

Cohen, S. (**1980**). "Aftereffects of stress on human performance and social behavior: A review of research and theory," Psychol. Bull. **88**, 82–108.

De Kluizenaar, Y., Gansevoort, R. T., Miedema, H. M. E., and De Jong, P. E. (**2007**). "Hypertension and road traffic noise exposure," J. Occup. Environ. Med. **49**, 484–492.

Dempster, F. N. (**1992**). "The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging," Dev. Rev. **12**, 45–75.

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.

Ellermeier, W., and Zimmer, K. (**1997**). "Individual differences in susceptibility to the 'irrelevant speech effect'," J. Acoust. Soc. Am. **102**, 2191–2199.

Enmarker, I., and Boman, E. (**2004**). "Noise annoyance responses of middle school pupils and teachers," J. Environ. Psychol. **24**, 527–536.

Evans, G. W., Hygge, S., and Bullinger, M. (**1995**). "Chronic noise and psychological stress," Psychol. Sci. **6**, 333–338.

Fields, J. M. (**1993**). "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M. (**2001**). "An updated catalog of 521 social surveys of residents' reactions to environmental noise (1943–2000)," Report No. NASA/CR-2001-211257, National Aeronautics and Space Administration (NASA), Hampton, VA.

Fozard, J. L., and Gordon-Salant, S. (**2001**). "Changes in vision and hearing with aging," in *Handbook of the Psychology of Aging*, 5th ed., edited by J. E. Birren and K. W. Schaie (Academic, San Diego), Chap. 10, pp. 241–266.

Gisselgård, J., Petersson, K. M., and Ingvar, M. (**2004**). "The irrelevant speech effect and working memory load," Neuroimage **22**, 1107–1116.

Groothuis-Oudshoorn, C. G. M., and Miedema, H. M. E. (**2006**). "Multilevel grouped regression for analyzing self-reported health in relation to environmental factors: The model and its application," Biom. J. **48**, 67–82.

Guski, R. (**2004**). "How to forecast community annoyance in planning noisy facilities," Noise Health **6**, 59–64.

Guski, R., Felscher-Suhr, U., and Schuemer, R. (**1999**). "The concept of noise annoyance: How international experts see it," J. Sound Vib. **223**, 513–527.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., and Rypma, B. (**1991**). "Age and inhibition," J. Exp. Psychol. Learn. Mem. Cogn. **17**, 163–169.

Hygge, S. (**2003**). "Noise exposure and cognitive performance: Children and the elderly as possible risk groups," in WHO Technical Meeting on Noise and Health Indicators, Brussels, Belgium.

Hygge, S., Evans, G. W., and Bullinger, M. (**2002**). "A prospective study of some effects of aircraft noise on cognitive performance in schoolchildren," Psychol. Sci. **13**, 469–474.

ISO (**2002**). "Acoustics: Description, measurement, and assessment of environmental noise—Part 2: Determination of environmental noise levels," Report No. ISO/CD 1996-2, International Standards Organization (ISO), Geneva, Switzerland.

Job, R. F. (**1988**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.

Langdon, F. J. (**1976**). "Noise nuisance caused by road traffic in residential areas: Part III," J. Sound Vib. **49**, 241–256.

Lundberg, U., Mårdberg, B., and Frankenhaeuser, M. (**1994**). "The total workload of male and female white collar workers as related to age, occupational level, and number of children," Scand. J. Psychol. **35**, 315–327.

McKennell, A. C. (**1963**). "Aircraft noise annoyance around London (Heathrow) Airport" (Her Majesty's Stationery Office, London).

Michaud, D. S., Keith, S. E., and McMurchy, D. (**2005**). "Noise annoyance in Canada," Noise Health **7**, 39–47.

Michaud, D. S., Keith, S. E., and McMurchy, D. (**2008**). "Annoyance and

disturbance of daily activities from road traffic noise in Canada," J. Acoust. Soc. Am. **123**, 784–792.
Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.
Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.
Miedema, H. M. E., and Vos, H. (**1999**). "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344.
Miedema, H. M. E., and Vos, H. (**2003**). "Noise sensitivity and reactions to noise and other environmental conditions," J. Acoust. Soc. Am. **113**, 1492–1504.
Murphy, D. R., McDowd, J. M., and Wilcox, K. A. (**1999**). "Inhibition and aging: Similarities between younger and older adults as revealed by the processing of unattended auditory information," Psychol. Aging **14**, 44–59.
Nijland, H. A., Hartemink, S., Van Kamp, I., and Van Wee, B. (**2007**). "The influence of sensitivity for road traffic noise on residential location: Does it trigger a process of spatial selection?" J. Acoust. Soc. Am. **122**, 1595–1601.
Nivison, M. E., and Endresen, I. M. (**1993**). "An analysis of relationships among environmental noise, annoyance and sensitivity to noise, and the consequences for health and sleep," J. Behav. Med. **16**, 257–276.
Ouis, D. (**2001**). "Annoyance from road traffic noise: A review," J. Environ. Psychol. **21**, 101–120.
Plomp, R., and Mimpen, A. M. (**1979**). "Speech-reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66**, 1333–1342.
Stansfeld, S. A. (**1992**). "Noise, noise sensitivity and psychological studies," Psychol. Med. Monogr. Suppl. **22**, 1–44.
Stansfeld, S. A., Berglund, B., Clark, C., Lopez-Barrio, I., Fischer, P., Öhrström, E., Haines, M. M., Head, J., Hygge, S., Van Kamp, I., and Berry, B. F. (**2005**). "Aircraft and road traffic noise and children's cognition and health: A cross-national study," Lancet **365**, 1942–1949.
Tafalla, R. J., and Evans, G. W. (**1997**). "Noise, physiology, and human performance: The potential role of effort," J. Occup. Health Psychol. **2**, 148–155.
TNO-PG and RIVM (**1998**). "Hinder, slaapverstoring, gezondheids- en belevingsaspecten in de regio Schiphol: Resultaten van een vragenlijstonderzoek [Annoyance, sleep disturbance, health, and experience aspects in the Schiphol region: Results of a survey] (summary in English)," Report Nos. TNO: 98.052 and RIVM: 441520011, TNO-PG and RIVM, Leiden/Bilthoven, The Netherlands.
Tomasi, D., Caparelli, E. C., Chang, L., and Ernst, T. (**2005**). "fMRI-acoustic noise alters brain activation during working memory tasks," Neuroimage **27**, 377–386.
Tun, P. A. (**1998**). "Fast noisy speech: Age differences in processing rapid speech with background noise," Psychol. Aging **13**, 424–434.
Van Kempen, E. E. M. M., and Van Kamp, I. (**2005**). "Annoyance from air traffic noise: Possible trends in exposure-response relationships," Report No. 01/2005 MGO EvK, National Institute of Public Health and the Environment, Bilthoven, The Netherlands.
Wallenius, M. A. (**2004**). "The interaction of noise stress and personal project stress on subjective health," J. Environ. Psychol. **24**, 167–177.

Policy discourse, people's internal frames, and declared aircraft noise annoyance: An application of Q-methodology

Maarten Kroesen[a)]
Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, P.O. Box 5015, 2600 GA Delft, The Netherlands

Christian Bröer
Faculty of Social and Behavioral Sciences, University of Amsterdam, OZ Achterburgwal 185, 1012 DK Amsterdam, The Netherlands

(Received 19 August 2008; revised 18 March 2009; accepted 27 April 2009)

Aircraft noise annoyance is studied extensively, but often without an explicit theoretical framework. In this article, a social approach for noise annoyance is proposed. The idea that aircraft noise is meaningful to people within a socially produced discourse is assumed and tested. More particularly, it is expected that the noise policy discourse influences people's assessment of aircraft noise. To this end, Q-methodology is used, which, to the best of the authors' knowledge, has not been used for aircraft noise annoyance so far. Through factor analysis five distinct frames are revealed: "Long live aviation!," "aviation: an ecological threat," "aviation and the environment: a solvable problem," "aircraft noise: not a problem," and "aviation: a local problem." It is shown that the former three frames are clearly related to the policy discourse. Based on this observation it is argued that policy making is a possible mechanism through which the sound of aircraft is turned into annoyance. In addition, it is concluded that the experience of aircraft noise and, in particular, noise annoyance is part of coherent frames of mind, which consist of mutually reinforcing positions and include non-acoustical factors. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3139904]

PACS number(s): 43.50.Rq, 43.50.Qp, 43.50.Sr, 43.50.Lj [BSF] Pages: 195–207

I. INTRODUCTION

Exposure to aircraft noise in residential areas is a prime focus of protests and policy in many countries. In Europe it is estimated that in 2006 2.2×10^6 people were exposed to annual aircraft noise exposure levels of Lden 55 dB(A) or more and 3.0×10^6 Europeans were exposed to night-time noise levels of Lnight 45 dB(A) or more (MPD, 2007). In addition, the population within the Lden 55 dB(A) is expected to increase to 2.3–2.4 in 2010 and to 2.6–2.7 in 2015 (MPD, 2007).

While aviation generally increased over the past decades, noise tolerance seems to decrease. Today less noise is necessary to have an equal portion of highly annoyed people (Guski, 2002, 2004; Bröer and Wirth, 2004; Van Kempen and Van Kamp, 2005; Schreckenberg and Meis, 2007). In an updated review of Van Kempen and Van Kamp (2005), Schreckenberg and Meis (2007) showed that exposure-response functions of the period 1990–2008 are different from those collected in the period 1965–1992 on which EU policy is based (Miedema and Oudshoorn, 2001; European Communities, 2002). The older "EU-curve" is found to structurally underestimate the negative community response observed presently.

Several explanations for this trend have been provided. One is the change in the structure of the noise load: The average noise load of single events has decreased, but the number of events has increased (Guski, 2004). This change is concealed by annual energy equivalent noise metrics, which are generally used to predict noise annoyance, and the new structure might be experienced as more annoying. Other explanations focus on changes in individual characteristics (e.g., noise sensitivity) or on changes in attitudes such as trust in the noise source authorities, which might have come about due to the advent of the risk averse society (Wirth and Bröer, 2004). Guski (2004) provided yet another reason in arguing that recent aircraft noise studies have been done in the context of step changes in noise exposure levels, which are known to cause so-called excess negative response on top of the response to be expected from exposure-response curves derived from steady-state situations.

In this study, however, we focus on a different explanation, one which has received little attention in previous research. This explanation focuses on the policy discourse at airports. A policy discourse is defined as the way policy actors socially and publicly define and handle problems. We hypothesize that public definitions of aircraft noise are internalized in frames, which people adopt to evaluate aircraft noise. For example, if the policy discourse identifies aircraft noise as an important problem, we expect that people will internalize this definition, and in doing so, become more annoyed by the noise. The explanation for the trend toward higher annoyance then lies in changes in the policy discourse.

In this article we propose a social explanation for declared noise annoyance. Based on previous work of Bröer (2006) the main hypothesis of the present study is that policy making is a possible mechanism through which the sound environment due to aircrafts is turned into noise annoyance. The main assumptions underlying this hypothesis are that (1) people make use of already existing frames to appraise an

[a)]Author to whom correspondence should be addressed. Electronic mail: m.kroesen@tudelft.nl

environmental stimulus such as aircraft noise (cf. Nijhof, 1995, 1998, 2003) and (2) one of the most influential sources of these frames is the policy discourse. More specifically, the hypothesis can be decomposed into two distinctive processes: an internalization process of the policy discourse in internal frames of people and, second, using this internal frame, an appraisal process of aircraft noise. It is assumed that the policy discourse (and subsequently also the internal frame) contains "feeling rules" (Hochschild, 1979): It legitimizes or delegitimizes concerns, complaints, or fears. This can be modeled like the following: The policy discourse treats aircraft noise as a problem and (de)legitimizes annoyance → cognition and feeling rules are internalized by people around the airport → people feel annoyed by aircraft noise.

We do not claim that these relationships are unidirectional. A policy discourse can develop within a field of multiple actors, including citizens. Furthermore, people's frames can depend on personal characteristics such as gender, age, or noise sensitivity. Their role, however, is not the focus of the present study.

Focusing on the criterion of association the present article will investigate the relationship between the policy discourse and the internal frames of people. To that effect the following approach is adopted. First, the policy discourse at one airport, namely, Amsterdam Schiphol (the largest airport in The Netherlands), is characterized. This particular airport is chosen for two reasons. First, the policy discourse at Amsterdam Schiphol explicitly defines aircraft noise annoyance as a problem, a necessary condition if the aim is to investigate whether this definition resonates with the internal frames of people. And second, sufficient previous research is already available to provide a satisfactory description of the policy discourse. Second, the different perspectives used to study aircraft noise annoyance will be reviewed. This review shows that to assess subjectivity, Q-methodology is well-suited. Third, we reveal the frames people adopt to evaluate aircraft noise and how these relate to the policy discourse and to the declared level of noise annoyance. The rationale behind the approach described here is that if (1) a resemblance is found between the internal frames and the policy discourse (at a single moment in time) and (2) noise annoyance response is found to be intrinsically related to the revealed internal frames, there will be sufficient evidence to support the hypothesis that there is a strong relation between the policy discourse and aircraft noise annoyance.

II. NOISE POLICY DISCOURSE AT AMSTERDAM SCHIPHOL

Hajer (1995) (p. 264) defined a discourse[1] as "an ensemble of ideas, concepts, and categories through which meaning is given to social and physical phenomena, and which is produced and reproduced through an identifiable set of practices." Hence, a policy discourse can be regarded as the way policy actors (socially) define and handle public problems. Useful elements to guide these definitions are policy concepts, story-lines, and metaphors. In addition, although multiple discourses surrounding an issue can be identified, only one of those is (usually) dominant. Hajer (2006) defined dominance using two criteria, namely, discourse structuration and discourse institutionalization. The former relates to the degree a particular discourse dominates a given social unit (e.g., a policy domain). It refers to the degree a discourse is shared among multiple actors, the so-called discourse-coalition.[2] The latter relates to the degree a discourse is institutionalized in policy processes and policy measures. When both conditions are satisfied a discourse is said to be dominant. The current description of the policy discourse will only focus on the dominant discourse. Although alternative discourses can be identified, this focus is justified by the argument that this discourse is most visible to residents around the airport.[3]

The present description of the policy discourse related to the issue of aircraft at Amsterdam Schiphol is based on several existing studies (Dierikx and Bouwens, 1997; Van Eeten, 1999, 2001; Abma, 2001; Wagenaar and Cook, 2003; Bröer, 2006). It is meant to identify the dominant policy discourse for noise annoyance in The Netherlands.

Before aircraft became a problem of noise annoyance, aviation had been introduced to The Netherlands as an economic asset and as a part of national development since 1919. In policy documents Schiphol airport and aviation were placed in a historical perspective, relating them to the image of The Netherlands as a successful seafaring nation in the golden age. Based on this analogy the airport should be regarded as something to be trusted and accepted and the government should strive to develop an airport that plays a role on a global scale.

In the mid-1950s aircraft noise was first identified as a (potential) problem. In the following decades this problem was, in line with the physical expansion of the airport, treated as a spatial planning problem. The fundament of the noise policy was to fit the airport, with its noise footprints, in the residential environment surrounding the airport, such that the flight routes avoided living areas. Other (implicit) assumptions followed from this central planning perspective. First, human response to aircraft noise was expected to be uniform. The physical noise level therefore became the central outcome of interest for policy regulation. Second, since spatial planning was a matter of centralistic control, a major role was given to national governmental bodies and (acoustical) experts in the development of the airport, while residents surrounding the airport were assumed to be passive. Third, planning and noise policy focused on long term developments, which were expressed in statistics, maps (showing noise contours), and scenarios. And lastly, solutions proposed by policy makers and advisory commissions to solve the noise problem were spatial and technocratic in nature (e.g., repositioning runways or flight routes, improving aircraft engines, restrictive land-use policies, and relocation of the airport to the sea).

However, the planning discourse failed because flight operations and housing more and more overlapped. From the 1960s onwards, therefore, policy makers accepted noise pollution in residential areas. Citizens around Amsterdam Schiphol, however, following the discourse's own premise that aircraft noise is an important problem, did not settle in their role as passive receivers. In the period between 1965

and 1995 the history of Schiphol knows many citizens' protests. In these protests the disciplinary effect of the policy discourse can be observed. Although citizens oppose the policy they still express themselves in terms of the planning discourse by advocating for solutions such as the repositioning of runways and relocation of the airport. The unsolvable conflict caused by the planning discourse (i.e., "noise is an avoidable problem" versus "some noise needs to be accepted") as well as the (resulting) protests led to a deadlock. To escape it a new (international) story-line was introduced in the 1990s, called "ecological modernization" (Weale, 1992; Mol and Spaargaren, 1993; Hajer, 1995). The combination of this story-line with the existing discourse has led to the policy discourse that exists to the present day, which Bröer (2006) termed the "mainport and environmental discourse."

The basic assumption of this new story-line was that economy and environment could be developed at the same time; the attainment of both economical and ecological goals should be regarded as a positive-sum game. The promise of ecological modernization relied strongly on developments in science and technology and market-based policy instruments (e.g., environmental taxes). Related to Amsterdam Schiphol the economic benefits of aviation became known under the umbrella of the "mainport," which was considered a vital entity to The Netherlands if it were to play a role in the globalizing economy. Schiphol should be seen as an "engine of the economy." The ecological negative externalities, most notably noise, but also risk and pollution, became known under the umbrella of the "environment."[4] From 1990s the mainport and environment discourse was spread among citizens through extended participatory processes. Repeatedly, citizens were called upon to be alert, to be informed, and to express their interests. In 1995, the mainport and environment discourse was institutionalized, when the decision was made to construct Schiphol's fifth runway (mainport) and to implement noise contours (environment).

Although the principle of ecological modernization seems to have provided a viable new perspective, it can actually be seen as an explicit reformulation of the existing problem conceptualization (i.e., the planning discourse) in modern (neo-liberal) terms. Policy makers seek to accommodate growth of the airport while trying to avoid its negative effects on the environment via traditional planning instruments. The only difference is the explicit acknowledgment of both economical and environmental effects/values.

III. THREE PERSPECTIVES TO STUDY AIRCRAFT NOISE ANNOYANCE

In studying noise annoyance three perspectives and related research approaches can be distinguished: the acoustical aggregate model, the (non-)acoustical disaggregate approach, and the discourse approach. In the following these three perspectives will be briefly discussed and their suitability to our research aim indicated.

The acoustical aggregate model has focused on the most obvious determinant of noise annoyance: the physical level of noise exposure. The effects of this variable are presented as exposure-response relationships, e.g., the percentage of highly annoyed people, at a given level of noise exposure, calculated or measured with energy-based noise metrics such as Lden. Schultz (1978), who was the first to integrate the results of 11 community surveys, developed a general exposure-response relationship for transportation noise, which was updated by Fidell *et al.* (1991) and Miedema and Vos (1998). The physical level of noise exposure can reveal community response but cannot account for all individual variability in noise annoyance. Based on review of 39 surveys Job (1988) concluded that only 9–29% of the variation in negative reaction (i.e., noise annoyance) can be explained by noise exposure. Since the aim of the present study is to elucidate the different frames of people, this model, which focuses on community response, does not suit our purpose well. The disaggregate non-acoustical approach [also termed the individual or situational difference model (Lercher, 1996)], which developed in response to the limitations of the acoustical model, is more in line with our aim, but it still does not fully suffice.

Within this disaggregate non-acoustical modeling approach the effects of personal and situational variables on individual levels of noise annoyance are studied via survey research or experiments, controlling for the level of noise exposure. Several non-acoustical factors have been identified. Borsky (1961), McKennell (1963), and Leonard and Borsky (1973) showed that noise annoyance is associated with source evaluation, misfeasance in relation to the authorities, fear of an aircraft crash, and concern about health effects. Job (1988) found that the attitude to the noise source and sensitivity to the noise account for more variance in annoyance than noise exposure does. A meta-analysis of Fields (1993), based on 136 surveys, revealed that socioeconomic and demographic variables (age, sex, social status, income, education, home ownership, dwelling type, length of residence, and personal benefit) had no influence on the level of noise annoyance. Instead, annoyance was related to the amount of insulation from sound at home, fear of danger from the noise source, noise prevention beliefs, general noise sensitivity, beliefs about the importance of the noise source, and annoyance with non-noise impacts of the noise source. Similar results were obtained by Miedema and Vos (1999). Overviews of relevant non-acoustical factors are given by Lercher (1996), Guski (1999), and Kroesen *et al.* (2008). The last mentioned authors identified 28 (potentially relevant) non-acoustical factors.

The disaggregate approach uncovered a wide range of factors empirically related to aircraft noise annoyance. In addition, it seems well-suited to investigate the causal structure, which underlies noise annoyance. Still, for our aim, it is unfit. In the first place, we are not interested in the statistical associations between variables, but in the frames people adopt to evaluate aircraft noise. (Linear) combinations of variables can be used to predict (or explain) annoyance response, but they are not suited to capture or qualify the frames we hope to reveal.

Second, the disaggregate approach recognizes that annoyance is partly based on subjectivity, but (implicitly) assumes that all people have the same understanding of non-acoustical factors such as trust in the source authorities or

noise sensitivity. Hence, the approach generally assumes an objective and unchanging frame of reference when different people respond to different questions. A concept such as noise annoyance, however, can be subject to a host of different definitions, each of which may be sensible within a specific social context. An *a priori* meaning of the concept introduces arbitrary subjectivity in the measurement process, which carries the risk of missing or misinterpreting meaning from the respondents own frame of reference.

A study of King *et al.* (2004) is illustrative for the way a social or political context can cause differences in internal frames of reference. They measured the perceived level of political efficacy within a Mexican and Chinese sample with the following question: "How much say do you have in getting the government to address issues that interest you?" It turned out that 50% of the Mexicans, while living in a democratic country, reported to have no say, in contrast to 30% of the Chinese, while living in a non-democratic (communist) country, reported to have no say. According to King *et al.* (2004) the explanation lies in the fact that Chinese have lower standards for what counts as satisfying the level described by any given response category. Hence, although their "actual" level of political efficacy is lower, the difference in the frame of reference between Mexicans and Chinese is cause for the found opposite result. This exemplifies the need to have an understanding and operationalization of an issue, which is grounded in specificities of a field.

The aggregate model or disaggregate modeling approach provide valuable insights on their own terms. In addition, it has been shown possible to make inferences about the internal frames of people with traditional questionnaire techniques and statistical analysis (Raimbault *et al.*, 2003). Yet, we want to put forward now a different approach, which pays more attention to differences in frames of reference.

A step toward an alternative approach was taken by Bröer (2006, 2007a, 2007b). His main thrust was to understand aircraft noise annoyance from subjects' own frame of reference. Instead of testing an already existing theory, his aim was to develop a new theory, which is grounded in the meaning people attribute to sound (Glaser and Strauss, 1967; Blumer, 1969; Charmaz, 2006). In line with the present study he assumed that sound is meaningful within a coherent frame, a concept which is connected to discursive psychology (Billig, 1987; Potter and Wetherell, 1987; Edwards and Potter, 1992). Here, a frame is defined as a discourse that operates at the individual level a coherent set of beliefs and attitudes that people use to observe and give meaning to reality. In general, frames guide the extraction of relevant cues from ongoing flows of events and act as filters through which we (selectively) observe the world, attribute meaning to it, and act on it (Goffman, 1974; Rein and Schön, 1993; Schön and Rein, 1994; Weick, 1995). Bröer (2006) argued that phenomena labeled "non-acoustical factors" can be part of such a frame. Furthermore, Bröer (2006) assumed that people learn or internalize frames socially and hypothesized that the frames' subjects develop to give meaning to the experience of aircraft noise are influenced by the policy discourse related to the issue of aircraft noise at an airport.

If the policy discourse influences people's attitude to aircraft noise, one would find different kinds of noise annoyance in different political settings. Therefore Bröer (2006) studied the policy discourses and people's frames of aircraft noise at two European airports: Amsterdam Airport Schiphol in The Netherlands and Zurich Kloten in Switzerland. He found that at similar sound levels the aircraft noise was indeed experienced differently *between* the two cases and that those differences can be traced back to different noise policies. Different attitudes toward noise *within* a case were related to the dominant policy discourse too: Typically people strongly adopted part of the dominant policy discourse and rejected or downplayed other parts. In general, people were found to evaluate noise policy when they heard aircraft sound and to have internalized the language and the logic of the policy. Based on these results Bröer (2006) concluded that noise annoyance is shaped by the policy discourse.

This third perspective is most closely related to our formulated aim. However, Bröer (2006) worked with an interpretative approach, which begs the questions if the frames he found can be objectified. Therefore, in contrast to Bröer's (2006) qualitative methodology, we use Q-methodology. In line with Bröer's (2006) approach this method assumes that subjectivity is anchored in self-reference. However, in contrast to Bröer's (2006) approach, the Q-method *can* be used to render internal frames of people manifest in an objective way (Brown, 1980; McKeown and Thomas, 1988).

The three perspectives are summarized in Fig. 1 and Table I. The present study will be in line with the discourse model and will further investigate the hypothesis that the policy discourse surrounding a particular airport becomes internalized in the frames people adopt to evaluate the meaning of aircraft noise. Yet, in contrast to Bröer's (2006) qualitative methodology, we use Q-methodology to render the internal frames of people visible. Lastly, we acknowledge the influence of personal determinants (e.g., age, gender, and noise sensitivity) and the physical level of aircraft noise exposure on people's frames, but these influences are not assessed.

IV. Q-METHOD

The basic idea of Q-methodology (Brown, 1980) is that people rank-order statements derived from everyday communication and that these rank-orderings (i.e., so-called Q-sorts), instead of traits related to the individual, are correlated and factor analyzed. When two Q-sorts are shown to correlate, the persons who constructed them are said to share a similar frame. By factor-analyzing a correlation matrix of $n \times n$ persons/Q-sorts, shared frames can be extracted. Underlying this procedure is the premise that subjectivity is anchored in self-reference. Subjects are encouraged to actively construct their opinion on the topic at hand. In addition, by letting the subjects rank-order the statements (on a single scale), they are evaluating and interpreting them in relation to each other. If, like in our study, subjects sort 48 statements, this involves, at least implicitly $(\frac{1}{2})(48)(48-1) = 1128$ judgments. This procedure is based on the assumption that meaning is relational: A specific statement cannot be seen in isolation but derives meaning from its relation to

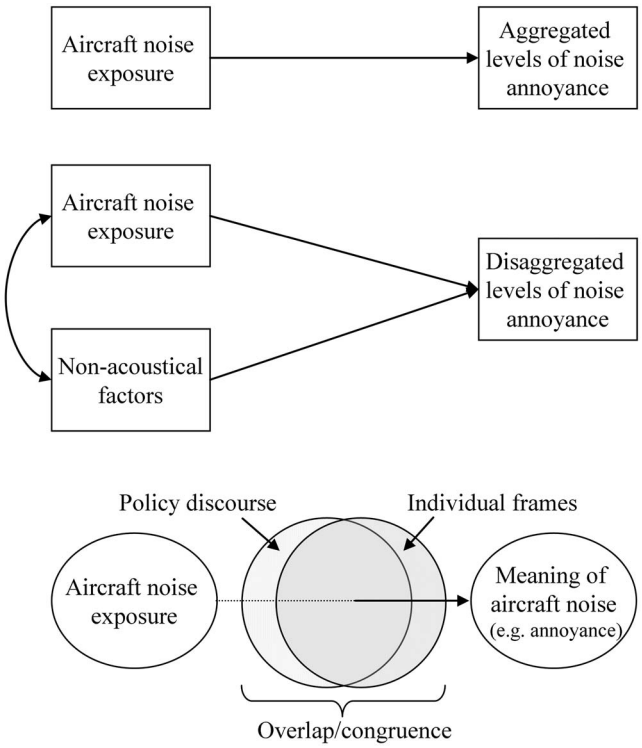


FIG. 1. Model structures of the three perspectives. The aggregate acoustical model (top figure), the disaggregate approach (middle figure), and the discourse approach (lower figure). The discourse approach assumes that meaning is provided to aircraft noise through an individual frame, which is schematized here as a filter. Second, it assumes that the individual frames are congruent with the policy discourse. We schematized the entities in the discourse approach as ovals to indicate their nature as *fixed qualities*. In contrast to the other two approaches were the entities (rectangles) relate to *variable quantities*.

other statements (a position common in Gestalt psychology, philosophy of language, discourse analysis, and large parts of interpretive social science). For example, when two people strongly agree with the statement "I am annoyed by aircraft noise," survey research treats those expressions as part of the same category. In a relational perspective, the statement might refer to disturbance or to unfair treatment and can therefore constitute two (or even more) different kinds of annoyance. Since the aim of the present study is to explore the different frames in which people are (not) annoyed by aircraft noise, Q-methodology seems well-suited for this task. Below we describe the way Q-method is applied to our case.

A. Defining the Q-sample

First, one has to define the "concourse:" the whole of statements of opinion, related to a certain topic that can be found among members of a social group (Stephenson, 1978; Brown, 1980). In this case the concourse encompasses all expressions by residents living in the vicinity of Schiphol Airport related to the topic of aircraft noise. Based on previous research of Bröer (2006) statements were theoretically sampled (Glaser and Strauss, 1967; Charmaz, 2006) from four diverse sources: thematically structured interviews with residents living in Amsterdam Osdorp related to the topic of aircraft noise ($n=47$), complaints to the Commission Regional Dialogue Schiphol ($n=130$), letters to the editor from residents around Schiphol published in three national newspapers ($n=71$), and statements from residents during public inquiry procedures ($n=18$). This provided us with 240 different statements about aircraft noise.

To select a Q-sample from these statements we used academic literature to identify four key themes: (1) perceptions of aircraft noise (covering statements related to hearing aircraft, being disturbed by aircraft, fear, noise annoyance, etc.), (2) non-acoustical factors (covering statements related to trust in noise source authorities, perceived control, expectations, etc.), (3) policy story-lines (covering statements related to economic benefits, ecological costs of aviation, complaining, etc.), and (4) "autonomous noise annoyance definitions," which are not covered in one of the first three and are rather unusual (covering statements such as "people have the right for silence"). To arrive at a representative sample, statements within each category were selected until all (sub)categories were covered. The final Q-set consisted of 48 statements and can be found in Table II (Sec. V). The final sample is *naturalistic* in the sense that the statements were derived from participants' own communications about aircraft noise and *structured* in the sense that theoretical (sub)themes were used to categorize the concourse, which ensured coverage of all relevant issues related to aircraft noise in the final sample (McKeown and Thomas, 1988).

B. Participants and procedures

We presented the selection of statements to residents of part of Amsterdam Osdorp, in The Netherlands. This was also the area were the initial interviews were held. All respondents were exposed to the same aircraft noise. The av-

TABLE I. Three perspectives to study the effect of aircraft noise on humans.

Acoustical aggregate model (top figure in Fig. 1)	(Non-)acoustical disaggregate approach (middle figure in Fig. 1)	Discourse approach (bottom figure in Fig. 1)
Main objective: (given the noise level) to predict aggregated levels of noise annoyance (i.e., community response).	Main objective: to predict/explain variation in individual levels of noise annoyance.	Main objective: to study the link between policy discourses and the internal frames, which people adopt to qualify aircraft noise (non-acoustical factors can be part of the internal frames).
Limitations: (1) Large portions of variance in (community) reaction remain unexplained and (2) unable to reveal internal frames.	Limitations: (1) Difficult to reveal internal frames and (2) implicit assumption of an objective frame of reference.	Limitation: difficult to generalize the results to a larger population.

TABLE II. Factor arrays of the five rotated factors.

No.	Statement	A	B	C	D	E
1	It is convenient to live near Schiphol.	3	−1	1	−1	−1
2	Schiphol should be allowed to stay: Long live aviation!	4	−1	0	2	0
3	I regularly hear aircraft.	3	4	3	−2	5
4	I fear that aircraft noise will increase.	−1	2	1	−4	5
5	I have the feeling that aircraft noise is forced on me.	−2	4	−3	−3	−4
6	The sound of aircraft belongs to this day and age.	3	−2	0	2	1
7	If you cannot stand aircraft noise, you should go and live somewhere else.	1	−3	0	−3	−4
8	It is acceptable that people are disturbed by aircraft noise in their dwelling.	0	−5	−4	−2	−3
9	It is acceptable that people have to interrupt a conversation due to aircraft noise.	−2	−5	−2	0	−1
10	I am annoyed by aircraft noise.	−5	2	−1	−4	2
11	Air traffic is a hazard for public health.	−2	1	1	0	3
12	The growth of Schiphol goes at the expense of the quality of life of many citizens.	−1	5	−3	−1	2
13	I cannot control the noise and this makes me feel angry and powerless.	−4	0	−2	−3	−3
14	If you do not pay attention to it (i.e., the noise) then you will not be bothered by it.	3	−4	1	0	−3
15	I can do something against the noise.	1	−2	−1	−2	−5
16	If I could I would move to a quiet neighborhood.	−4	−1	−4	−1	4
17	I am afraid that one day an aircraft will crash nearby.	−2	−1	−5	1	−3
18	As citizen you are powerless against Schiphol.	−3	1	0	3	−2
19	It does not help to complain about aircraft noise.	−1	0	−1	3	0
20	There is sufficient consideration for residents around Schiphol.	2	−4	−2	3	0
21	Citizens should collectively move up against aircraft noise.	−1	0	−3	−4	2
22	If people complain about aircraft noise they mainly serve their self-interest. They do not realize how important Schiphol is to The Netherlands.	1	−3	−2	5	2
23	There is too much attention for a small group of complainants around Schiphol.	0	−3	1	2	0
24	You cannot solve the "annoyance" problem. Schiphol has been around for a long time and this is something we have to deal with.	2	−1	−1	3	0
25	Flying is too cheap.	−4	0	−3	−1	−5
26	More technology will be developed that will reduce the noise.	4	1	3	0	3
27	Schiphol acts as a free-state making its own rules and regulations.	−1	1	−5	0	−4
28	I believe that Schiphol always gets his way.	−3	3	0	0	2
29	Schiphol does enough to reduce the noise.	0	−4	−2	−1	−2
30	The government does enough to reduce the noise.	0	−3	−1	−5	−2
31	The government does not live up to their promise to reduce the noise.	−1	1	−4	3	1
32	It is a good thing that the environmental movement and local action groups stand up for residents living around Schiphol.	1	3	2	−1	3
33	They always expand the airport first, and then raise the norms for the allowed levels of noise.	0	2	0	2	1
34	Schiphol is an engine of the economy.	5	0	5	1	0
35	We should be proud of our national airport.	4	0	2	1	−2
36	Aviation is important for the employment.	5	3	5	1	1
37	Noise annoyance from aircraft is an important problem.	0	4	2	−3	−1
38	Aviation is a threat to the environment.	0	5	2	1	1
39	The government should strive for reducing noise annoyance.	2	2	3	0	4
40	The government should strive for growth of Schiphol.	1	−2	−1	−5	−1
41	Economic interests are more important than reducing the level of noise annoyance.	1	−2	2	−2	0
42	Schiphol is big enough and should not be allowed to grow any further.	−3	1	3	4	−1
43	The double-sided aim (more growth but not more annoyance) of the government has failed. In the end the choice is always made to accommodate growth.	2	2	1	4	0
44	People have the right for silence.	0	3	4	2	4
45	Aircraft noise is "meaningless" (Dutch: zinloos) noise.	−3	−1	0	0	−1
46	I think it is a good idea to have an "aircraft-free-Sunday" every now and then.	−2	−2	4	0	3
47	Schiphol should be relocated to the sea.	−5	0	4	−2	−2
48	Further away from Schiphol aircraft noise is not really a problem.	2	0	0	5	1

erage noise level in this neighborhood, calculated over the period of 1 year, is approximately Lden 53 dB(A) (Bröer, 2007a). It is located approximately 5 km from the center of Amsterdam Schiphol. For the Q-method, 40–60 subjects are sufficient (Watts and Stenner, 2005). Respondents are chosen strategically: based on criteria derived from theory. In this case we included people who are highly, moderately, and not annoyed.

The data were collected at people's homes, by students under the close supervision of both authors in the period March–April 2008. We asked respondents to rank-order the 48 statements according to the following: "To which extent

do you agree/disagree with the following statements?" The scale ranged from −5 (most disagree) to +5 (most agree). In total 43 respondents completed the Q-sorting task and participated in a short interview afterwards. The interview asked for reasons behind respondents' rankings, additional topics, and noise annoyance, measured with the first item of the standardized noise annoyance scale developed by Fields *et al.* (2001).

C. Analysis

To identify similarly patterned Q-sorts, a correlation matrix of $n \times n$ Q-sorts (n=43 subjects) was calculated and factor analyzed using the method of centroid factor analysis (Brown, 1980). The PQMETHOD software (Schmolck, 2002) was used for this purpose. Based on Brown's (1980) recommendation seven factors were initially extracted. Next, the varimax rotation method was used to approximate simple structure. In line with standard Q-methodological practice only factors with two or more significant loadings[5] and an eigenvalue greater than 1 were considered acceptable. After rotation it was found that two factors did not satisfy these criteria. These were therefore disregarded from further analyses.

Next, factor exemplars to compute the composite factor arrays are identified. These are participants' Q-sorts that significantly and solely load on a factor and can therefore be considered as representative for the thought pattern present in the factor on which they load. Via the formula $2.58(1/\sqrt{N})$ and with N=48 (i.e., the number of statements) it can be calculated that loadings greater than ± 0.37 are significant at the 0.01 level. However, following the approach described by Watts and Stenner (2005), the confounding of participants (i.e., the number of participants that load on two or more factors) is minimized by raising this level to ± 0.40. At this level 37 participants load solely on one factor, 3 participants load on two factors, and 3 participants load on none of the factor. Hence, 86% of the data are used in the final analysis of the factors.

Lastly, the factor exemplars are merged into factor arrays, which represent "idealized" Q-sorts of hypothetical persons loading 100% on the factors.

V. RESULTS

A. Frames of residents around Schiphol

In the following the factors will be interpreted based on the computed factor arrays (Table II). For each factor, we indicate its relation to the noise policy discourse. Central to the first three factors is their relation to the mainport and environment policy discourse. In line with our theoretical argument, the factors are called frames below.

1. Frame A: Long live aviation! (the economic stance)

This frame is shared by 14 subjects and can account for 17% of the total variance of the correlation matrix.[6] In line with the policy discourse it strongly emphasizes the economic benefits of Schiphol airport (34: 5; read: statement 34, score 5) and of aviation in general (36: 5). According to this account we should be proud of our national airport (35: 4) and be cheerful about it (2: 4). Schiphol should grow (42: −3) and certainly not be relocated to the sea (47: −5). In this frame, one is optimistic about the future: Technology will reduce aircraft noise (26: 4) and aircraft noise is not expected to increase (4: −1).

While this frame strongly subscribes to the economic argument of the noise policy, it plays down the ecological arguments: Aviation is not considered a threat to the environment (38: 0) and noise annoyance is not considered a major problem (37: 0). Subjects tend to disagree with statements that aircraft noise is a hazard to public health (11: −2) and are indifferent about the statement that growth of Schiphol reduces the quality of life (12: −1).

In line with playing down the ecological arguments, complaining about noise is not supported: Subjects are indifferent about the statement that those who complain about noise are selfish and do not see the bigger picture (22: 1). They believe that residents around the airport receive sufficient consideration (20: 2) and they have no intention to engage in a collective action to address the noise problem (21: −1).

Given the support for economic reasoning, subjects are indifferent about the efforts of the government and Schiphol to reduce the noise (30: 0 and 31: 0). The relationship with the noise source authorities is mildly positive to neutral. Subjects do not believe Schiphol always gets its way (28: −3) and are indifferent about the statement that this actor makes its own rules and regulations (27: −1). This indifference can also be observed in relation to the statement that the government does not live up to its promise to reduce the noise (31: −1).

In this frame, the aim of the government to combine economic growth and ecology has failed (43: 2), but this does not go together with an overall negative attitude toward authorities.

Subjects subscribing to this frame do not consider themselves to be annoyed by the aircraft noise (10: −5), although they do regularly hear aircraft (3: 3). In addition, they have no intention of moving to a quieter place (16: −4).

Lastly, the frame acknowledges that we live in modern times: The sound of aircraft belongs to this day and age (6: 3) and aviation is just something we need to deal with (24: 2). This is typical for a "go with the flow" attitude toward modernity.

Altogether, frame A has a clear structure: It strongly favors economic arguments and plays down everything related to ecology.

2. Frame B: Aviation: An ecological threat (the environmental stance)

This frame is shared by 15 subjects and can explain 18% of the total variance. In contrast to frame A, this frame emphasizes that aviation is an environmental threat (38: 5), that growth of Schiphol goes at the expense of the quality of life of many citizens (12: 5), that disturbance by noise is completely unacceptable (8: 5, 9: 5), and that aircraft noise annoyance is an important problem (37: 4), which cannot be

ignored (14: −4). In line with the policy discourse, this account subscribes to the conceptualization of aviation as an important environmental problem.

While the frame stresses "ecology" it is less supportive of "economy." Subjects neither confirm nor disconfirm that Schiphol is an engine of the economy (34: 0). Aviation, however, is considered to be important for employment (36: 3). Compared to frame A, there is a strong support for one half of the policy discourse, but less criticism toward the other half.

Like subjects in frame A, subjects in frame B agree with the statement that the double-sided aim has failed and that in the end the government always chooses to accommodate growth (43: 2). But, different from frame A, in frame B this is combined with an elaborate negative attitude toward authorities. One believes that there is insufficient consideration for residents around Schiphol (20: −4) and that the government and Schiphol are not putting in enough effort to reduce the noise (29: −4 and 30: −3). Subjects believe Schiphol always gets its way (28: 3) and that the noise norms are purposively manipulated following expansion of the airport (33: 2). Consequently and in contrast to all other frames, subjects feel that aircraft noise is forced on them (5: 4), that something which is net undesirable (38: 5 versus 34: 0 and 36: 3) is unwillingly/forcefully and unasked (20: −4) being imposed on them. Policy has failed in the sense that noise annoyance is out of control. It is only in this frame that subjects do not think that aviation belongs to this day and age (6: −2). Instead, it is a runaway train, which threatens citizens and the environment.

Within the account people support complaining (22: −3 and 23: −3) and environmental movements (32: 3). This support is stronger than in all other frames. This is of course in line with the ecological stance. It might also be interpreted as a way to counter the criticism often raised against complainants in The Netherlands.

Subjects within this frame consider themselves moderately annoyed by aircraft noise (10: 2) and claim they regularly hear aircraft (3: 4).

Altogether, frame B has a clear structure: It strongly favors ecology, puts less emphasis on economy, is strikingly critical about noise policy, and portrays noise as an uncontrolled ecological threat.

3. Frame C: Aviation and the environment: A solvable problem (the technocratic stance)

This frame is shared by three subjects and can explain 5% of the total variance. This particular frame closely resembles the policy discourse with regard to Schiphol. It underlines the benefits of aviation for the economy (34: 5) and employment (36: 5), but also mildly agrees with the statements that aviation is a threat to the environment (38: 2) and that noise annoyance is an important problem (37: 2). Environmental pressure groups are viewed positively (32: 2).

Complaining, in this frame, is necessary and useful in general (19: −1, 22: −2), but subjects are indifferent about the statement that there is too much attention for a small group of serial complainers (23: 1).

This frame accurately reproduces the dominant policy and supports the government's policy stronger than any other frame. Subjects strongly disagree with the statement that the government does not live up to its promise to reduce the noise (31: −4) and with the statement that Schiphol acts as a "free-state" (27: −5). Subjects do not feel powerless (13: −2) and do not have the idea that the sound is forced on them (5: −3). Still, subjects weakly disagree with the statements that the government and Schiphol do enough to reduce the noise (29: −2 and 30: −1, respectively). So even in this frame achievement of the double-sided aim of the government is not supported (43: 1).

It seems as if in this frame, subjects have internalized the dominant policy, but feel disappointed with the results. Subjects strongly agree with statements that Schiphol should be relocated to the sea (47: 4) and that it would be a good idea to have an "aircraft-free-Sunday" every now and then (46: 4). The first measure has been debated since the 1960s; the second one is in no way part of the dominant policy discourse.

In addition, subjects have faith in technology to reduce noise (26: 3) as well as in technology in general. This latter remark is supported by the fact that subjects within the frame are least fearful of a nearby aircraft crash (17: −5). It is plausible that the acknowledged failure of the double-sided aim does not lie in subjects' belief that this is a wrong aim to strive for but probably lies in subjects' belief that wrong or too few solutions are being implemented.

Lastly, although subjects do regularly hear aircraft (3: 3), they are not particularly annoyed by aircraft noise (10: −1). They do, however, find it unacceptable that people are disturbed by aircraft noise in their dwelling (8: −4) or that people have to interrupt a conversation due to the noise (9: −2).

The structure of this frame closely resembles the dominant policy. In this frame, a "technological fix" is the prime solution for the still existing tension between economy and ecology.

4. Frame D: Noise is not a problem (the anti-government stance)

This frame is shared by two subjects and can explain 4% of the total variance. This account neither strongly concurs with the policy discourse's propagation of aviation as an important driver of the economy (35: 1 and 36: 1), nor with its propagation of aviation as an important environmental threat (38: 1). Moreover, subjects even disagree with the statement that noise annoyance is an important problem (37: −3). The denial of aircraft noise as an important problem also becomes apparent from other statements: Subjects are not annoyed by aircraft noise (10: −4), they do not believe that the government should strive for reducing noise annoyance (39: 0), nor do they fear that aircraft noise will increase (4: −4), and they strongly agree with the statement that farther away from Schiphol aircraft noise is not really a problem (48: 5). In addition, subjects in this frame do not regularly hear aircraft (3: −2) in contrast to the other frames in which subjects agree to this statement.

The attitude that aircraft noise is not a problem is consistent with the strong non-complaining attitude present in this frame. Subjects strongly agree with the statement that people who complain about aircraft noise only serve their self-interest and wrongfully neglect the importance of Schiphol to The Netherlands (22: 5). In addition, they do not believe that citizens should move up collectively against the noise (21: −4) and agree with the statement that there is sufficient consideration for residents around Schiphol (20: 3).

Still, subjects believe that the government does not do enough to reduce the noise (30: −5), that the double-sided aim of the government has failed (43: 4), and that the government does not live up to its promise to reduce the noise (31: 3). Since subjects in this frame do not subscribe to the ecological or the economic arguments, their dissatisfaction is derived from a different argument. In this frame, subjects most strongly state that government should not strive for growth of the airport (40: −5) and that Schiphol is big enough and should not be allowed to grow any further (42: 4). Subjects probably fear the growth of the airport for which they blame politicians, not the industry. They do not believe Schiphol always gets its way (28: 0) or that it acts as a free-state (27: 0).

As mentioned earlier, subjects adhering to this frame do not find themselves annoyed by aircraft noise (10: −4), nor do they regularly hear aircraft (3: −2). As in frame A, subjects in frame D are rather indifferent about the acceptability of being disturbed by aircraft noise (8: −2 and 9: 0).

This frame is structured around the idea that the physical growth of the airport is insufficiently controlled by politicians, but this problem is not connected to either environmental or economic arguments. It might relate to a conservative anti-government frame in which the airport as such is big enough.

5. Frame E: Aviation, a local problem (the a-political stance)

This frame is shared by three subjects and can explain 5% of the total variance. Subjects in frame E are, similar to those in frame D, not very concerned with the positive economic effects (34: 0 and 36: 1) or the negative environmental effects (38: −1 and 39: 1) of the airport. Instead the consistent theme in this frame is that subjects evaluate the statements in terms of the direct consequences they hold to their personal situations. Therefore, subjects do not take a strong position in the wider public controversy related to the economy-ecology conflict, but instead react with strong agreement to the statements such as "I fear aircraft noise will increase" (4: 5) and "Air traffic is a hazard for public health" (11: 3).

The most striking feature of frame E is the subjects' desire to move to a quieter neighborhood (16: 4). In addition, subjects strongly disagree with the statement "I can do something about the noise" (15: −5). Only in this frame, people do not think that one should be proud of the airport (34: −2).

Subjects strongly believe that the government should strive for noise reduction (39: 4) and deny that noise annoyance is an important problem at the same time. They weakly believe that the government and Schiphol are not putting in enough effort to reduce noise (30: −2 and 31: −2) and that Schiphol always gets its way (28: 2). They support an "aircraft-free-Sunday" (46: 3), but relocation of the airport is not considered a good idea (47: −2). Although such a measure would of course result in direct positive effects (i.e., no more aircraft noise) it also has its direct disadvantages, for it would probably raise the price for air travel. This goes against subjects' desire to travel by air, which can be inferred from subjects' strong disagreement with the statement that flying is too cheap (25: −5).

Similar to subjects in frame B, subjects within this frame consider themselves to be moderately annoyed by aircraft noise (10: 2) and subjects regularly hear aircraft (3: 5). Lastly, they find it unacceptable to be disturbed by aircraft noise (7: −4 and 8: −3).

The line of reasoning in this frame is difficult to interpret. It does not resemble the dominant policy and seems inherently contradictory. What seems to stand out is a fear of personal damage, a desire to move away from the neighborhood, and no identification with the airport. This might be seen as an a-political stance. The ecology-economy conflict is turned into a local and personal problem, which can be solved with a local solution, i.e., moving to a quieter place.

B. The relation between the policy discourse and internal frames

We expected that the ways people approach aircraft noise (described in Sec. V A) are related to the way this noise is approached in policy discourse (Sec. II). Based on the results it can be concluded that the first three frames are clearly related to the policy discourse. Frame A follows the economic argument, and frames B and C follow both the economic and environmental arguments. Moreover, none of the frames denies the economic or environmental trains of thought. Frame A, the economic frame, does not acknowledge the environmental problems posed by aviation, but also does not deny them. Statements related to environmental concerns receive a neutral score, not a negative one. Frame B, the environmental frame, moderately agrees with part of the economic reasoning (i.e., employment). Lastly, frame C also sides with both arguments, but, in contrast to frame B, emphasizes the economic values. In addition, since the first three frames account for the major part of the total portion of explained variance (cumulative 40% of the total 49%), it can be concluded that the lines of reasoning expressed within the policy discourse interact with most of the participants' beliefs. Hence, the way the problem is framed in the policy discourse becomes internalized in the internal frames of people.

C. Noise annoyance response within the frames

Next, the noise annoyance response within each frame is assessed. This is done through examination of the position of statement 10, "I am annoyed by aircraft noise," in the factor arrays (see Table II). In addition, this information is supplemented with results from the standardized noise annoyance question posed in the short interview conducted after the

TABLE III. Position of statement 10 and the descriptive statistics of the standardized noise annoyance item.

Frame	Noise annoyance (0–10)						
	Position s10	Mean	Min	Max	Median	SD	*N*
A—Long live aviation!	−5	1.43	0	5	1	1.45	14
B—Aviation: an ecological threat	2	6.00	3	10	7	2.37	15
C—Aviation and the environment: a solvable problem	−1	4.00	3	5	4	1.00	3
D—Aircraft noise: not a problem	−4	2.50	2	3	2.5	0.71	2
E—Aviation: a local problem	2	6.33	4	10	5	3.21	3

Q-sorting exercise. The Q-methodological and traditional survey results are both reported to cross-validate the observations.[7] From Table III it can be deduced that the position of statement 10 for the different frames is overall consistent with the mean scores of the standardized noise annoyance item.

Differences greater than 2–3 between statement scores can be treated as significant (Brown, 1980). Based on this rule-of-thumb it is concluded that several annoyance scores vary significantly across frames. More specifically, the following comparisons are significant: frames A and D versus the other frames, frames B and E versus the other frames, and frame C versus the other frames.

Within frames A and D annoyance is strongly denied. For frame A the denial of aircraft noise as annoying is consistent with the belief that aviation has only economic benefits and is not associated with environmental costs. Frame D even explicitly denies aircraft noise as a problem. On the contrary, for frames B and E, annoyance is (moderately) justified. Frame B prioritizes ecological concerns over the economic benefits. Aircraft noise is regarded as a serious problem. Frame E does not relate to the environment-economy dichotomy. However, here, the local conflict justifies a negative response to noise. It is important to note, however, that frames B and E do not legitimize an extreme annoyance response. After all, benefits of aviation (being national or individual) *are* acknowledged, so one cannot totally oppose aviation/Schiphol. Lastly, frame C strongly supports economic benefits but also acknowledges environmental values. This goes together with an average noise annoyance score, which deviates significantly from the annoyance scores in the other frames.

Overall, it can be concluded that annoyance response is intrinsically related to the frames and that the frames legitimize or delegitimize different degrees of annoyance response. The variance in annoyance response (i.e., after controlling for the level of noise exposure by keeping its level constant) aligns well with the variation in frames. The present approach therefore provides an adequate means of understanding this variation.

VI. DISCUSSION

Lastly, we would like to reflect on the results of our analysis and focus our attention on two issues: the observed variation in frames and the noise annoyance response within the frames.

The first issue relates to the finding that people's frames and the policy discourse indeed overlap. With respect to this observation it can be questioned why we did not find one frame that fully resembles the policy discourse. In the following an argumentation will be provided why this finding would have been unlikely.

It could be speculated that a frame fully reflective of the policy discourse would position both economical and environmental arguments on the right side of the scale; after all, both are considered very important in the policy discourse. In line with the policy discourse, subjects would trust central planning authorities. However, such a frame was not found. Instead, subjects across all frames (mildly) agree with the statement that the government has failed to achieve the double-sided aim (statement 43: to let the airport grow and restrict environmental impacts at the same time). This critical evaluation can be explained by an inherent contradiction present within the policy discourse because, on the one side, the policy discourse relies strongly on technological advances, which are said to "fix" the problem, but, on the other side, these technological advances contribute to the growth of aviation. Hence, the situation remains that some aircraft noise will have to be accepted. The policy does not provide a clear solution to the economy-ecology conflict. Therefore, an inconsistency can be perceived within the policy discourse because it reproduces the contradiction it claims to solve.

Subsequently, in line with Festinger's (1957) theory of cognitive dissonance, which postulates that inconsistency among beliefs will cause an uncomfortable psychological tension, it can be argued that people are forced to resolve this inconsistency. It can be observed that each frame related to the policy discourse (i.e., frames A–C) has a distinct way of doing this. Frame A simply resolves the inconsistency by playing down the environmental arguments. For frame B, which prioritizes environment over economy, but indeed subscribes to both arguments, the inconsistency is resolved by "adding" other cognitions and feelings: a negative attitude toward the authorities, distrust that they will successfully handle the noise problem and feelings of a lack of control. The government makes a promise (less noise) but does not keep it (aviation and Schiphol keep growing), and is therefore not to be trusted. The well-established "non-acoustical factors" such as trust and control serve the purpose of resolving the perceived dissonance. Lastly, frame C, which prioritizes economy over environment, but also subscribes to both arguments, resolves the inconsistency via two ways. Like frame B it "adds" cognitions that the authorities fail to do

their job, but distinctively, it also places high hopes on possible future solutions, most notably, the relocation of the airport to the sea.[8]

Altogether, it can be argued that subjects experience an inconsistency within the policy discourse. The different ways to resolve the perceived dissonance lead to different frames. As can be observed from the lines of reasoning expressed in the frames, each has developed its own distinctive way of doing this. In addition, established "non-acoustical factors" such as trust and control are internalized as part of the frames and hence as part of an argumentative relation with policy makers. In fact, they can hardly be treated as isolated variables, but should be approached as part of specific discourses.

A second issue on which we want to focus relates to the noise annoyance response within the frames. In the present study it is assumed that the position of aircraft noise annoyance follows from the lines of reasoning present within the frames. However, it can be argued that the varying levels of disturbance, which people experience, dictates the adoption of specific policy arguments. A person who regularly feels disturbed by aircraft noise (e.g., who is interrupted in a conversation or awakened during sleep) might be selective in the adoption of the arguments that are congruent with this state. We hold the (preliminary) belief that people "construct" their experience of aircraft noise on the basis of the disturbances they experience as well as under influence of socially sanctioned arguments provided by the policy discourse. It can be argued that it is unlikely that people will become annoyed by the noise if they are not disturbed by it in any way and that, the other way around, people who have to interrupt a conversation due to the noise might not classify this as particularly annoying if the policy discourse would not legitimate such concerns. To substantiate this point further, it can be observed that in frame D, a person claims not to hear aircraft with any regularity (see statement 3 in Table II). This particular frame selectively ignores aircraft noise as relevant. This observation is consistent with a literature review of Stallen (2008), which suggests that (even) the perceived loudness of a stimulus is not determined by its physical characteristics alone but also by its (social) context, an insight which already existed in relation to noise annoyance (Maris *et al.*, 2007a, 2007b).

VII. CONCLUSION

In this study the hypothesis is investigated that policy making is a possible mechanism through which the sound environment due to aircraft is turned into annoyance. To this effect, the policy discourse is described and the internal frames of people are revealed via Q-methodology. The factor analysis revealed five frames, which residents around Amsterdam Schiphol adopt to evaluate aircraft noise. We showed that the three main frames are related to the policy discourse. Based on these results it is concluded that the policy discourse is a source of arguments, which plays a role in structuring the frames of people. Second, it is shown that the experience of aircraft noise, and, in particular, noise annoyance, is intrinsically related to whole and consistent frames: the meaning of sound depends on a large set of mutually reinforcing positions. Non-acoustical factors should be regarded as part of these specific comprehensive frames and serve the purpose of making these frames internally consistent. Lastly, it can be concluded that our approach has been effective in explaining the variation in annoyance response controlled for the level of noise exposure. The analysis has provided a better understanding of the (negative) experience of aircraft noise.

Finally, we can relate our findings to our point of departure, namely, the observable trend that presently people are more annoyed than several decades ago at equal (annual equivalent energy) noise levels. Our analysis suggests that this trend can be explained by the fact that today's policy discourses explicitly recognize aircraft noise as an important problem. This definition becomes internalized by people affected by aircraft noise and structures the experience of noise as negative.

To investigate our hypothesis further, the following directions for further research can be formulated. First, our research focused on the relationship between the policy discourse and individual frames at one moment in time without considering which of the two takes causal precedence. Bröer's (2006, 2007b) research provides data, which point at least to a historic precedence of policy arguments before people's frames. But the issue of causality remains. One might argue that annoyance is part of a field in which multiple actors (including policy makers, stakeholders, and citizens) together construct annoyance policy and frames. Further research should focus on this process. Particularly, one should focus on the micro-processes in which people develop perceptions of aircraft sound. By studying this process insights could be gained as to whether these coherent frames are built around experienced disturbances due to aircraft noise (which subsequently dictate the adoption of specific policy arguments) or around the arguments put forward by the policy discourse (which facilitates the formation of negative feelings and increases the proneness of being disturbed) or whether it is, in fact, a co-evolutionary process in which both processes mutually reinforce each other.

The second possible focus of future research is the distribution of the frames over the population. A mixed-method approach, combining Q-methodology with traditional survey methods, would have to be followed to gain information about the exact distribution. Within such a mixed-method model the effects of the physical level of aircraft noise exposure (which presently is not part of our model) could also be investigated. For example, it could be hypothesized that the distribution of different frames is different for varying levels of noise exposure.

A third direction is to study the policy discourse and individual frames at other airports. In this study the relationship between the policy discourse and the individual frames are studied for one airport only. To find further support for our hypothesis that the policy discourse shapes individual frames this relationship should be studied at multiple airports where different problem definitions exist. Airports where no well-defined noise policies exist would be even more interesting cases. In such instances one might find little negative

response to aircraft noise, find that those who are annoyed might need to go at great length to develop comprehensive frames that rationalize their negative experiences (since no pre-existing frames are available), find that other institutions provide people with a framework to interpret noise, or find that a much larger variety of individual frames exist (since no common frame is available). In short, research focused on such cases can yield interesting results.

Lastly, we would like to relate our findings to the policy practice. The analysis shows that the conceptualization of aircraft noise as an important problem by policy makers disciplines the way aircraft noise is evaluated. Should policy makers therefore stop treating aircraft noise as a problem? We do not believe so. In the first place, as we have seen in our analysis, there are frames that do no relate to the policy discourse and in which annoyance response to aircraft noise is still present. In addition, next to the disciplinary effect of the policy discourse on community response, we believe that the policy discourse also serves the function of channeling response. As mentioned in the previous paragraph, without this common discourse that people can fall back on in qualifying the sound of aircraft, it can be speculated that the variety of frames would probably be much larger and maybe more extreme. We believe that denial of aircraft noise as a problem should therefore not be regarded as a successful strategy.

However, the way policy deals with aircraft noise after acknowledging it as a problem is another issue. At Amsterdam Schiphol the mainport and environmental discourse is based on the premise that technological development is able to uncouple the divergent goals. Yet, in all of the revealed frames, whether pro-economy, pro-environment, or its combination, it is believed that achievement of the double-sided aim (growth *and* reduction in annoyance) has failed. If we relate this observation to Dryzek's (2001) (p. 652) notion of discursive legitimacy, which he defined as "the degree that collective outcomes are responsive to the balance of competing discourse in the public sphere," it can be concluded that the policy discourse's main premise is inconsistent with the frames shared among the public. This inconsistency undermines the legitimacy (and credibility) of the noise policy. Along this line of reasoning it would be better to let go the idea of a technological fix and explicitly choose for either economy or ecology.

ACKNOWLEDGMENTS

The authors would like to thank Thijs Bol and Pita Spruijt for their assistance during the data gathering phase as well as for their inspiration and critical comments. In addition, the authors owe their gratitude to Job van Exel and Michel van Eeten for their useful methodological advises and Eric Molin, Bert van Wee, Rainer Guski, Pieter Jan Stallen, and two anonymous reviewers for their comments on an earlier draft of this paper.

[1]There are several variants of discourse analysis, even within social psychology. What matters most to this study is the fact that "language in use" structures what can and cannot be said and thought in a specific situation. Discourse is different from "discussion" in the sense that it points to a pattern of the discussion.

[2]Hajer (1995) indicated that the term discourse-coalition differs from Sabatier's (1988) advocacy-coalition, which is a coalition of actors that share similar normative beliefs and/or interests. The essence of the term discourse-coalition is that actors with different and even competing goals (who by definition do not form an advocacy-coalition) can still be united under the flag of a discourse (in the sense that they share similar ways of thinking and acting).

[3]For the purpose of readability the term "dominant policy discourse" is therefore, in the remainder of this paper, equated and replaced with "policy discourse."

[4]Therefore, the remainder of this paper will treat the terms ecological and environmental interchangeably.

[5]Unlike in traditional applications of factor analysis the aim is not to account for as much variance as possible, instead its primary aim lies in finding unique shared viewpoints. At minimum, such a shared viewpoint can be identified based on two subjects.

[6]This value is calculated via the following formula: 100× (factor eigenvalue/number of subjects) (Brown, 1980).

[7]We acknowledge that the sample is too small to provide reliable estimates for the means and standard deviations. These figures are regarded as indicative.

[8]Here, a nice analogy between frame C and a particular smoker can be drawn. A smoker, who feels an inconsistency between smoking behavior and the cognition that smoking is bad for health, can neutralize this inconsistency by resolving to stop smoking in the (near) future. This postpones the feeling of being inconsistent.

Abma, T. (**2001**). "Narratieve infrastructuur en fixaties in beleidsdialogen, de Schiphol-discussie als casus (Narrative infrastructures and fixations in policy dialogues, the case of Schiphol)," Beleid Maatsch. **28**, 66–79.

Billig, M. (**1987**). *Arguing and Thinking: A Rhetorical Approach to Social Psychology* (Cambridge University Press, Cambridge).

Blumer, H. (**1969**). *Symbolic Interactionism: Perspective and Method* (Prentice-Hall, Englewood Cliffs, NJ).

Borsky, P. N. (**1961**). "Community reactions to Air Force noise I: Basic concepts and preliminary methodology. II: Data on community studies," WADD Technical Report No. 60-689 (I), National Opinion Research Center of the University of Chicago, Chicago, IL.

Bröer, C. (**2006**). "Beleid vormt overlast, hoe beleidsdiscoursen de beleving van geluid bepalen (Policy shapes annoyance, how policy discourses shape the experience of aircraft sound)," Ph.D. thesis, Aksant, Amsterdam; English summary available at http://home.medewerker.uva.nl/c.broer/bestanden/schijn.pdf (Last viewed 1/1/2009).

Bröer, C. (**2007a**). "Noise annoyance and policy: How policy shapes non-acoustical factors," in Internoise 2007, Istanbul.

Bröer, C. (**2007b**). "Aircraft noise and risk politics," Health Risk Soc. **9**, 37–52.

Bröer, C., and Wirth, K. (**2004**). "More annoyed by aircraft noise than 30 years ago? Some figures and interpretations," in Internoise 2004, Prague.

Brown, S. R. (**1980**). *Political Subjectivity: Applications of Q Methodology in Political Science* (Yale University Press, New Haven, CT).

Charmaz, K. (**2006**). *Constructing Grounded Theory* (Sage, Thousand Oaks, CA).

Dierikx, M., and Bouwens, B. (**1997**). *Building Castles of the Air, Schiphol Amsterdam and the Development of Airport Infrastructure in Europe, 1916–1996* (Sdu, The Hague).

Dryzek, J. S. (**2001**). "Legitimacy and economy in deliberative democracy," Polit. Theory **29**, 651–669.

Edwards, D., and Potter, J. (**1992**). *Discursive Psychology* (Sage, London).

European Communities (**2002**). *Position Paper on Dose Response Relationships Between Transportation Noise and Annoyance* (Office for Official Publications of the European Communities, Luxembourg).

Festinger, L. (**1957**). *A Theory of Cognitive Dissonance* (Stanford University Press, Stanford, CA).

Fidell, S., Barber, D. S., and Schultz, T. J. (**1991**). "Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233.

Fields, J. M. (**1993**). "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., Lercher, P., Vallet, M., Yano, T., Guski, R., Felscher-Suhr, U., and Schumer, R. (**2001**). "Standardized general-purpose noise reaction

questions for community noise surveys: Research and a recommendation," J. Sound Vib. **242**, 641–679.
Glaser, B. G., and Strauss, A. L. (**1967**). *The Discovery of Grounded Theory; Strategies for Qualitative Research* (Aldine, Chicago, IL).
Goffman, E. (**1974**). *Frame Analysis* (Harvard University Press, Cambridge).
Guski, R. (**1999**). "Personal and social variables as co-determinants of noise-annoyance," Noise Health **3**, 45–56.
Guski, R. (**2002**). "Status, tendenzen und desiderate der lärmwirkungsforschung (Results, trends and needs of research on community noise effects)," Zeitschrift fur Larmbekampfung **49**, 219–232.
Guski, R. (**2004**). "How to forecast community annoyance in planning noisy facilities," Noise Health **6**, 59–64.
Hajer, M. (**1995**). *The Politics of Environmental Discourse, Ecological Modernization and the Policy Process* (Clarendon, Oxford).
Hajer, M. (**2006**). "Doing discourse analysis: Coalitions, practices, meaning," in *Words Matter in Policy and Planning—Discourse Theory and Method in the Social Sciences*, edited by M. Van den Brink and T. Metze [Royal Dutch Geographical Society (KNAG), Utrecht], pp. 65–74.
Hochschild, A. R. (**1979**). "Emotion work, feeling rules and social structure," Am. J. Sociol. **85**, 551–575.
Job, R. F. S. (**1988**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.
King, G., Christopher, J. L. M., Joshua, A. S., and Ajay, T. (**2004**). "Enhancing the validity and cross-cultural comparability of measurement in survey research," Am. Polit. Sci. Rev. **98**, 191–205.
Kroesen, M., Molin, E. J. E., and Van Wee, B. (**2008**). "Testing a theory of aircraft noise annoyance: A structural equation analysis," J. Acoust. Soc. Am. **123**, 4250–4260.
Leonard, S., and Borsky, P. N. (**1973**). "A causal model for relating noise exposure, psychosocial variables and aircraft noise annoyance," *Proceedings of the International Congress on Noise as a Public Health Problem*, Dubrovnik, pp. 691–705.
Lercher, P. (**1996**). "Environmental noise and health: An integrated research perspective," Environ. Int. **22**, 117–129.
Maris, E., Stallen, P. J., Vermunt, R., and Steensma, H. (**2007a**). "Evaluating noise in social context: The effect of procedural unfairness on noise annoyance judgments," J. Acoust. Soc. Am. **122**, 3483–3494.
Maris, E., Stallen, P. J., Vermunt, R., and Steensma, H. (**2007b**). "Noise within the social context: Annoyance reduction through fair procedures," J. Acoust. Soc. Am. **121**, 2000–2010.
McKennell, A. C. (**1963**). *Aircraft Noise Annoyance Around Heathrow Airport* (Her Majesty's Stationary Office, London).
McKeown, B., and Thomas, D. (**1988**). *Q Methodology* (Sage, Beverly Hills, CA).
Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.
Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.
Miedema, H. M. E., and Vos, H. (**1999**). "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344.
Mol, A. P. J., and Spaargaren, G. (**1993**). "Environment, modernity and the risk-society—The apocalyptic horizon of environmental reform," Int. Sociol. **8**, 431–459.
MPD (**2007**). "Study of aircraft noise exposure at and around community airports: Evaluation of the effect of measures to reduce noise" (European Commission, Directorate-General for Energy and Transport, Directorate F—Air Transport), available at http://ec.europa.eu/transport/air_portal/environment/studies/doc/aircraft_noise_exposure_en.pdf (Last viewed 8/1/2008).
Nijhof, G. (**1995**). "Parkinson's-disease as a problem of shame in public appearance," Sociol. Health Illn. **17**, 193–205.
Nijhof, G. (**1998**). "Naming as naturalization in the medical encounter," J. Pragmat. **30**, 735–753.
Nijhof, G. (**2003**). *Tekstsociologie (Text Sociology)* (Aksant, Amsterdam).
Potter, J., and Wetherell, M. (**1987**). *Discourse and Social Psychology: Beyond Attitudes and Behavior* (Sage, London).
Raimbault, M., Lavandier, C., and Berangier, M. (**2003**). "Ambient sound assessment of urban environments: Field studies in two French cities," Appl. Acoust. **64**, 1241–1256.
Rein, M., and Schön, D. (**1993**). "Reframing policy discourse," in *The Argumentative Turn in Policy Analysis and Planning*, edited by F. Fischer and J. Forester (Duke University Press, Durham, NC), pp. 145–166.
Sabatier, P. (**1988**). "An advocacy coalition model of policy change and the role of policy-oriented learning therein," Policy Sci. **21**, 129–168.
Schmolck, P. (**2002**). "PQMethod (version 2.11)," available at http://www.lrz-muenchen.de/~schmolck/qmethod/ (Last viewed 8/1/2008).
Schön, D., and Rein, M. (**1994**). *Frame Reflection: Toward the Resolution of Intractable Policy Controversies* (Basic Books, New York).
Schreckenberg, D., and Meis, M. (**2007**). "Noise annoyance around an international airport planned to be extended," in Internoise 2007, Istanbul, Turkey.
Schultz, T. J. (**1978**). "Synthesis on social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.
Stallen, P. J. (**2008**). "When exposed to sounds, would perceived loudness not be affected by social context?," in Acoustics '08, Paris, France.
Stephenson, W. (**1978**). "Concourse theory of communication," Communication **3**, 21–40.
Van Eeten, M. J. G. (**1999**). *Dialogues of the Deaf: Defining New Agendas for Environmental Deadlocks* (Eburon, Delft).
Van Eeten, M. J. G. (**2001**). "Recasting intractable policy issues: The wider implications of The Netherlands civil aviation controversy," J. Policy Anal. Manage. **20**, 391–414.
Van Kempen, E. E. M. M., and Van Kamp, I. (**2005**). "Annoyance from air traffic noise. Possible trends in exposure-response relationships" [National Institute for Public Health and the Environment (RIVM), Bilthoven].
Wagenaar, H., and Cook, S. N. (**2003**). "Understanding policy practices: Action, dialectic and deliberation in policy analysis," in *Deliberative Policy Analysis: Understanding Governance in the Network Society*, edited by M. Hajer and H. Wagenaar (Cambridge University Press, New York,), pp. 139–171.
Watts, S., and Stenner, P. H. D. (**2005**). "Doing Q methodology: Theory, method and interpretation," Qual. Res. Psychol. **2**, 67–91.
Weale, A. (**1992**). *The New Politics of Pollution* (Manchester University Press, New York).
Weick, K. E. (**1995**). *Sensemaking in Organizations* (Sage, Thousand Oaks, CA).
Wirth, K., and Bröer, C. (**2004**). "Mehr belästigung bei gleichem pegel. Wieso flugzeuggeräusche heute möglicherweise lästiger sind als vor 40 jahren (More annoyance at equal noise levels. Why aircrafts noise today is more annoying than 40 years ago)," Zeitschrift für Lärmbekämpfung **51**, 118–121.

Evaluating standard airborne sound insulation measures in terms of annoyance, loudness, and audibility ratings

H. K. Park
Chonnam National University, Gwangju, 500–757, Korea

J. S. Bradley[a)]
National Research Council, Montreal Road, Ottawa K1A 0R6, Canada

(Received 12 January 2009; revised 18 March 2009; accepted 11 May 2009)

This paper reports the results of an evaluation of the merits of standard airborne sound insulation measures with respect to subjective ratings of the annoyance and loudness of transmitted sounds. Subjects listened to speech and music sounds modified to represent transmission through 20 different walls with sound transmission class (STC) ratings from 34 to 58. A number of variations in the standard measures were also considered. These included variations in the 8-dB rule for the maximum allowed deficiency in the STC measure as well as variations in the standard 32-dB total allowed deficiency. Several spectrum adaptation terms were considered in combination with weighted sound reduction index (R_w) values as well as modifications to the range of included frequencies in the standard rating contour. A STC measure without an 8-dB rule and an R_w rating with a new spectrum adaptation term were better predictors of annoyance and loudness ratings of speech sounds. R_w ratings with one of two modified C_{tr} spectrum adaptation terms were better predictors of annoyance and loudness ratings of transmitted music sounds. Although some measures were much better predictors of responses to one type of sound than were the standard STC and R_w values, no measure was remarkably improved for predicting annoyance and loudness ratings of both music and speech sounds. [DOI: 10.1121/1.3147499]

PACS number(s): 43.55.Hy [NX] Pages: 208–219

I. INTRODUCTION

Airborne sound transmission between spaces is measured in standardized tests. In North America the ASTM E90 (Ref. 1) procedure is used in the laboratory and the ASTM E336 (Ref. 2) procedure is used in field situations. In most other countries the ISO 140 procedures[3] are usually followed to measure airborne sound transmission through walls and floors. The two approaches are very similar and include single number ratings to reduce the results at a number of frequencies to a single numerical value. The sound transmission class (STC) from the ASTM E413 standard[4] and the weighted sound reduction index (R_w) from the ISO 717-1 standard[5] are quite similar and are widely used to specify the required sound insulation in various situations such as between homes. Both quantities are based on shifting a prescribed rating contour to match the measured values of sound transmission loss versus frequency following rules that specify the maximum allowed sum of all deficiencies below the contour. For the STC rating, a limit on the maximum allowed deficiency below the rating contour in a single frequency band is also specified.

The results of our previous research[6,7] showed that the ISO R_w and ASTM STC ratings were not good predictors of the intelligibility of transmitted speech sounds. Although the total allowed deficiency of 32 dB in the STC and R_w measures was found to be acceptable, the maximum allowed deficiency of the 8-dB rule was not helpful for predicting the intelligibility of transmitted speech. The intelligibility of transmitted speech was found to be well related to measures that are related to the intelligibility of speech, such as the articulation index,[8] the speech intelligibility index,[9] and the articulation class[10] as well as arithmetic averages of transmission loss values over frequency.

It is well known that many different types of sounds such as music, speech, television/radio, vacuum cleaners, etc., can be disturbing when transmitted through walls to neighbors.[11] Because the initial study only considered the intelligibility of transmitted speech, further investigations were needed to consider other ratings of airborne sound insulation and other types of sounds. Music sounds from neighbors are often said to be a prime cause of annoyance and complaints. For example, a survey of the occupants of refurbished flats in England[12] confirmed that, of airborne sounds from flats either side, the highest annoyance was due to the category "television/amplified music sounds." Bradley's[13] analyses of his survey results showed stronger negative responses for music sounds from neighbors either side than for the sounds of voices or radio and TV.

Although the disturbance caused by music sounds is a well-known problem, there have been very few studies to consider how well sound insulation measures predict subjective ratings of transmitted music sounds or of other common types of sounds. A review of a number of earlier studies is included in Ref. 14.

In an examination of the STC procedure, Clark[15] evaluated the merits of the STC contour shape and the 8-dB rule for music and speech sounds as well as vacuum cleaner

[a)]Electronic mail: john.bradley@nrc-cnrc.gc.ca

noise. In the first part of the work subjects adjusted the level of octave and 1/3-octave bands of the three test sounds to match the loudness of a broadband versions of the same sound that was frequency-weighted by a STC-contour shaped filter. The variation in the adjusted levels of the octave and 1/3-octave band speech and music sounds matched the shape of a STC contour quite well. However, this may simply indicate that the STC contour is a good approximation to the inverse of an equal loudness contour or an *A*-weighting contour, and subjects were rating the loudness of each octave or 1/3-octave band of the test signals. No attempt was made to test other contour shapes in sound insulation measures for the test sounds. The playback system was said to be flat ±3 dB from 100 Hz to 8000 Hz and presumably the test sounds did not include strong very low-frequency components.

The merits of the 8-dB rule were investigated by introducing dips electronically into a STC shape frequency response. The dips occurred at 500, 1000, and 2000 Hz and the depth of the dips was varied in 5-dB steps. Subjects adjusted the levels of the test sounds played through these simulated walls with dips until the test sounds were equally annoying to the same sound played through a simulated wall without a dip. Their results led them to conclude that adding the 8-dB rule did not improve the prediction of annoyance ratings of the test sounds. However, the 8-dB rule is most commonly applied for wall constructions at low frequencies well below the frequencies of Clark's dips. For example, of the 20 different real wall transmission loss versus frequency results used in the current study, in 12 of them the 8-dB rule was applied and always at lower frequencies. Thus it is not clear how well Clark's results apply to real walls with dips at lower frequencies.

Vian *et al.*[14] carried out a quite detailed investigation of sound insulation measures. Subjects rated the annoyance of 12 different music sounds each transmitted through 12 different simulated walls. The music sounds included different types of music and varied playback bandwidth. The simulated wall transmission loss characteristics were constructed by systematically varying slopes and dips and were not copied from the measurements of real walls. This made it possible to explore a broader range of conditions but they were not necessarily realistic representations of commonly occurring walls. Some included dips in the transmission loss at 160, 630, and 2500 Hz, but only the 160-Hz dip was representative of the property of wood and steel stud walls that most commonly triggers the application of the 8-dB rule. Vian *et al.* concluded that the measured level differences of walls to a pink noise test sound, that were frequency limited (125–4000 Hz), correlated well with annoyance ratings. They found that including music sounds with increased low-frequency components (40–10 000 Hz compared to 300–5000 Hz) increased annoyance ratings and that increases in the slopes of the transmission loss versus frequency between 100 and 1000 Hz were most important and the increased slopes led to reduced annoyance ratings.

Quite recently Rindel[16] reported the results of a study intended to confirm the merits of spectrum adaptation terms for music sounds with strong low-frequency components. Subjects rated the annoyance of transmitted music sounds heard via five simulated transmission loss characteristics. The five cases were based on systematic variations in the slope of the low-frequency portion of simulated transmission loss versus frequency characteristics. Adding the corrections of the spectrum adaptation term for the frequency range of 50–3150 Hz seemed to provide improved predictions of annoyance ratings of the music sounds but statistical test results of the data were not provided. There was a focus on impact sounds and no other types of airborne sounds were included. The results were based on only five artificial characteristics that may not be close approximations to those of real walls.

Because our initial study only considered the intelligibility of transmitted speech and because there have only been limited numbers of other previous investigations of the merits of measures of airborne sound insulation for specific types of sounds, the current study was carried out. The new study in the present paper compared standard sound insulation measures in terms of how well they predicted mean subjective annoyance and loudness ratings of transmitted music and speech sounds. The focus was to look for measures that better predict average subjective ratings, and individual subject variables such as sensitivity to noise were not considered. Subjects listened to music and speech sounds, which were modified to simulate transmission through 20 different walls with sound transmission characteristics based on the measured values of a wide range of walls. The annoyance and loudness ratings of these sounds were related to the standard R_w and STC ratings that are based on rating contours as well as variations in these standard procedures. Evaluation of other measures will be considered in a future publication.

II. EXPERIMENTAL DETAILS

A. Conditions common to previous study

All tests were conducted in the same sound isolated room as in the previous experiment.[6,7] The same equipment and procedures were used to playback sounds to the listeners. However, in the new experiments both speech and music sounds were used in combination with simulated constant ambient noise. The speech and music sounds were filtered to represent transmission through the same 20 walls used in the previous study[6,7] that had STC values evenly distributed from STC 34 to STC 58. These were representative of transmission through walls as measured in standard laboratory tests. They are described in Fig. 1 and Table I of the previous study.[7]

Listeners heard speech and music sounds from loudspeakers positioned 2 m in front of them and simulated ambient noise from another set of loudspeakers positioned above the ceiling, directly above the listener. As in the previous study,[6,7] the loudspeaker response was equalized to be flat ±1 dB from 60 Hz to 12 000 Hz. The simulated ambient noise had a −5-dB/octave spectrum shape with an overall level of 34.7 dBA, as illustrated in Fig. 1.

The speech tests used the Harvard sentences,[17] which were recorded by a clear speaking male in anechoic conditions. They are phonetically balanced and of low predictability. Listeners heard three different sentences played through

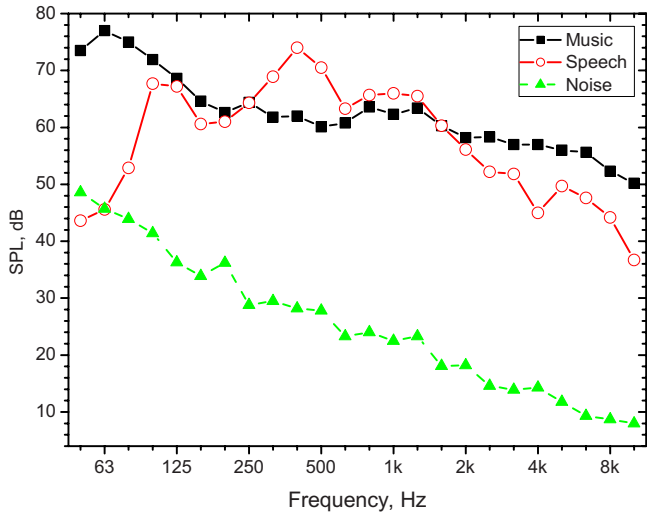


FIG. 1. (Color online) Average spectra of speech and noise sources before transmission through the walls as well as the spectrum of the simulated ambient noise.

each simulated wall. Three music samples were used which were selected from a large number of pieces as being potentially annoying and of different musical styles. They were rap music (The Roots, "I Remain Calm"), house music (Dizzee Rascal, "Stand Up Tall"), and pop music (Cyndi Lauper, "She Bop"). The average spectra of the speech and music sounds, before modification to simulate transmission through various walls, are shown in Fig. 1. The music is seen to have had a broad spectrum extending from 63 Hz to at least 6300 Hz and gradually decreasing in level with increasing frequency. The speech had a more restricted spectrum with highest levels between 100 and 1250 Hz.

In the first test, in which the subjects rated the annoyance of the sounds, the average sound pressure levels (SPLs) of the sound sources were 83.0 dB (72.1 dBA) for the music and 80.9 dB (75.2 dBA) for the speech. The source levels and the difference in source levels between speech and music sounds were fixed at levels such that the transmitted sounds varied from barely audible to clearly audible and quite loud for the range of simulated wall transmission characteristics. This was intended to maximize the range of responses.

The source sound levels for the second test, in which subjects rated the loudness of the sounds, were reduced by 6 dB relative to those for the annoyance test to make it possible to also determine the threshold of audibility of the transmitted sounds.

B. Procedures for annoyance and loudness ratings

To familiarize subjects with the types of sounds they would hear, they first did a practice test consisting of 12 sounds made up of 6 different test sentences and 6 music samples each played through one of 6 different simulated walls that varied from very low to very high STC rating. They were told that the practice examples were representative of the full range of conditions that they would hear in the full test. The practice test followed an initial hearing sensitivity test.

In the full test, listeners heard 3 different Harvard sentences and 3 different music samples through each of the 20 simulated walls for a total of 60 sentences and 60 music samples. The order of the speech and music samples and of the walls was randomized so that subjects heard conditions in one of three different randomized orders. In the experimental procedure, only the simulated transmission characteristics of the walls were varied. The effective speech or music source level and the ambient noise level at the listener's position remained constant throughout the tests.

In the annoyance tests, subjects were asked to imagine they were at home trying to relax. In this context they were asked to rate how annoying they would find each of the transmitted test sounds. They rated annoyance on a seven-point scale with the end points labeled "not at all annoying" and "extremely annoying." The mid-point was labeled "moderately annoying."

In the second part of the main study, subjects rated the loudness of the same transmitted speech and music samples. In the loudness test subjects rated the loudness of the sounds on an eight-point scale. Point No. 1 was labeled "not at all loud," point No. 7 "extremely loud," and point No. 4 "moderately loud." In the loudness rating test, they could also give a 0 response which was labeled "not audible" making it an eight-point scale. To ensure that there were a number of inaudible cases, the source levels for the loudness experiment were reduced by 6 dB relative to those of the annoyance experiment. This made it possible to determine the threshold of audibility as the point at which 50% of the subjects could just hear speech or music sounds.

In both the annoyance and the loudness experiments, subjects heard the test speech or music sounds followed by a 5-s gap. In the 5-s gap they rated the annoyance or loudness using a numeric keypad and their results were written to a computer file.

The results were analyzed in terms of the average annoyance or loudness scores for all listeners and all test sentences, or for all music samples, for each wall. That is, each average annoyance or loudness rating of transmitted speech sounds was an average of the scores for three sentences by all subjects for each wall. Similarly, each average annoyance or loudness rating of transmitted music sounds was the average of the ratings of three music samples by all subjects for each test wall.

Ten subjects completed the annoyance test. They were all NRC employees who volunteered to do the test after being approached by an electronic mail request for volunteers with good hearing and with English as their first language. The research was carried out according to the procedure approved by the NRC Research Ethics Board (Protocol 2006-6). Twenty subjects were tested for the loudness test. These subjects were hired from a temporary employment agency and were shared with three other NRC/IRC projects (according to Protocol 2006-27).

All subjects were first given a hearing sensitivity test. Their pure tone average (PTA) hearing levels varied from −8.0 to +9.2 dB (with an average of +1.6 dB) for the annoyance test subjects and from −3 to +16 dB (with an average of +4.2 dB) for the loudness test subjects. [PTA values are

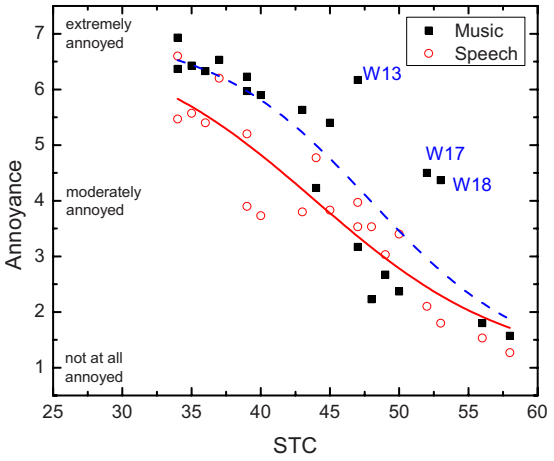


FIG. 2. (Color online) Mean annoyance ratings versus STC values for music and speech sounds (music: R^2=0.728; speech: R^2=0.856).

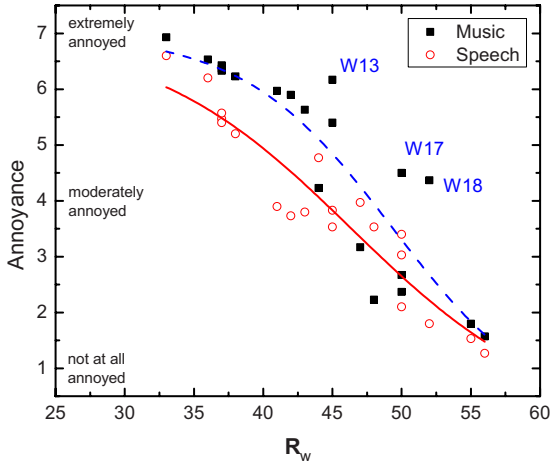


FIG. 3. (Color online) Mean annoyance ratings versus R_w values for music and speech sounds [music: R^2=0.798 (0.728); speech: R^2=0.890 (0.856)]. (Values in parentheses are R^2 for annoyance versus STC values from Fig. 2.)

averages of hearing level (HL) values over the test frequencies of 500, 1000, and 2000 Hz]. The means for these subjects were a little better than the 50th percentile hearing levels for normal hearing listeners.[18]

Most of the analyses were in the form of plots of the mean subjective ratings versus a sound insulation measure such as STC. To test the strength of the relationships between the subjective ratings and the sound insulation measures, Boltzmann equations were fitted to the plots and the related R^2 values calculated. The Boltzmann equation describes a sigmoidal-shaped relationship and was used because it fits the form of the responses, which approach asymptotically the maximum and minimum values of the response scales. It is defined by the following:

$$y = \frac{A_1 - A_2}{1 + e^{(x-x_0)/dx}} + A_2, \quad (1)$$

where A_1 is the y-value for $x=-\infty$ (1 for annoyance test), A_2 is the y-value for $x=+\infty$ (7 for annoyance test), x_0 is the x-value of mean y-value, that is, the x-value when $y=4$ for the annoyance test, and dx is related to the slope of the mid-part of the regression line.

Since there were always 20 data points and the same format of regression equation, the significance is simply related to the R^2 value (i.e., the coefficient of determination). Any R^2 value $\geq$0.193 is statistically significant at $p<0.05$ and an R^2 value $\geq$0.317 at $p<0.01$.

III. EVALUATION OF STANDARD MEASURES STC AND R_w

A. Annoyance ratings versus STC and R_w for music and speech

Figure 2 shows a plot of mean annoyance ratings versus the STC values of the 20 walls. The related R^2 values for the Boltzmann equation fits to these data are included in the figure title and indicate that STC values are better related to the annoyance to transmitted speech sounds than to the annoyance to transmitted music sounds.

The results for annoyance to music sounds for walls W13, W17, and W18 deviate more than other points from the regression line in Fig. 2. For these walls, the transmitted low-frequency sound levels were distinctly higher than for the other walls,[19] leading to higher annoyance ratings for music sounds and to the increased scatter in Fig. 2. Because the speech sounds did not include significant low-frequency sound, the same effect is not seen for speech sounds.

The relationships between annoyance ratings of music and speech sounds and R_w values in Fig. 3 are very similar to those versus STC values shown in Fig. 2. The related R^2 values shown in the title of Fig. 3 are slightly higher than those for STC values. Again the R^2 values indicated that annoyance ratings of music sounds were less well predicted than annoyance ratings of speech sounds and data for the same three walls (W13, W17, and W18) deviated more from the main trend of the annoyance to music ratings due to the increased low-frequency transmitted sounds.

B. Loudness ratings versus STC and R_w for music and speech

Figure 4 compares mean ratings of the loudness of the transmitted speech and music sounds plotted versus STC val-

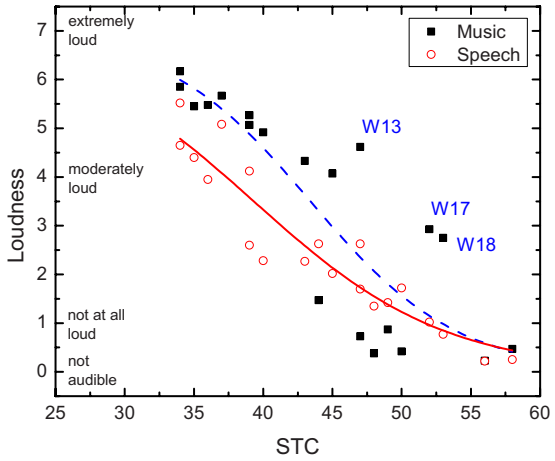


FIG. 4. (Color online) Mean loudness ratings versus STC values for speech and music sounds [music: R^2=0.734 (0.728); speech: R^2=0.886 (0.856)]. (Values in parentheses are R^2 values for annoyance versus STC from Fig. 2.)

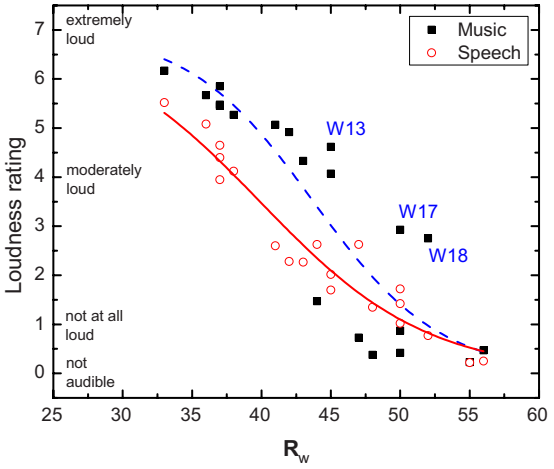


FIG. 5. (Color online) Mean loudness ratings for music and speech sounds versus R_w values [music: R^2=0.779 (0.798); speech: R^2=0.930 (0.890)]. (Values in parentheses are R^2 for annoyance versus R_w values from Fig. 3.)

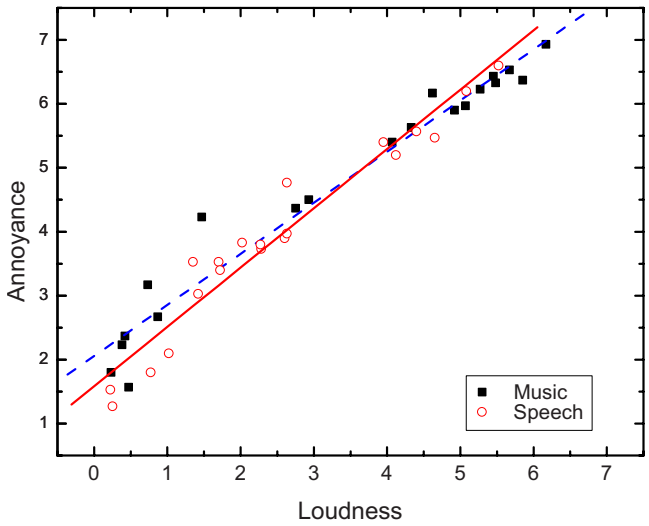


FIG. 6. (Color online) Relationship between mean annoyance and mean loudness ratings for both speech and music sounds [music: R^2=0.957; speech: R^2=0.943].

ues. Figure 5 shows similar results plotted versus the R_w ratings of the test walls. The results are similar to the previous plots of annoyance ratings in that STC and R_w are both better predictors of the loudness ratings of speech sounds than the loudness ratings of music sounds.

Comparisons of the various R^2 values from Figs. 2–5 indicate that R^2 values were higher for regression lines fitted to ratings of speech sounds than for ratings of music sounds and R_w values were a little better than STC values as predictors of annoyance and loudness ratings of both speech and music sounds. Loudness and annoyance ratings led to very similar relationships and neither loudness nor annoyance responses were distinctly better related to these two sound insulation measures. The regression coefficients of all regression equations are given in the Appendix.

C. Comparison of annoyance and loudness ratings

Comparing Figs. 2 and 4 shows similar annoyance and loudness ratings for both speech and music sounds when plotted versus STC values and the regression equations had similar associated R^2 values. Although the mean annoyance ratings were higher than the mean loudness ratings, the form of the regression lines was quite similar and the related R^2 values were also very similar. The results in Figs. 3 and 5 also show similar trends for annoyance and loudness ratings when plotted versus R_w values. For both speech and music sounds the Boltzmann equation fits for annoyance and loudness responses have similar slopes at the mid-points of the curves, as indicated by the similar dx values (see Appendix for regression coefficients). However, the x_0 values were between 3.8 and 4.5 dB larger for annoyance responses. As previously described, the response scales were a little different in that the loudness responses included a 0 value for subjects to indicate inaudible speech or music sounds. This, and the 6-dB difference in speech and music source levels, may largely explain the difference in average loudness and annoyance responses and the corresponding differences in x_0 values.

Figure 6 plots annoyance responses versus loudness responses. The near linear relationships and high R^2 values again indicate that these different concepts led to quite similar results. That is, sounds that were judged to be louder were also judged to be more annoying in these experiments. It is probably not necessary to assess both loudness and annoyance responses as they seem to provide almost the same information.

IV. VARIATIONS IN THE STC MEASURE

A. Variations in the 8-dB rule

In the previous study,[6,7] removing the 8-dB rule from the standard STC procedure led to slightly improved predictions of intelligibility scores. When mean annoyance responses were plotted versus STC_{no8} values (STC values calculated without an 8-dB rule), the related R^2 value was increased for annoyance ratings of speech sounds but was decreased for annoyance ratings of music sounds. The plots were very similar to those in Sec. III and can be found in Ref. 19. The regression coefficients and R^2 values are included in the Appendix of this paper. The R^2 values are also found in the summary given in Table I. Loudness ratings of speech sounds were better related to STC_{no8} values than to STC values, as indicated by the R^2 values in Table I. However, loudness ratings of music sounds were less well related to STC_{no8} values than to STC values.

The previous research[6,7] examined the benefits of varying the magnitude of the maximum allowed deficiency from the 8-dB value that is given in the standard.[4] Similar varia-

TABLE I. Comparison of R^2 values for Boltzmann equations fitted to annoyance and loudness ratings plotted versus values of the standard STC measure and STC_{no8} values obtained without an 8-dB rule. R^2 values equal to or greater than 0.95 are in bold font.

Measure	Annoyance speech	Annoyance music	Loudness speech	Loudness music
STC	0.856	0.728	0.886	0.734
STC_{no8}	**0.950**	0.670	**0.970**	0.654

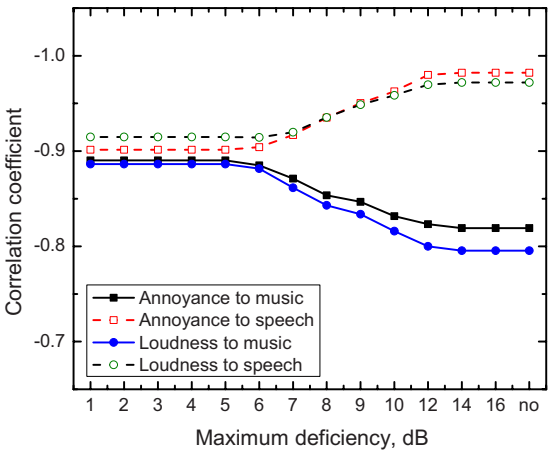


FIG. 7. (Color online) Correlation coefficients between annoyance and loudness ratings of speech and music sounds and modified STC values for which the maximum acceptable deficiency was varied for 1–16 dB including a no maximum deficiency limitation case.

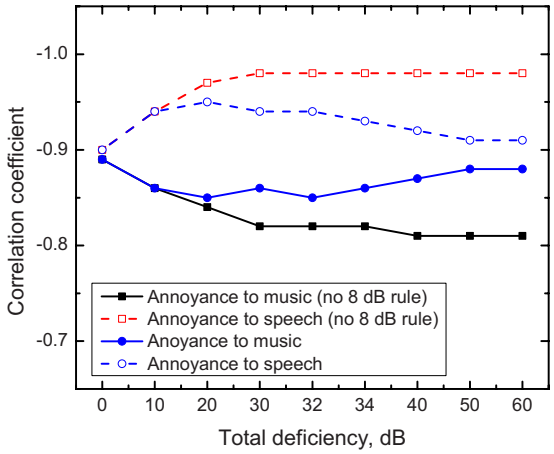


FIG. 8. (Color online) Correlation coefficients from correlations of annoyance ratings with modified STC values for which the total acceptable deficiency was varied from 0 to 60 dB.

tions in the allowed magnitude of the maximum deficiency were also examined in the current research. Figure 7 shows the results of linear correlations of mean loudness and annoyance ratings with STC values for varied maximum allowed deficiency values.

In Fig. 7, annoyance and loudness ratings indicate very similar correlation coefficients. However, responses to music sounds and those to speech sounds indicate different trends. For music sounds, higher magnitude correlation coefficients were found for smaller maximum allowed deficiencies and the same correlation coefficients were found for maximum allowed deficiencies from 1 to 5 dB inclusive. The inverse was true for speech; larger maximum acceptable deficiencies led to higher magnitude correlations with subjective ratings. For speech, the no 8-dB rule case led to the highest correlation coefficients because this was similar to a very large maximum allowed deficiency. The current value of 8 dB is a little more helpful for predicting responses to speech sounds than those to music sounds. The results in Fig. 7 indicate that reducing the 8 dB to a value between 1 and 5 dB would be a slightly better compromise for both speech and music sounds.

To better understand the results in Fig. 7, the frequency at which the 8-dB rule was applied was determined for all walls where the STC rating was limited by the 8-dB rule. The 8-dB rule was applied for 12 of the 20 walls and in all cases it was applied at lower frequencies (125–250 Hz). In nine of the walls it was applied in the 125-Hz 1/3-octave band. For the walls included in this study, the 8-dB rule functioned to better represent the effects of low-frequency dips in the transmission loss versus frequency responses. Where there was a large dip and low-frequency transmitted sounds could be particularly strong, the 8-dB rule limited the STC value so that it better indicated the effect of the reduced attenuation of the low-frequency transmitted sounds. Thus, for the music sounds with strong low frequency components, including the 8-dB rule led to better sound insulation ratings that were better correlated with subjective ratings of the music sounds. However, for speech sounds without strong low-frequency components, including the 8-dB rule distorted the rating of the more important mid- and higher-frequency components of the transmitted speech sounds.

The 8-dB rule is seen to be useful because it is one method of better including the influences of low-frequency dips in the transmission loss versus frequency characteristics when predicting subjective responses to sounds with significant low-frequency content such as music. Responses to speech sounds are better predicted without the 8-dB rule.

B. Variations in the total allowed deficiency

The effect of varying the total allowed deficiency in STC and R_w calculations was also examined by varying it from the 32-dB limit included in both the STC and R_w procedures. This was done for annoyance and loudness ratings of both speech and music sounds with and without the 8-dB rule being included. Figure 8 plots the results of correlations of annoyance ratings for speech and music sounds versus STC values with varied allowed total deficiency both with and without the 8-dB rule being included.

For responses to speech sounds and with no 8-dB rule, a maximum total deficiency corresponding to the current standard value of 32 dB worked well. If the maximum deficiency was much smaller than 32 dB, then the correlations with annoyance to speech sounds would decrease because the STC rating would become more influenced by prominent low-frequency dips in the transmission loss versus frequency characteristics. However, for annoyance to music sounds and without the 8-dB rule, the opposite was true. For music sounds, the highest correlations occurred for the minimum total deficiency (0 dB) because then the prominent low-frequency dips in the transmission loss versus frequency characteristics most influenced the STC rating.

When the 8-dB rule was included, the results were a little different. For annoyance ratings of speech sounds, the correlation coefficients increased until the total deficiency equaled 20 dB. Presumably at higher values the STC became more influenced by the application of the 8-dB rule because of prominent low-frequency dips. Because these low-frequency dips did not greatly influence annoyance to speech

TABLE II. R^2 values for Boltzmann equation fits to mean loudness and annoyance ratings plotted versus sound insulation measures based on R_w with various spectrum adaptation terms. R^2 values equal to or greater than 0.95 are in bold font.

Measure	Annoyance speech	Annoyance music	Loudness speech	Loudness music
R_w	0.890	0.798	0.933	0.779
R_w+C	0.741	0.918	0.821	0.900
$R_w+C_{\mathrm{tr(100–3150)}}$	0.566	**0.950**	0.676	**0.960**
$R_w+C_{\mathrm{tr(50–5000)}}$	0.388	0.943	0.482	**0.980**
$R_w+C_{\mathrm{tr\_mod}}$	0.541	**0.983**	0.634	0.991
R_w+C_{mod}	**0.975**	0.580	0.973	0.556
$R_{w(63–5000)}$	0.769	0.940	0.848	0.907
$R_{w(160–5000)}$	**0.972**	0.562	⋯	⋯

sounds, the correlation coefficients decreased for higher values of the allowed total deficiency. Conversely, for annoyance to music sounds with the 8-dB rule included, the highest correlations occurred for a 0 dB allowed total deficiency. In this case, the STC rating was determined totally by the low-frequency dips in the transmission loss versus frequency characteristics.

Plots of correlation coefficients between loudness ratings and STC values for which the total deficiencies were varied followed very similar trends to those in Fig. 8.[19] There is no indication that other combinations of allowed maximum and total deficiency would lead to greatly improved predictions of ratings of both speech and music sounds.

V. VARIATIONS IN THE R_w MEASURE

A. Evaluation of standard spectrum adaptation terms

The ISO R_w procedure includes two types of spectrum adaptation terms that can optionally be applied. The C-type spectrum adaptation term is intended to better represent responses to pink spectrum sounds and the C_{tr}-type correction is to improve the rating of outdoor sounds such as road traffic noise, which have strong low-frequency components. Both types of spectrum weighting were evaluated for annoyance and loudness ratings of speech and music sounds.

When annoyance ratings were plotted versus R_w+C values, plots similar to the previous relationships (Figs. 2–5) were obtained and Boltzmann equations were fitted to them. For ratings of music sounds, the resulting R^2 values were increased, relative to the corresponding R^2 values for the plots of ratings of music sounds versus R_w values (Figs. 3 and 5) but were decreased for ratings of speech sounds. That is, adding the C-type spectrum adaptation term improved the prediction of annoyance ratings of music sounds but reduced the accuracy of the prediction of annoyance ratings of speech sounds. The various R^2 values are summarized in Table II and the corresponding regression coefficients are included in the Appendix.

Adding the C_{tr}-type correction over the frequencies from 100 to 3150 Hz had a similar effect. When Boltzmann equations were fitted to plots of annoyance ratings of music sounds versus $R_w+C_{\mathrm{tr(100–3150)}}$ values, the related R^2 values were increased, relative to the corresponding R^2 values obtained when using only R_w values. Again R^2 values for the relationships with annoyance to speech sounds were decreased, as shown in Table II. Regression lines were also fitted to plots of the loudness ratings of the two types of sounds versus R_w values, combined with each of these spectrum adaptation terms. As shown in Table II, the R^2 values increased for both spectrum adaptation terms for loudness ratings of music sounds but decreased for loudness ratings of speech sounds. Both of these spectrum adaptation terms improved the prediction of ratings of sounds with strong low-frequency components such as the music sounds in this study.

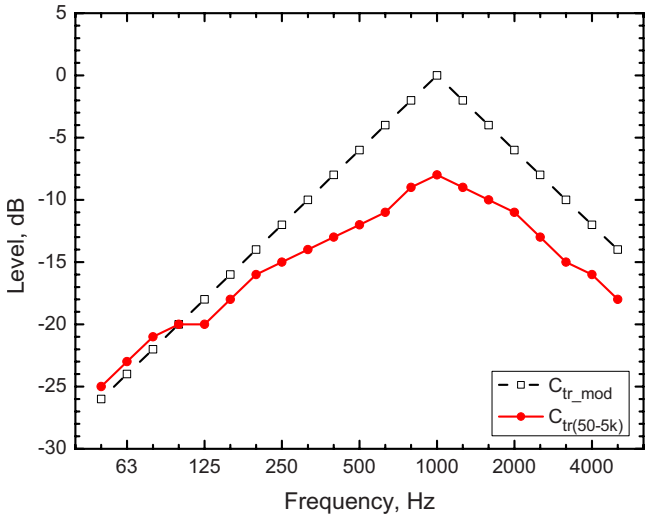


FIG. 9. (Color online) New spectrum weighting, $C_{\mathrm{tr\_mod}}$, and the standard $C_{\mathrm{tr(50–5000)}}$ weighting.

B. Evaluation of new spectrum adaptation terms

The ISO standard for R_w (Ref. 5) includes two variations in the C_{tr} spectrum weighting. One version, $C_{\mathrm{tr(100–3150)}}$, was considered in Sec. IV. It modifies only the frequencies from 100 to 3150 Hz. The other version modifies results over a broader range of frequencies from 50 to 5000 Hz and is illustrated in Fig. 9.

When mean annoyance and loudness ratings were plotted versus $R_w+C_{\mathrm{tr(50–5000)}}$ values, the extended frequency range led to different relationships than when the $C_{\mathrm{tr(100–3150)}}$ spectrum adaptation term was used. The R^2 values associated with the Boltzmann equation fits to the plots of mean ratings versus $R_w+C_{\mathrm{tr(50–5000)}}$ values are summarized in Table II. For both annoyance and loudness ratings of speech sounds the resulting R^2 values were reduced below those for $R_w+C_{\mathrm{tr(100–3150)}}$ values. For ratings of music sounds, the R^2 values increased a little for loudness ratings but decreased a little for annoyance ratings. That is, the extended frequency range in the $C_{\mathrm{tr(50–5000)}}$ weighting is not particularly helpful for predicting responses to speech and music sounds.

A modification of the C_{tr} spectrum weighting was created that simplified the shape of the weighting and enhanced the mid- and high-frequency weightings. This new weighting is referred to as $C_{\mathrm{tr\_mod}}$ and is illustrated in Fig. 9. When a Boltzmann equation was fitted to the plot of mean annoyance

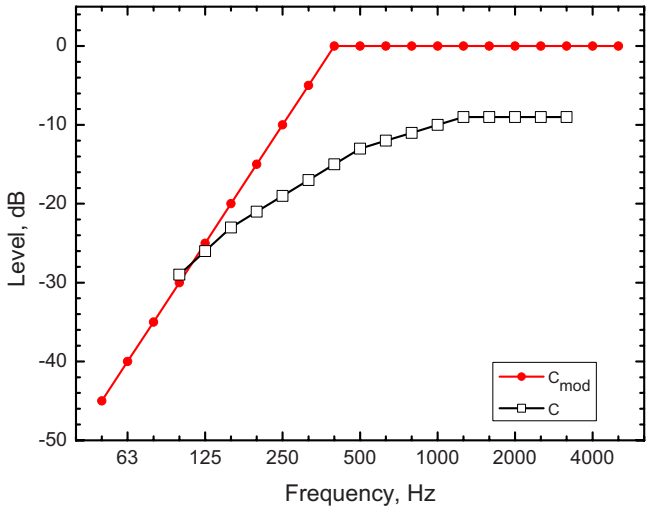


FIG. 10. (Color online) The modified C-type spectrum weighting term, C_{mod}, and the standard C-type spectrum weighting.

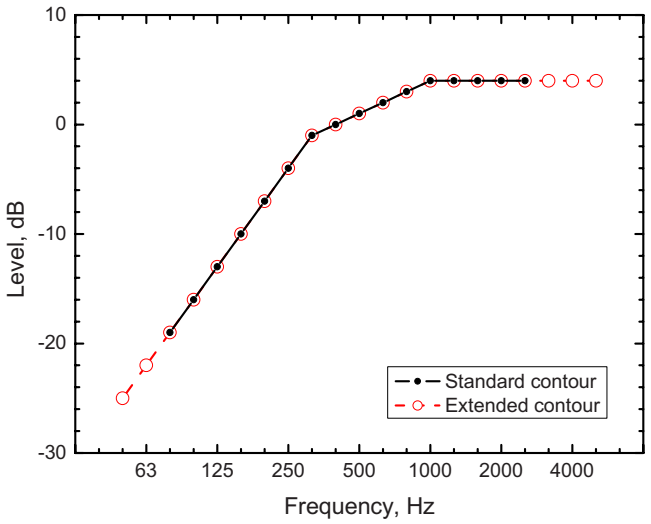


FIG. 11. (Color online) Open circle symbols show extensions to the standard R_w rating contour.

ratings of music sounds versus $R_w+C_{\text{tr\_mod}}$ values, the related R^2 value was 0.983. This was the highest R^2 value obtained for any relationship with ratings of the annoyance of music sounds. However, this combination of predictors was only moderately successful as a predictor to annoyance to speech responses with an R^2 of 0.541. The R_w+C_{mod} values were similarly successful as predictors of loudness ratings as they were for annoyance ratings.

Another modified spectrum adaptation term was produced to create a measure better related to responses to speech sounds. The new weighting, C_{mod}, is compared with the standard C-type spectrum adaptation terms in Fig. 10. The C_{mod} weighting emphasizes the higher frequencies most because the higher frequencies are more important for speech sounds. When mean annoyance ratings were plotted versus R_w+C_{mod} values, the new measure was seen to be a better predictor of ratings of speech sounds. For annoyance to speech sounds, the R^2 value associated with a Boltzmann equation fit was 0.975. This was the highest R^2 value for any predictor of responses to speech sounds in this paper. The summary of R^2 values in Table II shows that although R_w+C_{mod} values were a very successful predictor of responses to speech sounds, they were only a moderately successful predictor of ratings of annoyance to music sounds. The $R_w+C_{\text{tr\_mod}}$ values were a little better as predictors of loudness ratings of speech and music sounds (see Table II).

The merits of the various spectrum adaptation terms are summarized in Table II that lists the R^2 values associated with the various best fit Boltzmann equations to the plots of mean annoyance and loudness versus various combinations of R_w plus one of the spectrum adaptation terms. Only the spectrum adaptation term that limited the influence of the low frequencies (R_w+C_{mod}) improved R^2 values for ratings of speech sounds relative to that for just R_w values. All of the weightings that increased the effect of the low frequencies (R_w+C, $R_w+C_{\text{tr(100–3150)}}$, $R_w+C_{\text{tr(50–5000)}}$, and $R_w+C_{\text{tr\_mod}}$) improved the R^2 values for ratings of music sounds relative to those for R_w values. The most successful of these for music sounds was $R_w+C_{\text{tr\_mod}}$.

C. Variations in frequencies included in the R_w measure

Standard R_w values are based on a rating contour that extends from 100 to 3150 Hz. Because many results in this study have indicated that frequencies above and below this range may influence subjective ratings of the transmitted sounds, rating contours were evaluated that extended to both higher and lower frequencies than the standard one. Figure 11 shows the extensions to the rating contour, which were simple extrapolations of the slopes of the standard contour. The contour was extended to include frequencies from 63 to 6300 Hz. However, in some analyses more limited frequency ranges were considered.

Using a rating contour that included frequencies from 63 to 5000 Hz, $R_{w(63–5000)}$ values were found to lead to better predictions of responses to music sounds. When a Boltzmann equation was fitted to a plot of mean annoyance to music sounds versus $R_{w(63–5000)}$ values, the associated R^2 value was 0.940. However, for annoyance ratings of speech sounds $R_{w(63–5000)}$ values were a less successful predictor and the associated R^2 value was only 0.769. Extending the frequency range of the rating contour to include low frequencies improved predictions for sounds that included significant low-frequency content such as music but not for speech sounds.

By limiting the rating contour to include only speech frequencies (160–5000 Hz), a new rating, $R_{w(160–5000)}$, was created that was a better predictor of annoyance to speech sounds than the standard R_w measure. When Boltzmann equations were fitted to plots of mean annoyance versus $R_{w(160–5000)}$ values, a very high R^2 value resulted for annoyance to speech sounds ($R^2=0.972$) but not for music sounds ($R^2=0.562$).

To evaluate the many other possible frequency ranges using the same rating contour shape, linear correlations of annoyance ratings of speech and music sounds were made with $R_{w(f_1–f_2)}$ values. These are R_w values with the range of included 1/3 octave band frequencies extending from frequency f_1 to frequency f_2. All other aspects of the calculation

TABLE III. Correlation coefficients from correlations of mean annoyance ratings of music sounds with $R_{w(f_1-f_2)}$ values, where f_1 is the lowest included frequency and f_2 is the highest included frequency. The value in bold font shows the highest magnitude correlation coefficient.

		Upper frequency (f_2) (Hz)															
		200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000	6300
Lower frequency $\langle f_1 \rangle$ (Hz)	63	−0.97	−0.97	−0.97	**−0.98**	−0.97	−0.97	−0.96	−0.96	−0.96	−0.97	−0.96	−0.96	−0.96	−0.96	−0.96	−0.96
	80			−0.95	−0.94	−0.93	−0.92	−0.91	−0.91	−0.91	−0.91	−0.91	−0.93	−0.93	−0.93	−0.93	−0.93
	100				−0.90	−0.88	−0.86	−0.85	−0.85	−0.84	−0.84	−0.85	−0.87	−0.88	−0.88	−0.88	−0.88
	125					−0.78	−0.76	−0.74	−0.74	−0.73	−0.73	−0.76	−0.79	−0.81	−0.81	−0.81	−0.81
	160						−0.65	−0.61	−0.61	−0.61	−0.61	−0.64	−0.69	−0.74	−0.75	−0.75	−0.75
	200							−0.44	−0.43	−0.43	−0.44	−0.49	−0.58	−0.66	−0.70	−0.70	−0.70

of R_w were not changed. Table III lists the resulting correlation coefficients for annoyance to music sounds and Table IV gives the corresponding correlation coefficients for ratings of annoyance to speech sounds.

All of the correlation coefficients are negative because annoyance decreased as the sound insulation rating was increased. For annoyance to music sound responses, Table III shows that the larger magnitude correlation coefficients occurred when all low-frequency bands were included. Although the highest magnitude correlation coefficient occurred for the range from 63 to 400 Hz, the extended range from 63 to 5000 Hz led to a correlation coefficient that was only slightly smaller in magnitude. For annoyance ratings of speech sounds, Table IV shows that the highest magnitude correlation coefficients occur when the low-frequency bands were excluded. For an included frequency range from 160 to 5000 Hz, a correlation coefficient of −0.99 was obtained. Other similar frequency ranges led to the same value. These results indicate once again that the optimum range of included frequencies for predicting ratings of music sounds was different than that for speech sounds.

VI. AUDIBILITY OF TRANSMITTED SOUNDS

During the loudness rating test, subjects could give a score of zero if they could not hear any music or speech sounds. Therefore, the fraction of subjects scoring greater than 0 can be used as a measure of the audibility of the sounds. The point at which 50% of the subjects can just hear some transmitted test sound (i.e., with scores greater than 0) is referred to as the threshold of audibility for the sounds.

Figure 12 plots the fraction of subjects just able to hear at least some speech or music sound versus STC values separately for speech and music sounds. STC values are seen to be very good predictors of the audibility of speech sounds but not so good for predicting the audibility of music sounds. Figure 13 shows similar results when the same data are plotted versus R_w values. Both the standard sound insulation ratings, STC and R_w, were very good predictors of the audibility of speech sounds but are not as good for predicting the audibility of music sounds.

When spectrum adaptation terms were added to R_w values, predictions of the audibility of music sounds were improved (R_w+C: $R^2=0.757$, $R_w+C_{\mathrm{tr}(100-3150)}$: $R^2=0.920$) but predictions of the audibility of speech sounds were reduced in accuracy (R_w+C: $R^2=0.903$, $R_w+C_{\mathrm{tr}(100-3150)}$: $R^2=0.853$).

The standard sound insulation ratings STC and R_w are surprisingly good predictors of the audibility of speech sounds and with added spectrum adaptation terms can also be good predictors of the audibility of music sounds.

VII. SUMMARY AND DISCUSSION OF RESULTS

As shown in Figs. 2–5, R_w values were a little better as predictors of annoyance and loudness ratings of both speech and music sounds than were STC values. Both R_w and STC values were better predictors of responses to speech sounds than responses to music sounds. However, R_w and STC values were seen to be much more accurate predictors of the audibility of speech sounds than of annoyance and loudness ratings of these sounds.

When the 8-dB rule was removed from the calculation of STC values, the R^2 values in Table I showed that the resulting $\mathrm{STC}_{\mathrm{no8}}$ values were a much more accurate predictor of annoyance and loudness ratings of speech sounds than the standard STC values, but the opposite was true for ratings of music sounds.

Varying the maximum allowed deficiency from the standard 8-dB value indicated that reducing this value by several decibels would improve predictions of ratings of music

TABLE IV. Correlation coefficients from correlations of mean annoyance ratings of speech sounds with $R_{w(f_1-f_2)}$ values, where f_1 is the lowest included frequency and f_2 is the highest included frequency. The value in bold font show the highest magnitude correlation coefficient.

		Upper frequency (f_2) (Hz)															
		200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000	6300
Lower frequency $\langle f_1 \rangle$ (Hz)	63	−0.67	−0.71	−0.76	−0.81	−0.83	−0.85	−0.87	−0.87	−0.88	−0.88	−0.88	−0.88	−0.89	−0.89	−0.89	−0.89
	80			−0.82	−0.87	−0.89	−0.91	−0.92	−0.92	−0.93	−0.93	−0.93	−0.92	−0.92	−0.92	−0.92	−0.92
	100				−0.91	−0.93	−0.94	−0.95	−0.95	−0.96	−0.96	−0.96	−0.96	−0.95	−0.95	−0.95	−0.95
	125					−0.96	−0.97	**−0.98**	**−0.98**	**−0.98**	**−0.98**	**−0.98**	**−0.98**	**−0.98**	**−0.98**	**−0.98**	**−0.98**
	160						−0.95	−0.96	−0.96	−0.96	−0.96	**−0.98**	**−0.98**	**−0.99**	**−0.99**	**−0.99**	**−0.99**
	200							−0.91	−0.91	−0.90	−0.91	−0.93	−0.96	**−0.98**	**−0.99**	**−0.99**	**−0.99**

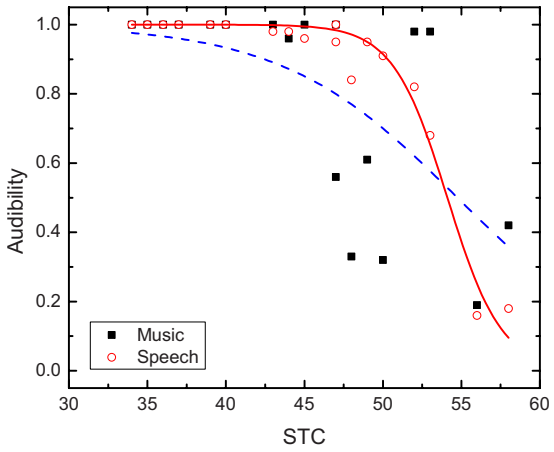


FIG. 12. (Color online) Fraction of subjects finding sounds just audible versus STC ratings (music: R^2=0.452; speech: R^2=0.968).

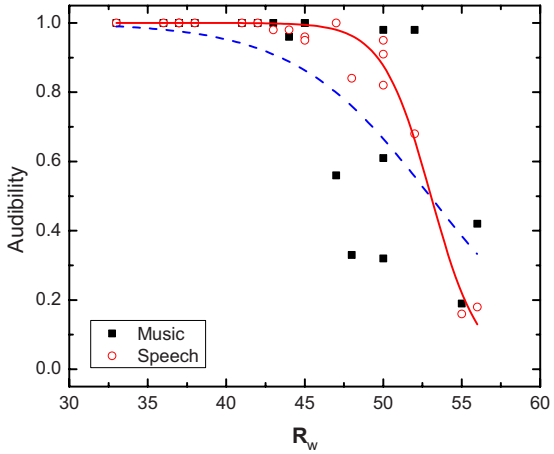


FIG. 13. (Color online) Fraction of subjects finding sounds just audible versus R_w ratings (music: R^2=0.526; speech: R^2=0.971).

sounds a little but would reduce the accuracy of predictions of the ratings of speech sounds a little. It could be argued that this would produce a better compromise for all types of sounds but the improvement would be small.

For the current wall transmission loss data, the 8-dB rule was applied at lower frequencies and mostly at 125 Hz. This is thought to be typical of many common constructions. However, if constructions could be found for which 8-dB limitations occurred at higher frequencies, the 8-dB rule might be beneficial for predictions of ratings of speech and other sounds without strong low-frequency content.

Varying the total allowed deficiency also had opposite effects for speech and music sounds. Reducing the total allowed deficiency to zero improved the accuracy of predictions of the ratings of music sounds but decreased the accuracy of the predictions of ratings of speech sounds. The standard value of 32 dB is more helpful for predicting ratings of music sounds than for those of speech sounds. If the 8-dB rule were removed, then the difference between the accuracy of ratings of music sounds and speech sounds would increase for all values of the total deficiency above zero.

Spectrum adaptation terms can be a successful means of increasing the prediction accuracy of particular types of transmitted sounds. Spectrum adaptation terms that increased the importance of lower-frequency sounds led to better predictions of annoyance, loudness, and audibility ratings of music sounds (e.g., R_w+C, $R_w+C_{\mathrm{tr(100–3150)}}$, $R_w+C_{\mathrm{tr(63–5000)}}$, $R_w\_C_{\mathrm{tr\_mod}}$). Conversely, spectrum adaptation terms that diminished the influence of low-frequency sounds led to improved predictions of ratings of speech sounds (e.g., R_w+C_{mod}). The best of these were extremely well related to ratings of one of the sound types but never very strongly related to both ratings of music and speech sounds.

An alternative approach in varying the range of included frequencies of transmission loss values and spectrum adaptation terms is to change the frequency range included in the rating contour. This was successful but not quite as good as the best spectrum adaptation terms and is less practical to implement than are new spectrum adaptation terms.

Comparison of ratings of the loudness and annoyance of transmitted sounds indicated that they provided similar information. Subjects seemed to be rating how much they did not like the sounds and annoyance and loudness ratings were used as parallel indicators of this dislike. However, audibility ratings tended to follow a different pattern of relationships.

The summary of the results of this paper in Tables V and VI of the Appendix shows that some combinations of R_w and spectrum adaptation terms were remarkably good predictors

TABLE V. Summary of regression coefficients and R^2 values for Boltzmann equation fits to ratings of music sounds in this paper. R^2 values equal to or greater than 0.95 are in bold font.

			Annoyance			Loudness			Audibility		
Measure	f_1	f_2	R^2	X_0	dx	R^2	X_0	dx	R^2	X_0	dx
STC	125	4000	0.728	47.916	5.667	0.734	43.417	5.275	0.452	54.742	5.593
R_w	100	3150	0.798	47.507	4.602	0.779	43.744	4.527	0.526	52.979	4.331
$\mathrm{STC_{no8}}$	125	4000	0.670	48.884	5.334	0.654	44.839	5.074	0.378	55.820	5.761
R_w+C	100	3150	0.918	44.624	4.371	0.900	40.591	4.638	0.757	49.789	2.910
R_w+C_{mod}	50	5000	0.580	37.605	5.313	0.556	33.735	5.106	0.297	44.594	5.949
$R_w+C_{\mathrm{tr(100–3150)}}$	100	3150	**0.950**	39.140	5.035	**0.960**	34.528	5.074	0.920	43.946	2.369
$R_w+C_{\mathrm{tr(50–5000)}}$	50	5000	0.943	36.812	5.753	**0.980**	31.574	4.543	0.948	43.143	1.878
$R_w+C_{\mathrm{tr\_mod}}$	50	5000	**0.983**	34.966	4.683	**0.991**	30.778	4.099	0.889	39.989	2.944
$R_{w\mathrm{(63–5000)}}$	63	5000	0.940	46.060	3.864	0.907	42.633	3.863	0.750	50.226	2.979
$R_{w\mathrm{(160–5000)}}$	160	5000	0.562	49.977	5.808	…	…	…	…	…	…

TABLE VI. Summary of regression coefficients and R^2 values for Boltzmann equation fits to ratings of speech sounds in this paper. R^2 values equal to or greater than 0.95 are in bold font.

			Annoyance			Loudness			Audibility		
Measure	f_1	f_2	R^2	X_0	dx	R^2	X_0	dx	R^2	X_0	dx
STC	125	4000	0.856	43.955	7.018	0.886	39.322	6.906	**0.968**	54.106	1.731
R_w	100	3150	0.890	43.946	6.176	0.933	39.891	6.013	**0.971**	53.045	1.551
STC_{no8}	125	4000	**0.950**	45.271	5.804	**0.970**	41.299	5.842	0.919	54.915	1.830
R_w+C	100	3150	0.741	40.552	7.459	0.821	36.050	6.818	0.903	50.763	1.777
R_w+C_{mod}	50	5000	**0.975**	34.198	4.994	**0.973**	30.716	5.142	0.863	42.877	1.720
$R_w+C_{tr(100–3150)}$	100	3150	0.566	34.370	9.529	0.676	29.197	8.063	0.853	46.252	0.465
$R_w+C_{tr(50–5000)}$	50	5000	0.388	32.276	11.912	0.482	26.447	9.560	0.856	44.888	0.090
$R_w+C_{tr\_mod}$	50	5000	0.541	31.084	8.840	0.634	26.422	7.167	0.866	42.593	0.945
$R_{w(63–5000)}$	63	5000	0.769	42.623	6.361	0.848	38.816	5.701	0.871	52.353	1.493
$R_{w(160–5000)}$	160	5000	**0.972**	46.301	5.359	…	…	…	…	…	…

of responses to one of the two types of sounds. However, no measure provided extremely accurate predictions of ratings of both speech and music sounds. The values of the R_w measure were a little better than STC values as predictors of responses and adding appropriate spectrum adaptation terms for specific types of sounds to R_w values would be a practical approach for improving the accuracy of predictions of subjective ratings of sound insulation. For example, this could include separate corrections for speech and music sounds.

If one looks for measures that were strongly related to all three types of responses (annoyance, loudness, and audibility) there were very few. If we define "strongly related" as corresponding to R^2 values of 0.9 or greater, only $R_w+C_{tr100–3150}$ and $R_w+C_{tr50–5000}$ values were strongly related to all three responses to music sounds and $R_w+C_{tr\_mod}$ values almost met this criterion. Only STC_{no8} values met this criterion for response to speech sounds and R_w+C_{mod} values almost met it.

These results are based on 20 transmission loss characteristics representative of standard laboratory test results of common wall constructions. In many real buildings the existence of flanking paths tends to limit the higher transmission loss values at mid- and higher frequencies.[20,21] Such modified characteristics or a different selection of wall transmission loss characteristics might lead to slightly different results. However, it seems unlikely that such differences would significantly change the overall trends.

VIII. CONCLUSIONS

The standard STC and R_w measures were not the best predictors of annoyance and loudness ratings of speech and music sounds. Although they did predict the audibility of speech sounds very well, this was not true for music sounds.

The standard R_w values can be improved by the addition of different spectrum adaptation terms for speech and music sounds. $R_w+C_{tr100–3150}$ and $R_w+C_{tr50–5000}$ values were good predictors of all responses to music sounds. R_w+C_{mod} values were good predictors of annoyance and loudness ratings of speech sounds. When the 8-dB rule was excluded from STC values, the resulting STC_{no8} values were good predictors of all three responses to speech sounds.

ACKNOWLEDGMENTS

The first author's contribution to this work was supported by a grant from the Korean Ministry of Education, Science and Technology (The Regional Core Research Program/Biohousing Research Institute) and a Korean Research Foundation Grant (MOEHRD) (KRF-2006-352-D00200). They would also like to thank Dr. Brad Gover for helpful discussions during this project.

APPENDIX: BOLTZMANN EQUATION REGRESSION DETAILS

This paper includes many regression fits of Boltzmann type equations [described in Eq. (1)] to plots of mean subjective ratings versus sound insulation measures. These included annoyance, loudness, and audibility ratings of both speech and music sounds. The regression coefficients describing the key regression equations discussed in this paper are summarized in the tables of this Appendix. Table V includes the regression coefficients and R^2 values for ratings of music sounds. Table VI includes the regression coefficients and R^2 values for ratings of speech sounds.

[1]ASTM E90, "Standard test method for laboratory measurement of airborne sound transmission loss of building partitions and elements," ASTM International, West Conshohocken, PA.

[2]ASTM E336, "Standard test method for measurement of airborne sound insulation in buildings," ASTM International, West Conshohocken, PA.

[3]ISO 140, "Acoustics—Measurement of sound insulation in buildings and of building elements"—Part 3, "Laboratory measurement of airborne sound insulation of building elements" (2004), Part 4, "Field measurements of airborne sound insulation between rooms" (1998).

[4]ASTM E413, "Classification for rating sound insulation," ASTM International, West Conshohocken, PA.

[5]ISO-717-1, "Acoustics—Rating of sound insulation in buildings and of building elements—Part 1: Airborne sound insulation," International Organization for Standardization (1996).

[6]H. K. Park, J. S. Bradley, and B. N. Gover, "Evaluation of airborne sound insulation in terms of speech intelligibility," IRC Research Report No. RR-228, National Research Council, Canada, March 2007.

[7]H. K. Park, J. S. Bradley, and B. N. Gover, "Evaluating airborne sound insulation in terms of speech intelligibility," J. Acoust. Soc. Am. **123**, 1458–1471 (2008).

[8]ANSI S3.5-1969, "American National Standard Methods for the Calculation of the Articulation Index," Standards Secretariat, Acoustical Society of America, New York.

[9]ANSI S3.5-1997, "Methods for calculation of the speech intelligibility

index," American National Standard, Standards Secretariat, Acoustical Society of America, New York.
[10]ASTM E1110, "Standard classification for determination of articulation class," ASTM International, West Conshohocken, PA.
[11]C. Grimwood, "Complaints about poor sound insulation between dwellings in England and Wales," Appl. Acoust. **52**, 211–223 (1997).
[12]C. J. Grimwood and N. J. Tinsdeall, "Occupant opinion of sound insulation in converted and refurbished dwellings in England and the applications for national building regulations," in Proceedings of the Inter Noise 98, Christchurch, NZ (1998), pp. 1727–1732.
[13]J. S. Bradley, "Deriving acceptable values for party wall sound insulation from survey results," in Proceedings of the Inter Noise 2001 (2001), pp. 1505–1510.
[14]J.-P. Vian, W. F. Danner, and J. W. Bauer, "Assessment of significant acoustical parameters for rating sound insulation of party walls," J. Acoust. Soc. Am. **73**, 1236–1243 (1983).
[15]D. M. Clark, "Subjective study of the sound-transmission class system for rating building partitions," J. Acoust. Soc. Am. **47**, 676–682 (1970).
[16]J. H. Rindel, "On the influence of low frequencies on the annoyance of noise from neighbours," in Proceedings of the Inter Noise 2003 (2003), pp. 1500–1503.
[17]IEE Recommended Practice for Speech Quality Measurements, IEEE Trans. Audio Electroacoust. **AU–17**, 225–246 (1969).
[18]ISO 7029-1984 (E), "Acoustics—Threshold of hearing by air conduction as a function of age and sex for otologically normal persons," ISO, Geneva, Switzerland.
[19]H. K. Park, J. S. Bradley, and B. N. Gover, "Rating airborne sound insulation in terms of the annoyance and loudness of transmitted speech and music sounds," NRC/IRC Research Report No. RR-242, National Research Council, Canada, November 2008.
[20]R. J. M. Craik, T. R. T. Nightingale, and J. A. Steel, "Sound transmission through a double leaf partition wall with edge flanking," J. Acoust. Soc. Am. **101**, 964–969 (1997).
[21]T. R. T. Nightingale, J. D. Quirt, F. King, and R. E. Halliwell, "Flanking transmission in multi-family dwellings: Phase IV," Institute for Research in Construction Research Report No. RR-218, National Research Council, Canada, March 2006.

Interference suppression for code-division multiple-access communications in an underwater acoustic channel

T. C. Yang[a)]
Naval Research Laboratory, Code 7120, 4555 Overlook Avenue, Washington, DC 20375

Wen-Bin Yang
National Institute of Standards and Technology, Stop 8920, 100 Bureau Drive, Gaithersburg, Maryland 20899

(Received 7 January 2009; revised 7 May 2009; accepted 7 May 2009)

In a code-division multiple-access communications network, the signal from a nearby user often creates a strong interference for the signal from a distant user. This is known as the near-far problem. Power control of source levels is ineffective in an underwater acoustic channel due to the slow sound speed. Interference rejection based on code orthogonality is ineffective using matched-filter processing due to the fact that multipath arrivals effectively destroy the code orthogonality and that the signal arrival times between different users are not synchronized. An algorithm, called hyperspace cancellation by coordinate zeroing, is used in this paper to remove/suppress interference. Using a fast Walsh–Hadamard transform (FWHT) based on the interferer's code sequence, the interference signal is enhanced and removed by coordinate zeroing. The residual signal is transformed back using an inverse FWHT. The filtered data, with the interference signal largely removed, are processed using the desired signal code sequence. Two methods previously developed for direct-sequence spread-spectrum communications in an underwater channel are used to extract the transmitted symbols. Low bit error rate ($<10^{-2}$) is found with the at-sea data for signal-to-interference ratio as low as -8 to -11 dB. [DOI: 10.1121/1.3147484]

PACS number(s): 43.60.Dh [EJS] Pages: 220–228

I. INTRODUCTION

For underwater acoustic sensors, either fixed or mobile, to work collaboratively, an underwater acoustic network is required. When the network traffic is high, such as for command and control of mobile platforms, and/or during periods of high activities, a multi-access communications network is desired.[1–4] When the propagation channel is time varying or when the source/receiver is moving, the travel time of the signals cannot be precisely controlled. The probability that more than one message will arrive at the receiver is not insignificant. In that case, the signal from a nearby transmitter (interferer) could mask the desired signal from a more distant transmitter, making it difficult for a receiver to decode the desired message. This is known as the near-far problem. In radio-frequency (rf) communications, a feedback message can be sent to the far-away transmitter to increase its source level. (The alternative will be to lower the interferer's power which may degrade the signal quality of the interferer's signal.) This is known as a closed-loop power control. Instantaneous power control is not possible in an underwater acoustic channel due to the slow sound speed. The reasons are twofold. First, the two-way travel time, before the feedback signal is received by the transmitter, is usually much longer than the signal duration (packet). Second, the signal often fluctuates by 5–10 dB over the packet duration, making power control ineffective. While one can still exercise some limited power control, substantial interference signal will remain and need to be removed by signal processing.

Using direct-sequence spread-spectrum (DSSS) signaling schemes where different users are assigned different spreading codes, one arrives at a multiple-access communication network, known as direct-sequence code-division multiple-access (DS-CDMA). In DSSS signaling, the information symbols are coded/multiplied with a code sequence, commonly known as chips. The signals are processed at the receiver using the code sequence as a matched filter to extract the information symbols. Using the processing gain derived from the matched filter, the desired signal can be enhanced against the noise. The desired signal can be discriminated from the interference signal if the spreading codes are orthogonal between the different users. Code "orthogonality" requires that the code sequence be almost orthogonal to any of the cyclically-shifted code sequences and to the code sequence of other users. With orthogonality, the matched-filtered output yields a low sidelobe level and thus ensures minimum interference between symbols as well as between signals of different users. This also assures accurate symbol synchronization.

Without code orthogonality, interference from other users, referred to as co-channel interference or multi-access interference (MAI), can induce significant bit errors for the desired signal. Symbol synchronization of the desired signal can be a problem. Multiuser detectors (MUDs) were proposed in rf communications which perform joint symbol detection and channel estimation for all users. The interference signal is regenerated based on the detected symbols and an

[a)]Author to whom correspondence should be addressed. Electronic mail: yang@wave.nrl.navy.mil

estimation of the channel impulse responses (CIRs) from the interfering sources and the signal power and subtracted from the received data to extract the desired signal. This is often done iteratively, either in sequence or in parallel, known as multistage successive interference cancellation (SIC)/parallel interference cancellation.[5,6] Another method minimizes the interference by finding the independent components of the signals by de-correlation or eigenvalue decomposition.[7] This is known as independent component analysis (ICA).

The underwater acoustic channel is known for extended multipath arrivals. The received symbols are no longer orthogonal (between themselves and between different users) when multipaths are present and not corrected for; code orthogonality is effectively destroyed by multipaths. Stojanovic and Freitag[1] proposed a single user detector based on hypothesis-feedback-equalization for suppression of multipath interference and dispreading (de-correlation) for suppression of MAI. Hypothesis-feedback assumes that the decision symbols are sufficiently reliable to perform symbol or chip-resolution decision-feedback equalization (DFE). Bit errors accumulate/propagate when the hypothesized symbols are in error, as when a strong interferer is nearby. Several MUDs were also proposed for underwater acoustic networks. Stojanovic and Zvonar[8] used crossover feedback filters to cancel out the MAI. Calvo and Stojanovic[3] proposed a minimum-mean-square-error detector for the transmitted symbols and signal parameters, where the cyclic descent method was used to find a local minimum of the cost function. Yeo *et al.*[9,10] applied recursive SIC to underwater acoustic communications. MUDs inherently require complex computations. Interference removal requires coherent processing (summing) of interference signals (including the signal phase). Bit error propagation can be severe when the CIRs (or equalizer coefficients) and signal power are inaccurately estimated. Symbol synchronization has to be done at the chip level.

Motivated by the need to communicate at low input signal-to-noise ratio (SNR), two methods have been proposed for DSSS signals in an underwater acoustic channel without requiring DFE. The first method cross-correlates the matched-filter outputs to determine the relation between adjacent symbols.[11] It will be referred to as the cross-correlation method in this paper. It works well for fixed communication nodes. The second method uses two transition energy detectors to determine the relation between the adjacent symbols.[12] It will be referred to as the energy-detector method. It works for both fixed and mobile nodes. These methods are simple and use only matched-filter processing. They have been demonstrated to work for input SNRs as low as -8 to -10 dB.

To mitigate the co-channel interference in the presence of multipaths, a method, called hyperspace cancellation by coordinate zeroing (HCC0),[13,14] is proposed and demonstrated in this paper. Compared with previously proposed methods, this method does not require channel estimation (the interferer's signal level and CIR). The method removes a strong (jamming) interference signal without affecting the desired signal by projecting the received signal into the interference signal-space using the HCC0 method. (It is similar to the eigenvalue decomposition as used in ICA, except that the eigenvectors are not the interference signals when the signals are not orthogonal.) When the interference signal is brought down to the level of the desired signal, the matched-filter (the cross-correlation and the energy detector) methods are often adequate to extract the desired signal with minimum (uncoded) bit errors. When multiple jamming signals are present, it can be removed one by one *independently*. Symbol synchronization is done at the symbol level. This method is a single user detector and involves simple computations.

This paper is organized as follows. The DS-CDMA system is briefly described in Sec. II. The HCC0 method is discussed in Sec. III, and the use of the fast Walsh–Hadamard transform (FWHT) in the Appendix. In Sec. IV, multi-user data collected at sea are processed using the HCC0 method, followed by the cross-correlation method. In Sec. V, the multi-user data are processed using the HCC0 method, followed by the energy-detector method. Section VI gives a summary of the results. Only one jamming signal is assumed in this paper.

Two effects of the interference removal are addressed: the accuracy of symbol synchronization and bit error rate (BER). Since symbol synchronization is often based on the probe signal or the beginning of the data symbols, the accuracy depends to a great extent on whether these components of the signal are masked by the interference signal, which in turn depends on the arrival time difference between the interference signal and desired signal. BER depends on the success of interference removal, which is affected by the interference between the interference signal and desired signal. BER varies depending on the phase or the time offset between the signals. To evaluate the performance, one needs to consider different arrival time offsets in the data analysis to derive an averaged BER. Thirty packets were used in the analysis.

II. SYSTEM DESCRIPTION

For a multi-access networking system using DS-CDMA signaling, K users could be simultaneously transmitting in the same band, where the kth user is assigned a spreading code, $c_l^k, l=0,1,\ldots,L-1$, with spreading factor L and chip duration T_c. The signal transmitted by the kth user (transmitter) is given in the baseband by

$$S^k(t) = \sum_{i=1}^{N} d_i^k C^k(t - iT_s), \tag{1}$$

where d_i^k is the ith symbol transmitted by user k; $C^k(t) = \sum_{l=0}^{L-1} c_l^k g(t - lT_c)$, g is a raised cosine or rectangular window centered at lT_c with a width of T_c; T_c is the chip duration; and T_s is a symbol duration ($T_s = L^* T_c$).

Let us study the interference when K users are transmitting at the same time. For the purpose of discussion, consider the received signal for one symbol, the ith symbol period. Let $d_i^k = 1$, and assume a continuous representation of the spreading code. One finds

$$r_i^k(t) = s_i^k C^k(t) \otimes h^k(t) \quad \text{for } iT_s \le t < (i+1)T_s, \tag{2}$$

where s_i^k is the received signal strength and $h^k(t)$ is the CIR from the kth transmitter to the given receiver. The total received signal at a given receiver is the sum of all the signals, i.e.,

$$r_i(t) = \sum_{k=1}^{K} r_i^k(t - \tau_k) + w(t), \tag{3}$$

where τ_k is the time delay with respect to a reference time, and $w(t)$ is the ambient noise at the receiver. Maximum interference occurs when all signals arrive at the receiver at the same time, i.e., $\tau_k = 0$, for all k, which will be assumed to be the case.

If one is interested in decoding the signal from user 1, treating the signals from users $2, \ldots, k$ as interference, the conventional approach is to use the code sequence of user 1 to extract the desired signal,

$$S_i^1 = C^1(t) \cdot r_i(t) = \sum_{k=1}^{K} s_i^k \rho^{1,k}(t) \otimes h^k(t) + n^1(t), \tag{4}$$

where $\rho^{1,k}(t) = C^1(t) \cdot C^k(t)$ is the cross-correlation function of the first code sequence with the kth code sequence ($\cdot$ denotes the cross-correlation operator) and $n^1(t) = C^1(t) \cdot w(t)$. If the code sequence of user 1 is orthogonal to the code sequences of other users, $\rho^{1,k}(t) = 0$ when $k \neq 1$. In other words, the interference from users 2 to K in Eq. (3) equals zero. But in reality, $\rho^{1,k}(t) \neq 0$ when $k \neq 1$. Hence, interference is nonzero in Eq. (3). This creates the near-far problem when the interference level is higher than the desired signal level.

In this paper, m-sequence is used as the code sequence. The m-sequence, with m integer, has a nice property that its auto-correlation function has a value $L = 2^m - 1$ at the peak and ± 1 everywhere else. For a given m, there exist many (different) m-sequences, but not all of them have good (near-orthogonal) cross-correlation properties. For a large value of m ($m \ge 9$) one can find many m-sequences created by selected law of code generation (LAW) which produces good cross-correlation properties. (This is not true in general for other LAWs.) Figure 1 shows such an example for $m = 9$. It shows the auto-correlation of the user-1 code sequence and the cross-correlation of the user-1 code sequence with the user-2 code sequence. One finds that the maximum value of the cross-correlation is 20 dB down from the peak value of the auto-correlation. Hence, the matched filter using m-sequence codes is able to suppress interference by ≥ 20 dB, as is the case when there is only a single path. In other words, if the interference is 10 dB higher than the desired signal in the received data, the interference is 10 dB lower than the desired signal in the matched-filter output. In this case the desired signal can be processed with minimal or no error. Figure 1 serves as the reference to compare with the case of extended multipath arrivals.

The inter-user (the co-channel) interference becomes significant in a multipath environment. To illustrate the origin of the interference, it is shown in Fig. 2(a) the correlation of the user-1 code sequence with data transmitted from user 1, denoted by CIR_1, the CIR for user 1. Likewise, Fig. 2(b) shows the correlation of the user-2 code sequence with data transmitted from user 2, denoted by CIR_2, the CIR for user 2. (The origin of the data is given in Sec. IV.) Figure 2(a) also shows the cross-correlation of the user-1 code sequence with data transmitted from user-2; this represents the interference to user 1 from user 2. Since symbol acquisition requires integrating the CIR over the multipaths, the amount of interference increases with the length of the multipath spread. In Fig. 2, the signal-to-interference ratio (SIR) is 0 dB (i.e., equal strength). One can see that the interference can prevent the decoding of the user-1 message when the interference is ten times larger than the desired signal, or SIR $= -10$ dB, as opposed to -20 dB for the signal path case. (The increase in interference due to extended multipaths can be viewed as an increase in interfering sources each arriving by signal path with a strength equal to the path amplitude.) Co-channel interference indeed creates a high BER, as will be shown by the at-sea data in Secs. IV and V. Correspondingly, to achieve a minimal BER ($<10^{-2}$), additional signal processing will be needed to suppress the interference which is discussed in Sec. III, using the HCC0 method. It is noted that multi-user codes cannot all have zero cross-correlation;[15] co-channel interference is always present.

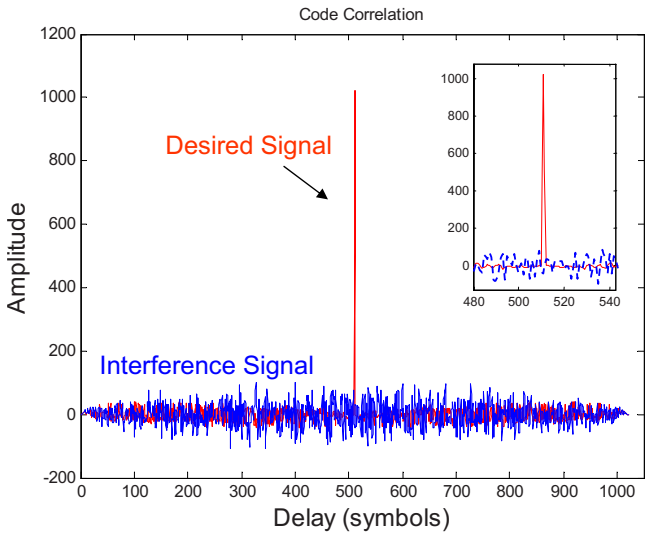


FIG. 1. (Color online) Code correlations between desired signal with itself (auto-correlation) and with the interference (cross-correlation).

III. INTERFERENCE ESTIMATION AND SUPPRESSION

The HCC0 method estimates the interference signal due to the kth transmitter using cyclic cross-correlation. To simplify the illustration, let the ith symbol data be represented by L samples, i.e., one sample per chip, in the form of a vector $\underline{r}_i$. The cyclic cross-correlation may be written in matrix form as follows:

$$\hat{\underline{r}}_i^k = M_L^k \underline{r}_i, \tag{5}$$

where M_L^k is an m-sequence matrix consisting of cyclic rotation of the kth code sequence $[c_0^k, c_1^k, \ldots, c_{L-2}^k, c_{L-1}^k]$ as given below

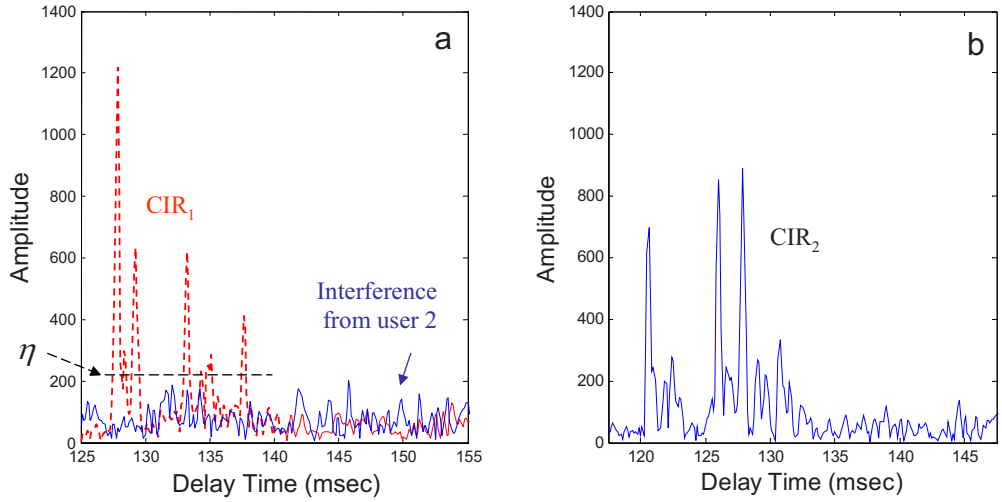


FIG. 2. (Color online) (a) Correlation of code 1 with data from user 1 (denoted by dashed curve and CIR_1) and user 2 (solid curve). (b) Correlation of code 2 with data from user 2 (CIR_2).

$$M_L^k = \begin{bmatrix} c_0^k & c_1^k & \cdots & c_{L-2}^k & c_{L-1}^k \\ c_{L-1}^k & c_0^k & \cdots & c_{L-3}^k & c_{L-2}^k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_2^k & c_3^k & \cdots & c_0^k & c_1^k \\ c_1^k & c_2^k & \cdots & c_{L-1}^k & c_0^k \end{bmatrix}. \tag{6}$$

The cyclic cross-correlation works like the conventional cross-correlation, in which the signal from the kth transmitter is enhanced by the matched-filter gain, whereas the signal components from the other transmitters, including the desired signal, are not. The signal from the kth transmitter (the interference signal) may then be removed by zeroing out the signal components (in $\hat{r}_i^k$) having an amplitude higher than a certain threshold, η. The signals from other transmitters, including the desired signal, having amplitudes below η, remain intact. This operation produces the following signal:

$$\breve{r}_i(l) = \begin{cases} \hat{r}_i^k(l) & \text{if } |\hat{r}_i^k(l)| < \eta \\ 0 & \text{elsewhere} \end{cases} \quad \text{for } l = 0,1,\ldots,L-1. \tag{7}$$

The signal $\breve{r}_i$, with the interference signal largely removed, is now transformed back using the inverse of the original matrix, M_L^k, as follows:

$$\tilde{r}_i = (M_L^k)^{-1} \cdot \breve{r}_i. \tag{8}$$

The desired signal can then be estimated by matched filtering the data $\tilde{r}_i$ with the code sequence of the desired source. The above processing is repeated for all N symbols, $i = 1,2,\ldots,N$.

Equation (5) can be illustrated graphically using Fig. 2(a). For this discussion, the signal from user 2 is the desired signal and the signal from user 1 is the interference. Figure 2(a) shows the result of Eq. (5) for both users, where η is set such that the signal from user 1 is mostly clipped, whereas the (transformed) signal from user 2 is unaffected. The signal from user 2 can then be recovered using the inverse transform.

Figure 2(a) can be interpreted as follows. Using Eq. (4) one has

$$S_i^1 = C^1(t) \cdot r_i(t) = s_i^1 \rho^{1,1}(t) \otimes h^1(t) + s_i^2 \rho^{1,2}(t) \otimes h^2(t) + n^1(t),$$

After cross-correlation, the interference signal (the first term of the above equation) is shown in Fig. 2(a) by the dashed curve, denoted by CIR_1, and the desired signal (the second term of the above equation) is shown by the solid curve. The threshold η should be set such that the desired signal is not affected by the hard clipping. Hence

$$\eta \geq \max(s_i^2 \rho^{1,k}(t) \otimes h^2(t)) \simeq \max(|r_i^k|)\max(\rho^{1,k}),$$

$k = 2$ the desired signal.

The interference signal with amplitude $|r_i^1|$ greater than $\eta/\rho^{1,1}$ is removed.

In practice, a chip may be fractionally sampled with m samples per chip. The ith symbol data $\underline{r}_i = [r_i(1), r_i(2), r_i(3), \ldots, r_i(mL)]$ may be reshaped to an $(L \times m)$ matrix, where the jth column vector of the matrix is given by $\underline{r}_{i,j} = [r_i(j), r_i(m+j), r_i(2m+j), \ldots, r_i((L-1)m+j)]^T$, $j = 1,2,\ldots,m$, where the superscript T denotes transpose. The above processing is repeated for each jth column vector. The results are recombined into

$$\tilde{R}_i = [\tilde{r}_{i,1} \quad \tilde{r}_{i,2} \quad \cdots \quad \tilde{r}_{i,m}] \tag{9}$$

and reshaped into a row vector. The final data are matched filtered with a fractionally sampled code sequence for user 1.

The above processing requires the calculation of an inverse matrix which can be computationally complex when the code sequence is large. One notes that a long code sequence is often required to achieve orthogonality between code sequences at arbitrary delay time. The m-sequence codes are used in this paper where the cyclic correlation can be performed using the FWHT[16,17] similar to the fast Fourier transform. The FWHT algorithm is briefly described in the Appendix.

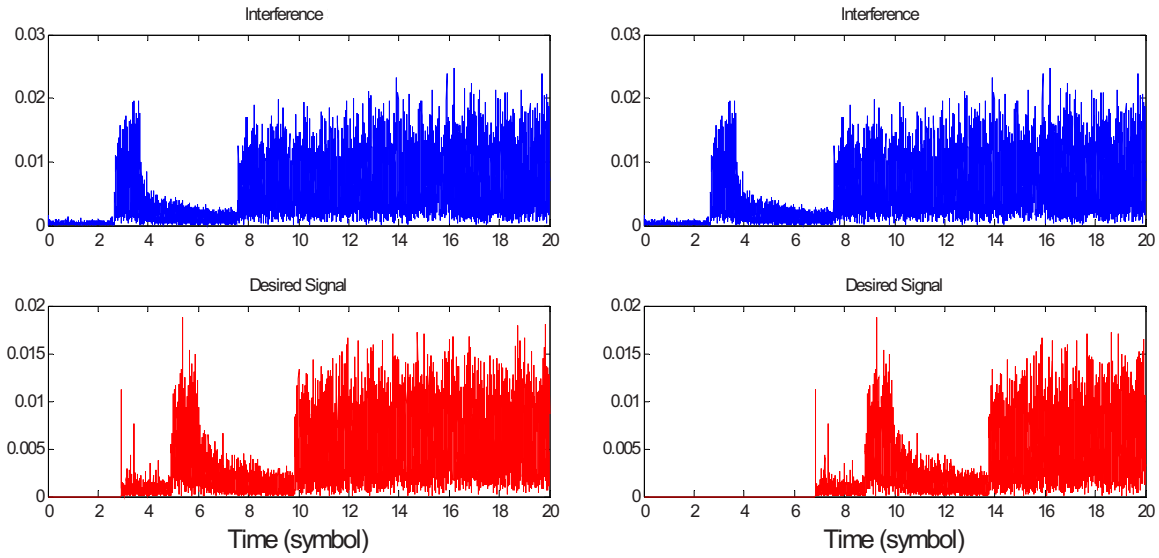


FIG. 3. (Color online) Time series showing different time offsets of desired signal and interference signal.

IV. DATA PROCESSING USING THE CROSS-CORRELATION METHOD: WITH AND WITHOUT INTERFERENCE REJECTION

The UNet06 experiment was conducted in the St. Margaret bay outside of Halifax, Canada in May 2006. The water depth was about 60 m. The sound speed profile presented a downward refractive channel. The sea was relatively calm. The source was at a depth of ~21 m and the receiver was at a depth of ~30 m. Only one of the eight available receivers on a vertical line was used. About 20 packets of DS-CDMA data were transmitted for each user with its own code sequence. The signal was centered at 17 kHz with a bandwidth of 4 kHz. Treating the data from one user as the desired signal and the other as the interference signal, the data are added together in postanalysis, with SIR varying from 0 to −12 dB to study the effect of interference on the BER of the desired signal. Each packet consists of a probe signal (a single *m*-sequence), followed by a gap and then followed by data consisting of 200 symbols. Each symbol is spread with 511 chips.[18]

The desired signal and interference signal can be offset in arrival time, as shown in Fig. 3. In Fig. 3(a), the probe signal used for symbol synchronization for the desired signal is largely uncontaminated by the interference signal; this will be referred to as case A. In this case, symbol synchronization is not an issue. In Fig. 3(b), the entire signal, including the probe signal, is contaminated by the interference signal; this will be referred to as case B. In this case, symbol synchronization becomes a problem for low SIR cases. The authors have also considered another case, i.e., where the two signals arrive at almost the same time and hence interfere maximally with each other. As the results are not much different than case B, this case will not be discussed in detail. For each case, in addition to the mean signal offset, 17 packets are generated with an additional near-random offset of up to one symbol period. The BER results presented below are averaged over 17 packets of data.

The data are processed in the baseband and sampled at two samples per chip. CIRs for the desired signal and interference signal were estimated from their respective probe signals and are shown in Figs. 2(a) and 2(b) to display their multipath structures.

The data which contain the sum of the desired signal and interference signal (with different SIRs) are first processed using the cross-correlation method,[11] as shown in Fig. 4. (For these data, the Doppler shift is near zero so that the cross-correlation method can be applied.) To decode the bits of the desired signal, the data are first matched filtered using the code sequence of the desired signal, symbol by symbol. The outputs of the matched filter are cross-correlated between adjacent symbols to determine their relation (or their relative phase, using differential phase shifting keying). The results of case A are shown in Fig. 5 by the diamonds. One sees a high BER. The reason is that the cross-correlation includes the co-channel interference covering the entire symbol length. To reduce the co-channel interference, the matched-filter outputs are gated to include only the portion covering the multipath spread of approximately 100 ms. The BER using the cross-correlation method is shown as "x" in Fig. 5. One finds that the BER improves significantly to $<10^{-2}$ for

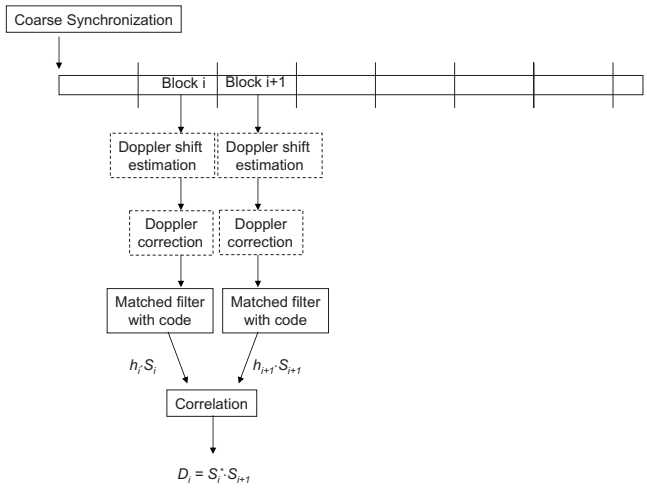


FIG. 4. Schematics of the cross-correlation method.

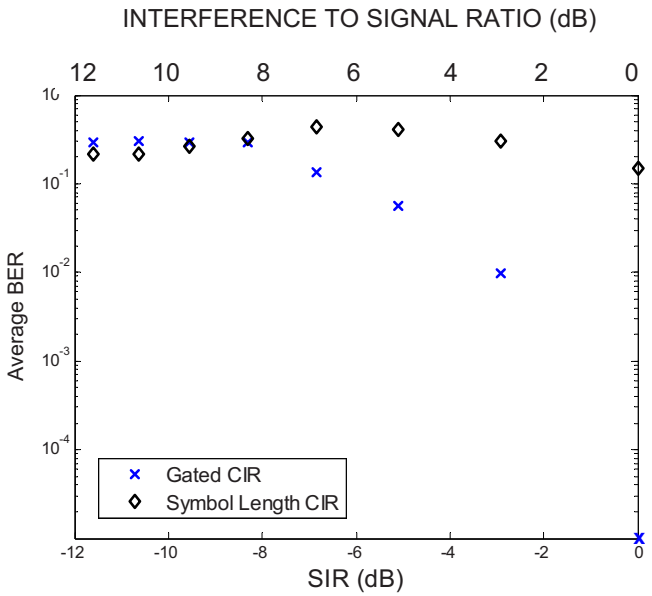


FIG. 5. (Color online) BER using the cross-correlation method with different lengths of CIR. Data averaged over 17 packets with time offsets as shown in Fig. 1.

SIR >-3 dB. For SIR between -3 and -8 dB, while there is some improvement, the BER is unacceptably high. For SIR <-8 dB, there is practically no improvement. To further improve BER when SIR <-3 dB, additional signal processing will be used to remove/suppress as discussed next.

To suppress the interference, the data are first processed using the HCC0 method and then processed using the cross-correlation method to decode the bits of the desired signal. As discussed above, the threshold value η should be set, in principle, slightly above the level of the matched-filter output of the desired signal with the interferer's code sequence, i.e., $\eta \geq \max(|r_i^1|(l))\max(\rho^{1,k})$. But since the level of the desired signal is hard to estimate when the interference is strong, η should be set in accordance with the interference level, i.e., $\eta \approx \max(|\hat{r}_i^k|(l))\alpha$, where α is a fractional number. (Note that the interference may be estimated accurately when the interference signal is strong.) Figure 2(a) suggests that the higher the interference-to-signal ratio, the smaller α should be in order to remove more of the interference signal. Since the SIR is not known *a priori*, a constant symbol α will be used. The data are processed with $\alpha = 1/7$.

Figures 6(a) and 6(b) show the BER with interference rejection and without interference rejection (raw data), using cross-correlations of gated CIR for cases A and B, respectively. One notes that without interference rejection, the case B result is significantly worse than the case A result; the difference is largely due to the symbol-synchronization error in the presence of strong interference for case B. After interference rejection, one finds in Fig. 6 that the BERs are similar for both cases, which is expected. With interference rejection, one finds that the (uncoded) BER is $<10^{-2}$ for SIR as low as -9 dB. As the range between different transmitters may differ (only) by a factor of ~ 2 or less within a group of local nodes (or ~ -3 dB in SIR), for most practical scenarios, one may be able to reject interference from several simultaneous transmitters.

V. DATA PROCESSING USING THE ENERGY-DETECTOR METHOD: WITH AND WITHOUT INTERFERENCE REJECTION

In this section the authors repeat the data processing, now using the energy detection method which can be applied to moving source data.[12] The energy detection method is shown in Fig. 7. It is assumed that the Doppler shift has been estimated and corrected for the data. (Doppler estimation for *m*-sequence signals from a moving source was discussed in Ref. 12. The measurements yield a Doppler resolution of 1 Hz or less.)

The energy-detector method uses two transition code sequences to match filter the data. The energies of the matched-filter outputs are compared between the two code sequences to determine the relationship between the adjacent symbols. As shown in Fig. 7, after initial symbol synchronization, the data are divided into blocks, each block having the length of

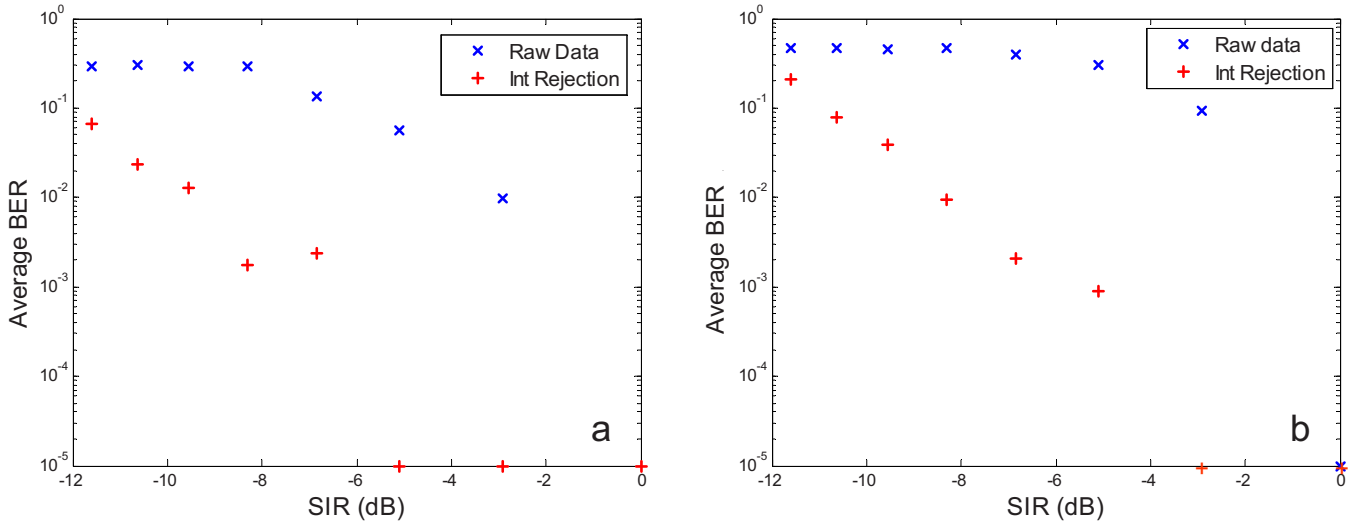


FIG. 6. (Color online) Average BER using the cross-correlation method with and without interference rejection for case A (a) and case B (b). Zero BER is represented by 10^{-5} on the logarithmic scale.

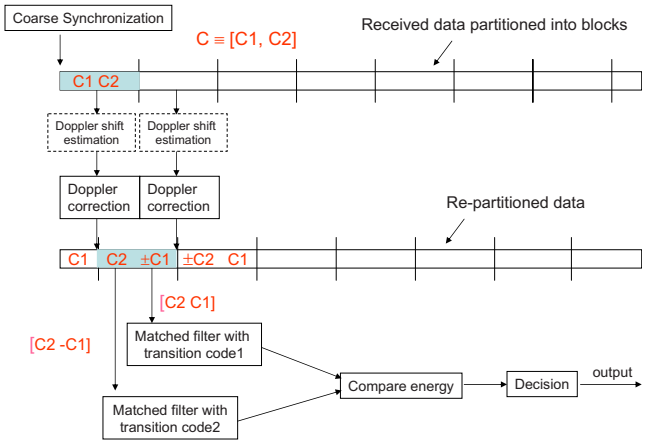


FIG. 7. (Color online) Schematics of the energy-detector method.

a symbol. This part is the same as before. Now the data are repartitioned into blocks which are shifted from the original partition by half a block. Now if the two adjacent symbols are of the same kind, the new partitioned block represents the symbol data with a code sequence [C2, C1], where C1 and C2 are the first half and second half of the original code sequence. Thus matched filtering this block of data with the code sequence [C2, C1] will yield the regular matched-filter gain. On the other hand, matched filtering this block of data with $[C2,-C1]$ will not. Likewise, if the two adjacent symbols are of the opposite kind, the block data represent the symbol data with a code sequence $[C2,-C1]$. Match filtering this block of data with the code sequence $[C2,-C1]$ will yield the regular matched-filter gain and match filtering this block of data with [C2, C1] will not. Consequently, by comparing the energy of the two matched-filter outputs, one can determine whether the adjacent symbols are of the same kind or opposite kind.

The data, which contain the sum of the desired signal and the interference signal (with different SIRs), are processed using the energy-detector method first without interference rejection. The results are shown in Figs. 8(a) and 8(b) for cases A and B. One sees that BERs are approximately similar. BER goes above 10^{-2} for SIR <-6 dB. With interference rejection, the BER is significantly improved. BER remains below 10^{-2} for SIR as low as -11 dB.

Figure 9 compares the results of the cross-correlation method and the energy-detector method, now averaged over many (30) packets with different time offsets. Comparing Fig. 9(b) with Fig. 9(a), one concludes that the energy-detector method performs much better than the cross-correlation method (by greater than an order of magnitude in BER reduction, or >3 dB in SIR for SIR >-10 dB) in the presence of strong interfering signals from other CDMA users.

The above results are for high level signals (SNR >5 dB). Figure 10 shows the similar results for low SNR signals (SNR $=-5$ dB). In this case, the energy-detector method performs significantly better than the cross-correlation method for SIR >-8 dB without interference rejection and slightly better than the cross-correlation method with interference rejection except at 0 dB. One achieves $<10^{-2}$ BER for SIR <-5 dB for the energy-detector method and for SIR <-3 dB for the cross-correlation method.

VI. SUMMARY AND CONCLUSIONS

This paper addresses the near-far problem for multi-access DS-CDMA in multipath underwater acoustic communications. Multipaths effectively destroy the code orthogonality rendering the conventional matched-filtering method ineffective. Because of the slow sound speed, closed-loop power control is not feasible in an underwater acoustic channel. While power can be adjusted on a gross scale, one must rely on signal processing to remove the interference in order to achieve a minimal uncoded BER. Toward this goal, the HCC0 method is used to suppress the interference by identifying the strong interference signal and transforming the data into the interferer's space using the interferer's code sequence. The interference signal is removed by hard clipping or coordinate zeroing and inverse transformed back to

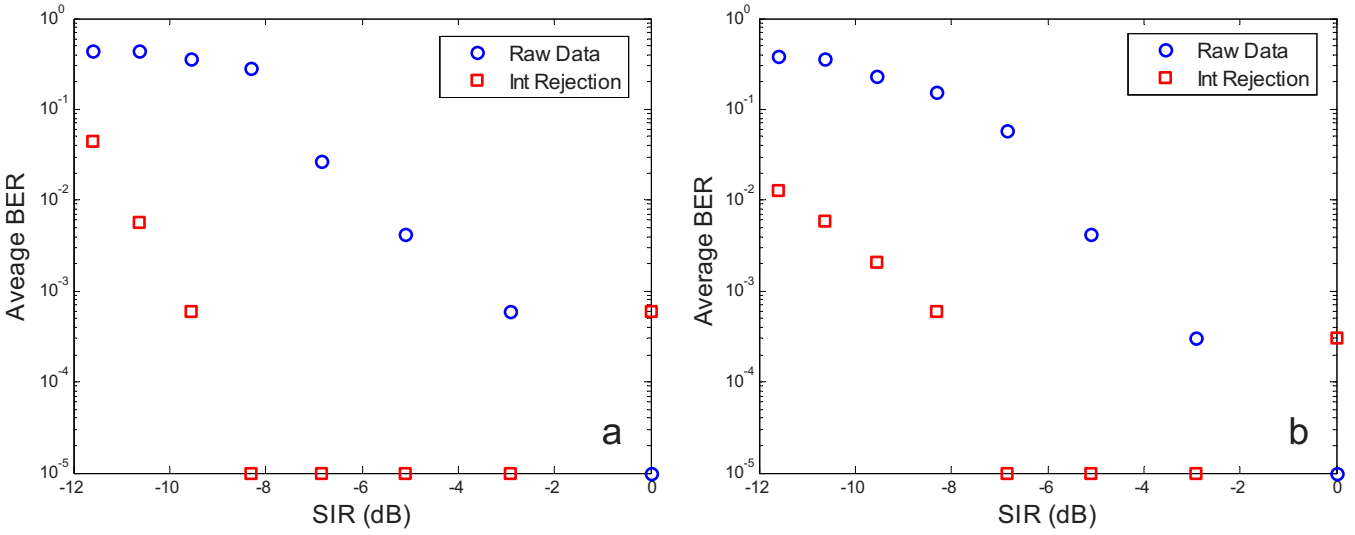


FIG. 8. (Color online) Average BER using the energy-detector method with and without interference rejection for case A (a) and case B (b). Zero BER is represented by 10^{-5} on the logarithmic scale.

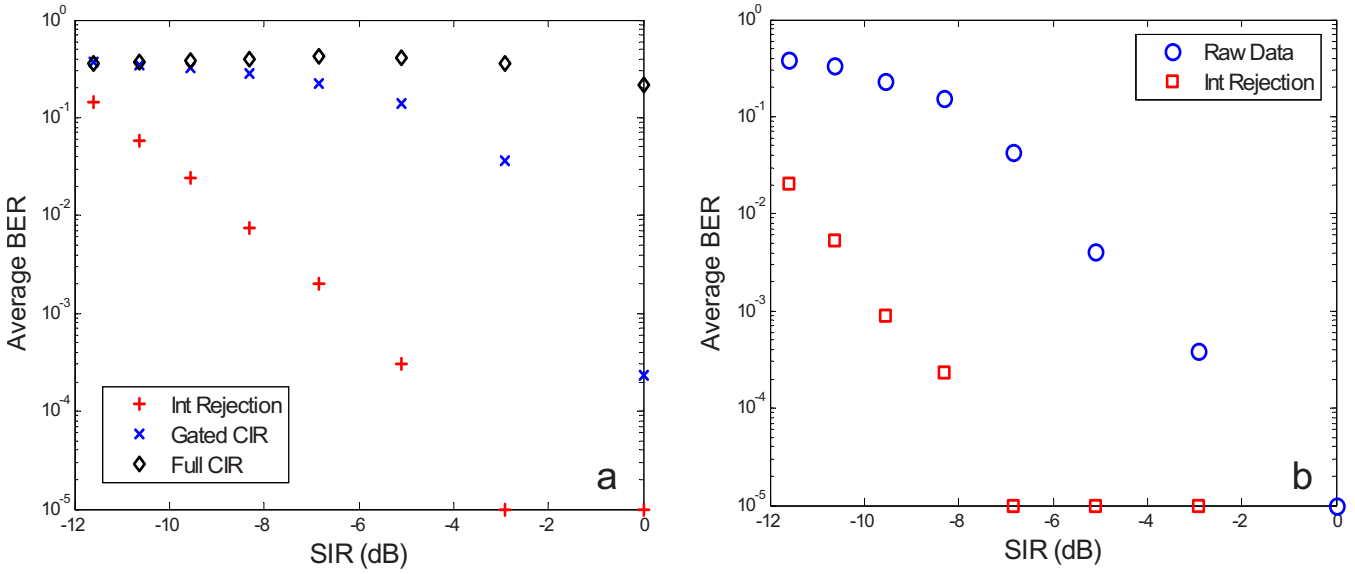


FIG. 9. (Color online) Comparison of the performance of the cross-correlation (a) and energy-detector methods (b). In this case, the BER is averaged over 30 packets with different time offsets between the desired and interference signals. Zero BER is represented by 10^{-5} on the logarithmic scale.

the original data space. This method is simple and efficient. It does not require joint symbol detection and signal parameter estimation (e.g., CIR and signal level) for each user as required in many published methods.

With the interference largely removed, the data are processed using two methods: the cross-correlation method and the energy-detector method, previously shown to work well with underwater acoustic data. One finds that for the case of two users, the energy-detector method works better than the cross-correlation method for SNR >-5 dB. For SNR >5 dB, the uncoded BER, after interference rejection, remains $<10^{-2}$ for SIR as low as -11 dB for the energy-detector method and for SIR as low as -9 dB for the cross-correlation method. For SNR $=-5$ dB, one finds BER $>10^{-2}$ for SIR <-2 dB for both methods without interference rejection. With interference rejection, the BER is significantly improved, favoring slightly the energy-detector method except for SIR $=0$ dB. One finds BER $<10^{-2}$ for SIR <-3 to -5 dB.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The UNet06 experiment was conducted under the auspices of The Technical Cooperative Program (TTCP) and the Office of Naval Research. The authors thank their colleagues at NRL and DRDC (Defense Research Development Canada) for their contributions to the UNet06 experiment.

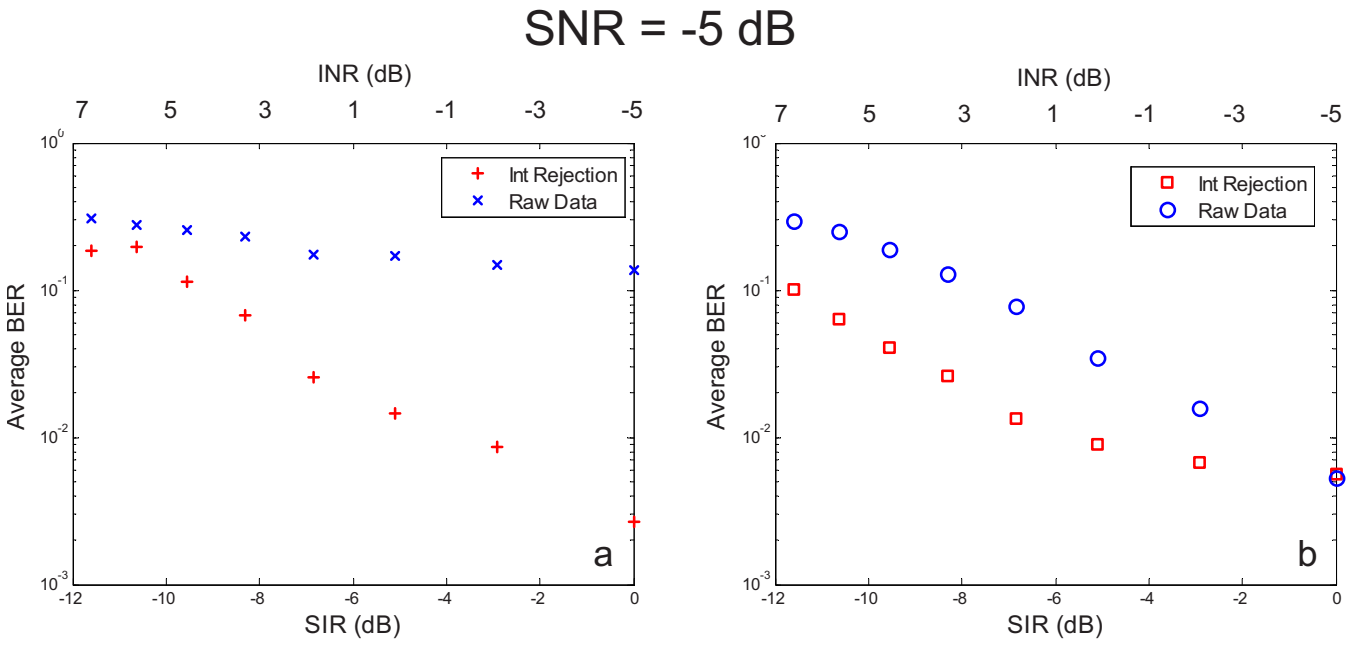


FIG. 10. (Color online) Comparison of the performance of the cross-correlation (a) and energy-detector methods (b) for SNR $=-5$ dB. INR stands for interference-to-noise ratio.

APPENDIX: CYCLIC CORRELATION USING FAST WALSH–HADAMARD TRANSFORM

The *m*-sequences possess a nice property which enables the cyclic correlation, Eq. (5), to be performed using an efficient algorithm, the FWHT, similar to the fast Fourier transform (FFT). This section will outline the computation algorithm of the FWHT for *m*-sequences. It hinges on the relation between the *m*-sequence matrix and the Hadamard matrix.[9,10] A good reference is the Appendix of Ref. 13.

$$M_L^k = V_R P_w H_{L+1} P_s V_I, \tag{A1}$$

where V_R is an $(L \times L+1)$ matrix to reduce the size of the processed vector from $L+1$ to L,

$$V_R = \begin{bmatrix} 0 & I_L \end{bmatrix}, \tag{A2}$$

where I_L is an identity matrix of dimension L. Similarly, V_I is an $(L+1 \times L)$ matrix to increase the size of the processed vector from L to $L+1$ given by

$$V_I = \begin{bmatrix} 0 \\ I_L \end{bmatrix}. \tag{A3}$$

P_w and P_s are permutation matrices which map an input vector to an output vector; their computation is negligible. The inverse of each of the permutation matrices is its transpose. For *m*-sequences, the mapping of the bit sequences is determined by the generation (bit sequence) of the *m*-sequence, referred to as the LAW of *m*-sequence. H_{L+1} is a Walsh–Hadamard matrix of size $(L+1 \times L+1)$. The Walsh–Hadamard matrix may be specified recursively by

$$H_1 = [1]$$

$$H_i = \begin{bmatrix} H_{i-1} & H_{i-1} \\ H_{i-1} & -H_{i-1} \end{bmatrix}, \tag{A4}$$

and these matrices exist only for orders of power of 2. Matrix multiplication by the Walsh–Hadamard matrix proceeds via an algorithm similar to the FFT algorithm.

The matrix multiplication in Eq. (5) can now be done by adding a zero to the input data, mapping the input data to a new vector, followed by multiplication by the Walsh–Hadamard matrix. The result is mapped back to the original data space and the first data sample is removed.

The inverse FWHT, as shown in Eq. (8), can be performed in a similar manner using the following equation:

$$(M_L^k)^{-1} = V_I^T P_s^T (H_{L+1})^{-1} P_w^T V_R^T, \tag{A5}$$

where the mapping matrices are the same as before. Multiplication by the inverse of the Walsh–Hadamard matrix can be done in an efficient manner like the inverse FFT. This allows Eq. (8) to be performed efficiently without calculating the inverse of a large matrix, e.g., of dimension (511×511) for an *m*-sequence of 511 elements.

[1] M. Stojanovic and L. Freitag, "Multi-channel detection for wideband underwater acoustic CDMA communications," IEEE J. Ocean. Eng. **31**, 685–695 (2006).

[2] F. Blackmon, E. M. Sozer, M. Stojanovic, and J. Proakis, "Performance comparison of RAKE and hypothesis feedback direct sequence spread spectrum techniques for underwater communication applications," in Proceedings of the MTS/IEEE OCEANS '02 (2002), Vol. **1**, pp. 594–603.

[3] E. Calvo and M. Stojanovic, "Efficient channel-estimation-based multiuser detection for underwater CDMA systems," IEEE J. Ocean. Eng. **33**, 502–512 (2008).

[4] C. C. Tsimenidis, O. R. Hinton, A. E. Adams, and B. S. Sharif, "Underwater acoustic receiver employing direct-sequence spread spectrum and spatial diversity combining for shallow-water multi-access networking," IEEE J. Ocean. Eng. **26**, 594–603 (2001).

[5] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," IEEE J. Sel. Areas Commun. **12**, 796–807 (1994).

[6] Y. F. Huang, J. H. Wen, and H. T. Wu, "An adaptive decision-feedback multiuser detector using parallel interference cancellation for CDMS systems," in Proceedings of the IEEE Vehicular Technology Conference (2001), Vol. **3**, pp. 1770–1774.

[7] T. Risteniemi and J. Joustansalo, "Advanced ICA-based receivers for DS-CDMS systems," in Proceedings of the IEEE PIMRC, London (2000).

[8] M. Stojanovic and Z. Zvonar, "Multichannel processing of broad-band multiuser communications signals in shallow water acoustic channel," IEEE J. Ocean. Eng. **21**, 156–166 (1996).

[9] H. K. Yeo, B. S. Sharif, A. E. Adams, and O. R. Hinton, "Performances of multi-element multi-user detection strategies in a shallow-water acoustic network (SWAN)," IEEE J. Ocean. Eng. **26**, 604–611 (2001).

[10] H. K. Yeo, B. S. Sharif, A. E. Adams, and O. R. Hinton, "Implementation of multiuser detection strategies for coherent underwater acoustic communication," IEEE J. Ocean. Eng. **27**, 17–27 (2002).

[11] T. C. Yang and W.-B. Yang, "Performance analysis of direct-sequence spread-spectrum underwater acoustic communications with low signal-to-noise-ratio input signals," J. Acoust. Soc. Am. **123**, 842–855 (2008).

[12] T. C. Yang and W.-B. Yang, "Low probability of detection underwater acoustic communications using direct-sequence spread spectrum," J. Acoust. Soc. Am. **124**, 3632–3647 (2008).

[13] H. S. Chang, "Detection of weak, broadband signals under Doppler-scaled, multipath propagation," Ph.D. thesis, University of Michigan, Ann Harbor, MI (1992).

[14] H. DeFerrari and A. Rodgers, "Eliminating clutter by coordinate zeroing," J. Acoust. Soc. Am. **117**, 2494 (2005).

[15] L. R. Welch, "Lower bounds on maximal cross correlation of signals," IEEE Trans. Inf. Theory **20**, 397–399 (1974).

[16] T. G. Birdsall and K. Metzger, "Factor inverse and matched filtering," J. Acoust. Soc. Am. **79**, 91–99 (1986).

[17] M. Cohn and A. Lempel, "On fast M-sequence transforms," IEEE Trans. Inf. Theory **23**, 135–137 (1977).

[18] The data the authors have use long (511) code sequences as the original purpose was for low signal-to-noise ratio (SNR) underwater acoustic communications. The methods discussed in this paper can be applied to DS-CDMA signals for short code sequences as well. However, as noted above, the available orthogonal codes with short *m*-sequences are very limited.

Modeling perceptual effects of reverberation on stereophonic sound reproduction in rooms

Thomas Zarouchas[a] and John Mourjopoulos
Department of Electrical Engineering and Computer Engineering, Audio and Acoustic Technology Group, Wire Communications Laboratory, University of Patras, Achaia 26500, Greece

(Received 20 March 2008; revised 6 April 2009; accepted 14 April 2009)

The proposed model derives time-frequency maps to estimate perceived alterations due to reverberation in stereo audio signals reproduced in rooms. These alterations relate to monaural masking due to reverberant decay, derived via a computational auditory masking model and to inter-channel cues for the formation of the spatial position of the aural objects, derived via an inter-channel cue mapping module. The maps illustrate in detail the varying nature of the perceptually-relevant alterations due to room reverberation. Quantitative metrics are also introduced which were found to be proportional to reverberation interference, to room-reverberation time and to depend on the specific audio signal. A statistical approach classifies room response properties via their histogram distributions. Corresponding distributions were also applied to the proposed signal-dependent perceptual maps. Such distributions were found to be useful for interpreting the perceived alterations with different kinds of signals, such as music or speech.
 [DOI: 10.1121/1.3129382]

PACS number(s): 43.60.Dh, 43.60.Uv [LMW] Pages: 229–242

I. INTRODUCTION

In recent years, so-called perceptual models have emerged within the audio and speech fields to complement the established signal processing methods. Such models employ descriptions of human hearing mechanisms and become increasingly important in applications such as audio and speech coding, multi-channel audio recording/coding/reproduction, audio and speech enhancement, speech recognition in adverse environments, etc. Room reverberation, being a complex phenomenon to be accurately described by signal processing representations and also being associated with sophisticated perceptual mechanisms, is a natural candidate for the evolution of such models. The present work is concerned with such models for stereophonic acoustic signals reproduced in rooms with reverberation.

As is well-known, room acoustics introduce reverberation to audio signals, which is formally described by the linear system response functions of the room. This approach has provided a system-dependent framework for modeling acoustic signal reproduction in such spaces, describing features that are important from a signal processing perspective.[1–3] Exploiting knowledge derived from psychoacoustics,[4] room reverberation [and room impulse responses (RIR)] are often analyzed in three distinct segments: (a) the acoustic signal that travels directly from source to receiver (so-called direct sound), (b) early reflections that arrive up to 80 ms after the direct signal, important for distance and spaciousness perception,[4–6] and (c) an exponentially-decaying stochastic reverberation tail with echo density increasing with the square of time.[7,8] It is also customary for engineers to extract well-established energy-related acoustical parameters [e.g., reverberation time, (RT), clarity, and definition] from such RIR analysis. Nevertheless, these parameters and functions can only provide a limited description of the perceptual effects of the rooms with respect to all kinds of signals. Furthermore, they appear to be limited and not sufficiently robust when more sophisticated signal processing methods have to be employed in such applications as sound field control or room acoustics correction/dereverberation.[9–14]

It is well-known that the perception of reverberation is a very complex phenomenon, resulting from time-frequency (T-F), delay, directional, and signal-dependent cues such as the signal's spectrum, signal onset, and offset and level.[4,15–18] Hence, an alternative approach to describe room reverberation is based on the perceptual approach considering that the human listener as a receiver is known not to perceive all the detailed information in the RIR. A model for such perceptual mechanism was introduced via the concept of "reflection masking,"[16] since many reflections are masked by the direct signal or by other reflections and are therefore inaudible.

Perception of acoustics signals in reverberation-charged spaces, especially when reproduced via stereo or multichannel audio systems, is necessarily linked to spatial hearing characteristics related to binaural fusion, position of the aural object, and the precedence effect.[4,17–19] It is also well-known that spatial hearing is highly determined by inter-aural differences of the signals at the listeners' ears (inter-aural cues)[4] and in any room, reflections from several different directions affect these inter-aural cues. Nevertheless for many recent engineering applications such as in multi-channel audio coding,[20,21] in binaural cue coding,[22,23] in directional audio coding,[24] and in parametric stereo coding,[25] it has been shown that for a typical acoustic signal reproduction setup in enclosed space, the inter-channel cues may also efficiently represent most attributes of the auditory event perceived by a

[a]Author to whom correspondence should be addressed. Electronic mail: thozar@wcl.ee.upatras.gr

human listener. As was described by Faller,[26] the use of different inter-channel cues for representation of auditory event properties can be interpreted as follows: "…summing localization[4] implies that perceptually-relevant audio channel differences for a loudspeaker signal channel pair are inter-channel time differences (ICTD) and inter-channel level differences (ICLD). ICTD and ICLD can be related to the perceived direction of auditory events…" Other auditory event attributes, such as apparent source width[5] and listener envelopment,[6,27–30] can be related to the inter-aural cross-correlation coefficient[5,26,29] (IACC). For loudspeaker pairs IACC is often directly related to inter-channel coherence[26,30] (ICC). Hence, the inter-channel cues[20,22,25] (ICTD, ICLD, and ICC) considered here are similar measures to the well-established inter-aural cues [ITD, ILD, and IC (Refs. 4 and 31)], but for such audio reproduction applications as the ones described here, are evaluated across the audio channels as opposed to ear entrance channels.

In order to represent such perceptual mechanisms, it is also becoming increasingly clear that it is convenient to derive two-dimension (2D) maps corresponding to each perceptual attribute. In most cases these maps illustrate perceptually-derived cues in the T-F domain, not unlike the spectrogram representations. To extend this discussion into methods of assessing such spatial attributes, it is useful here to recall earlier work in binaural activity maps generated by feeding a binaural computational model with binaural impulse responses which have been recorded in concert halls with the aid of an artificial head, the so-called "binaural room responses."[4,6] Such maps were further combined by means of a fuzzy temporal-cluster algorithm.[32] Subsequent work mainly from the speech community has integrated some additional binaural processing mechanisms into such maps. For example, binaural mechanisms in humans appear to adapt and to counteract (up to a certain degree) to the effects of reverberation, by suppressing echoes. A computational auditory scene analysis method was introduced by Harding *et al.*[33] and Brown *et al.*[34] which exploits binaural processing in order to improve the robustness of automatic speech recognition systems in multi-source, reverberation-charged environments.

Such perceptually-motivated methods for analyzing and processing reverberation-afflicted signals attempt to bridge the gap between the signal processing (system-dependent) framework and the perceptual (signal-dependent) methodology rooted in psychoacoustics. The present work complements such methods by extending an earlier method[16] and by introducing a compact and efficient perceptually-motivated model of stereo-signal reproduction in rooms. Specifically, a computational auditory masking model[35,36] (CAMM) is employed for modeling monaural reverberation masking effects, complemented by an inter-channel cue mapping module (ICMM) used for the description of the perceptual degradations of the stereo channel cues in a typical audio reproduction setup. The proposed method is not aiming at accurately modeling the physiological mechanisms involved in the perception of reverberation, but instead it is expanding the tools employed in engineering applications related to robust acoustic signal reproduction in rooms.[17,37]

The proposed method requires as inputs the source ("anechoic") audio signal(s) and the corresponding reverberation-loaded signal(s), recorded in a room or simulated via convolution with measured room impulse responses (RIRs). According to this approach, it is possible to locate, from the evaluated "internal representations" derived via the auditory model, T-F regions with significant alteration due to reverberation. Furthermore, the output of the ICMM displays the alterations in the relevant spatial cues due to reverberation. In both cases, the outputs of the CAMM and ICMM are presented in a form of T-F (2D) maps.[33,38,39] From such maps the present work introduces a number of monaural and inter-channel metrics, which are signal-dependent and time-varying in a way resembling the well-established noise to mask ratio[40] and other algorithmic quality measures such as the perceptual audio quality measure[41] employed for audio and speech signal-dependent evaluation. The proposed metrics are novel since up to now only few models have been introduced to provide useful signal-distortion measures[42–44] for engineering applications related to room acoustics assessment, and optimal audio signal reproduction in reverberation-charged spaces. Furthermore, preliminary results by the authors for appropriate modification of such maps to reduce reverberation-induced cues show promising results and improved robustness for acoustic signal reproduction in rooms.[38,45] To assess the performance of the proposed metrics with respect to the properties of typical systems (e.g., for different rooms) and with respect to different audio signals, a statistical analysis is introduced here acknowledging that earlier work in room acoustics has also evaluated room transfer function (RTF) statistics.[46–48] Recent speech and audio dereverberation applications are increasingly relying on such statistical methods in order to extract the differences between the audio-speech signal and the interfering noise-reverberation distortion.[49–54] For example, a dereverberation method proposed by Yegnanarayana *et al.*[51] and Wu *et al.*[52] employs appropriate statistical signal processing to linear-prediction residual values having lower kurtosis than the audio-speech signal itself. Another recent work by Extra *et al.*[55] applies block-by-block RIR kurtosis analysis for evaluating artificial reverberation algorithms.

The present study is organized as follows: in Sec. II the proposed structure for the generation of 2D T-F maps and related metrics due to reverberation is introduced (for both single-channel and inter-channel cues). Section III describes the statistical analysis that is performed for both system and signal-dependent parameters. In Sec. IV the methodology and the results of the various tests are presented to assess the efficiency of the proposed method. Finally, Sec. V presents discussion of the results and conclusions are drawn in Sec. VI.

II. MAPS OF PERCEPTION-BASED ALTERATIONS DUE TO REVERBERATION

A. Overview

As was discussed, the proposed approach is based on a combination of well-established system-dependent (algorithmic) methods together with known perception-based ap-

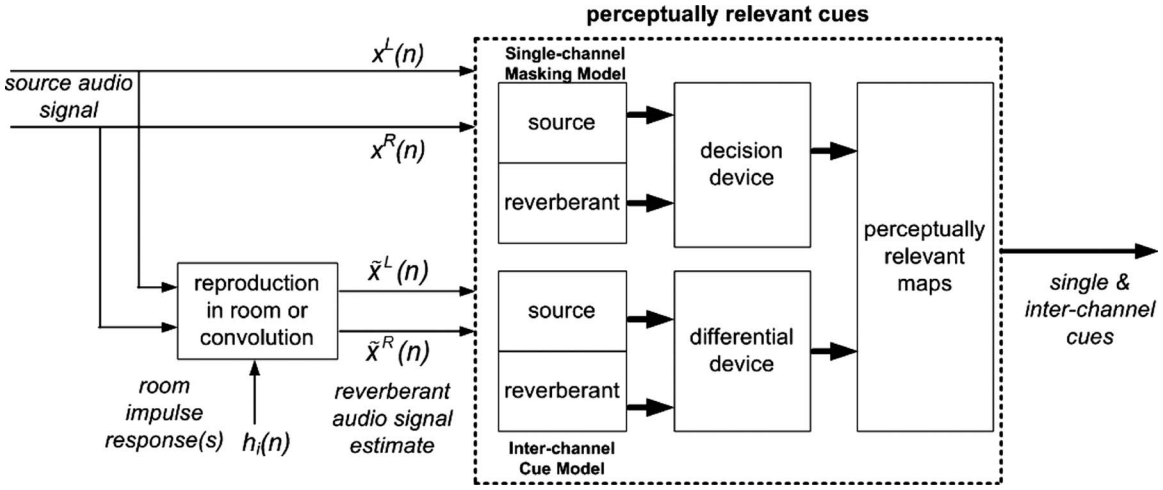


FIG. 1. Analysis scheme for the assessment of the perceptually-relevant reverberation cues.

proaches, leading to signal-dependent 2D maps of room-reverberation attributes. The structure for the derivation of such maps is shown in Fig. 1, for stereo reproduction. The concept can be easily extended and generalized to handle any number of audio channels, e.g., multi-channel audio signal reproduction for 5.1 channel applications.

The RIR function(s) $h_i(n)$ required in this description are assumed here to be always available from appropriate measurement, being also dependent on the specific source/receiver locations.[14] Future research will examine the feasibility of substituting such measurements with suitable RIR predictions or by simplified RIR models.

B. Filterbank

As is the case with most auditory models, a gammatone filterbank is employed in the signal pre-processing stage (see Fig. 3). Existing gammatone filterbank realizations can achieve analysis-to-synthesis reconstruction performance which is not sufficient for transparent 16-bit full bandwidth audio processing.[56] Given that transparent filterbank performance was required here,[38] a novel filterbank was utilized, with nearly-perfect reconstruction properties. This filterbank provides non-uniform analysis bands with sufficient frequency resolution in order to capture the perceptually-relevant cues at low frequencies, following closely the critical band equivalent rectangular bandwidth scale.

The frequency response of the first 15 frequency bands ($k=0,1,\ldots,14$) of the filterbank is illustrated in Fig. 2. The effective bandwidth of the frequency bands shown in Fig. 2 is approximately 43 Hz (for $k=0,\ldots,7$), 86 Hz (for $k=8,\ldots,11$), and 172 Hz (for $k=12,13,14$), for a sampling frequency $f_s=44\,100$ Hz.

C. The CAMM

The reverberation masking model used here was proposed by Buchholz and Mourjopoulos[35,36] and can emulate many aspects of the monaural signal processing of the auditory system, as inferred from psychoacoustic experimentation. This computational auditory masking model (CAMM) is based on the concept of signal dependent compression (SDC) (see Fig. 3), which assumes that the auditory system performs a compression of the input signal's amplitude and these characteristics mainly depend on the input audio signal's evolution.

Input to the auditory model are a single channel of the original anechoic audio signal $x(n)$ and the corresponding reverberation-afflicted audio signal $\tilde{x}(n)$ (tilde "~" as a superscript denotes features of reverberation-loaded signals) generated within a room, as can be measured at a specific position via an omnidirectional microphone. Alternatively, a simulated signal with reverberation may be obtained via convolution with a measured RIR. According to the CAMM, the input signals are first analyzed by a pre-processing module, consisting of the filterbank emulating the auditory frequency analysis performed on the basilar membrane, a full-wave rectifier, and a low-pass filter simulating the mechanical to neural transduction performed by the inner hair cells. The resulting signals ($s_k(n)$ and $\tilde{s}_k(n)$, respectively, k being the frequency band index, are processed by a SDC module which employs a static non-linear function.[35,36] The CAMM derives what we call internal representations $z_k(n)$ and $\tilde{z}_k(n)$

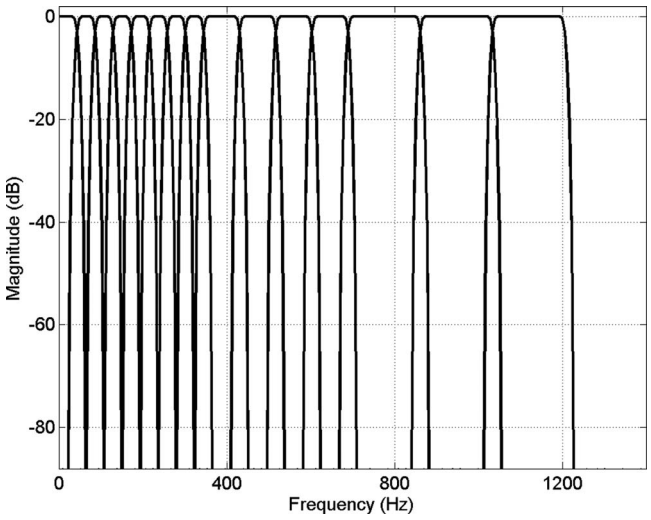


FIG. 2. Magnitude responses for the first 15 frequency bands of the filterbank.

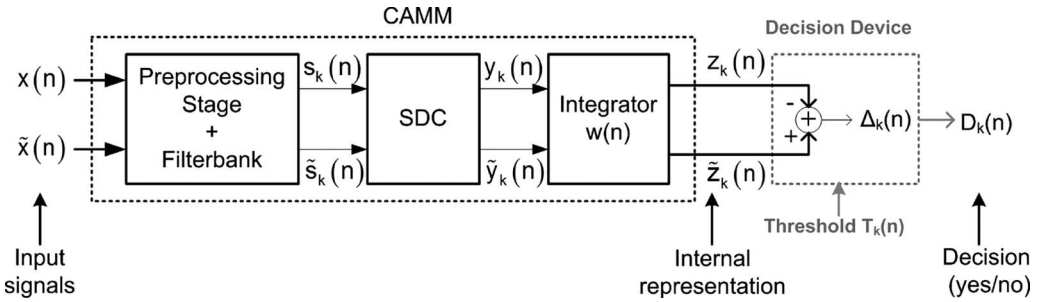


FIG. 3. Block diagram of the CAMM based on SDC module, for modeling room reverberation.

of the audio signals in a number of frequency bands which are fed onto a decision threshold device (DTD). The DTD module describes a higher-level process, which in this case is restricted to the specific task of generating appropriate decisions with respect to masking or signal/distortion detection. The concept of the DTD is based on the just noticeable intensity difference[35,36] of the internal signal representations. Therefore, the per-sample difference of the corresponding internal representations is calculated providing exact T-F regions with significant alterations due to reverberation. Formally, the output at the SDC module for the source signal $y_k(n)$ is described by

$$y_k(n) = 10\log_{10}\left(1 + \frac{s_k(n)}{s_{\text{op}}}\right) \quad \text{and} \quad s_{\text{op}} = 10^{\log_{10}(g_k \cdot s_k + 1) * h_{\text{norm}}}, \tag{1}$$

with g_k (see Appendix) being a constant gain factor defining the effective level of the operating point signal s_{op} and h_{norm} being the response of a linear filter (energy normalized running integrator) with the energy equal to 1.[35,36] A similar expression applies also for the reverberation-carrying signal $\tilde{y}_k(n)$. At the DTD, the following difference is derived:

$$\Delta_k^z(n) = \tilde{z}_k(n) - z_k(n) = \sum_{m=0}^{N-1} 10\log_{10}\left(\frac{1 + \frac{\tilde{s}_k(n)}{\tilde{s}_{\text{op}}}}{1 + \frac{s_k(n)}{s_{\text{op}}}}\right) \cdot w(n-m), \tag{2}$$

where $w(n)$ is first order lowpass filter with cut-off frequency f_g=4 Hz.[35,36,57]

D. Modeling monaural cues

The DTD, accompanied by a set of thresholds[58] $T_k(n)$, is utilized to extract the difference $\Delta_k^z(n)$, according to Eq. (2) and therefore to derive the reverberation masking index (RMI) function:

$$D_k^m(n) = \Delta_k^z(n) - T_k(n). \tag{3}$$

The RMI $D_k^m(n)$ represents an estimate of the perceived alterations due to reverberation above the specified threshold, i.e., when $0 \le D_k^m(n) \le d$, in the T-F domain for any filterbank channel signal, where d indicates the maximum alteration.

Typical results for the variation of the RMI $D_{k,n}^m$ (dB) and the corresponding 2D map for a single channel of a reverberation-afflicted acoustic signal (piano) recorded in room R3, a sports-hall, are shown in Fig. 4 (for details of this room, see Sec. IV). As can be observed in Fig. 4(c), the frequency-averaged metric $D_{k,n}^m$ increases during the reverberation decay of the piano note [shown without reverberation in Fig. 4(a)], illustrating the increase in the perceived

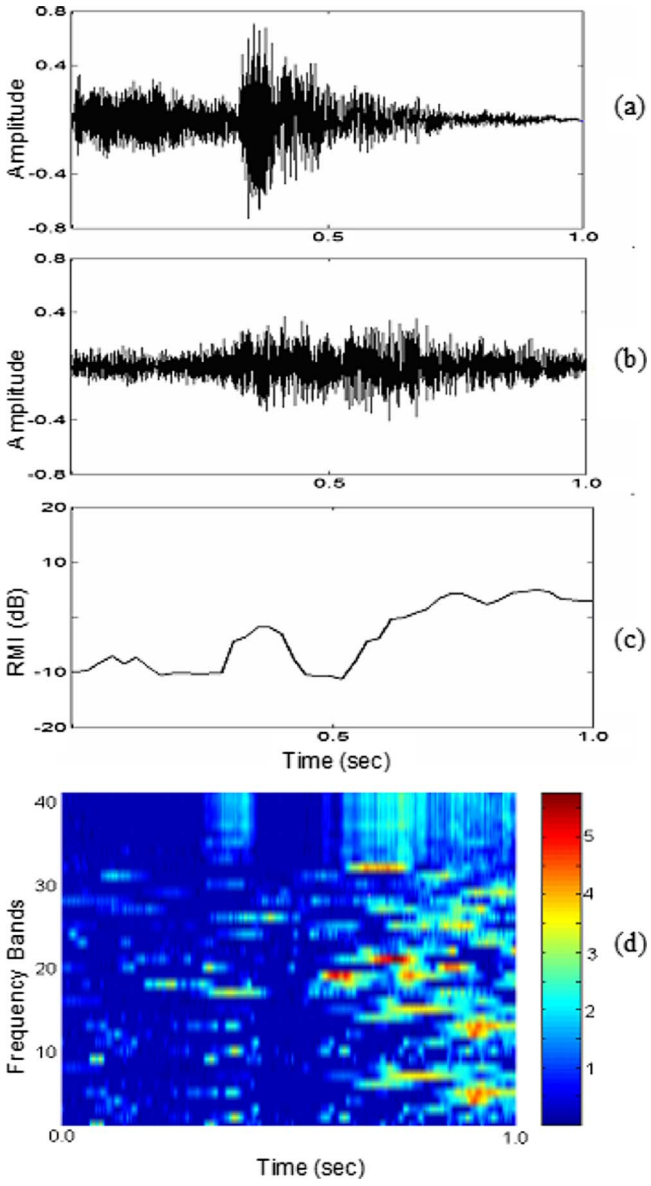


FIG. 4. (Color online) (a) Source audio signal segment piano, (b) corresponding reverberation-afflicted signal recorded in room R3, (c) frequency-averaged parameter $D_{k,n}^m$ (dB), and (d) 2D map derived from the CAMM.

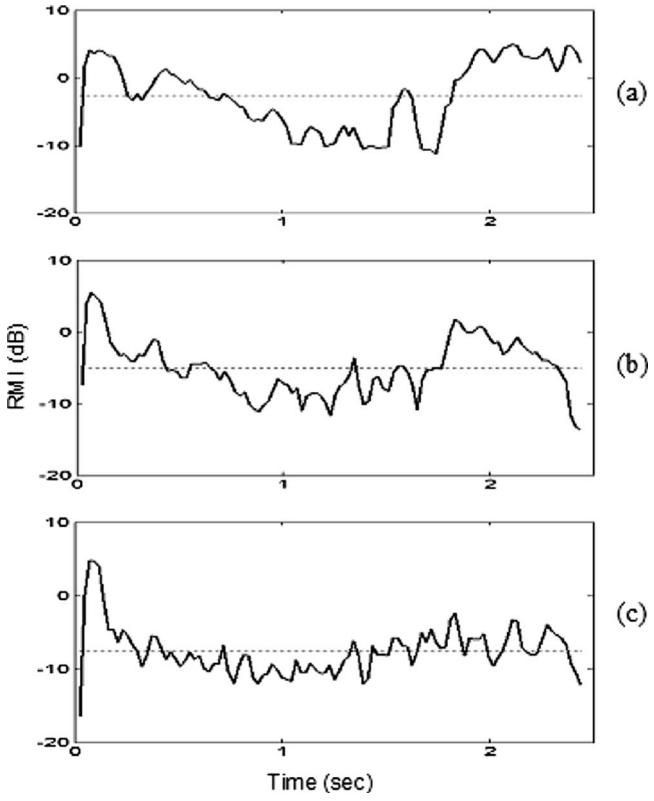


FIG. 5. Frequency-averaged RMI for an audio segment reproduced in different rooms: (a) room R3 (sports-hall), (b) room R2 (classroom), and (c) room R1 (laboratory). Dashed line indicates mean value over the complete segment.

alterations due to reverberation tail. The corresponding perceptually-motivated T-F map based on the CAMM for the reverberation-loaded audio signal segment is shown in Fig. 4(d). This 2D map gives a detailed illustration of the corresponding T-F alterations, indicating signal regions with a higher degree of perceived alterations.

Significantly, the proposed RMI appears to be sensitive to differences in room acoustics and this effect is shown in Fig. 5, where results for the same audio segment are shown after reproduction in three different rooms, the first having a RT of T_{60}=6.4 s, the second a RT of T_{60}=1.1 s, and the third a RT of T_{60}=0.368 s.

Note that the dashed line in each case indicates the mean value of the estimated perceived alterations over the duration of the audio segment. It is clear that the RMI varies along the time signal's evolution and that the mean of the RMI increases with RT so that the perceived effects of reverberation are more pronounced for the larger rooms [Figs. 5(a) and 5(b)] than for the acoustically-treated room [Fig. 5(c)]. Furthermore, heavier reverberation seems to bias the results toward long duration RMI peaks.

The CAMM may be extended to evaluate separately the RMI $D_k^m(n)$ for the two channels of a stereo signal. However, in its current state and due to the model being limited to monaural aspects of hearing/masking, any direction-dependent cues are not utilized and neither can be accommodated by this specific decision device. Such cues are only considered by the complementary ICMM described in Sec. II E.

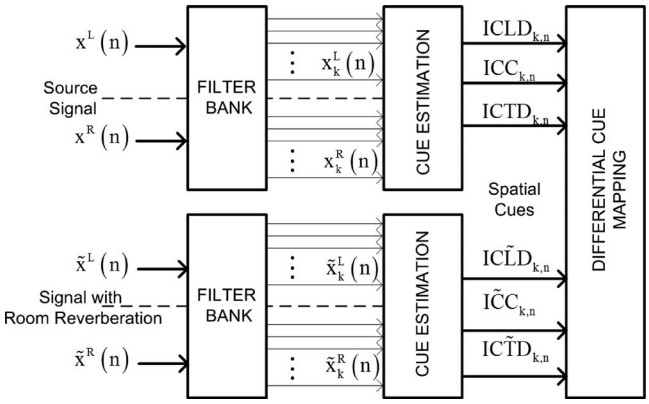


FIG. 6. Analysis scheme of the ICMM. Tilde "~" denotes reverberation-afflicted signals and cues.

E. Modeling inter-channel cues

Input to the ICMM, which is included in the scheme of Fig. 1, are the two stereo channels for both the source audio signal and the corresponding reverberation-afflicted signal as will be reproduced inside the room. The relevant spatial cues examined here are the ICLD, ICTD, and ICC. These are derived for all signals and channels independently, in each frequency band as a function of time as is shown in Fig. 6.

ICLD (in dB) denotes the level/intensity differences between two (left-right) channels:[4,20]

$$\mathrm{ICLD}(k,n) = 10\log_{10}\left(\frac{p_{x^R}(k,n)}{p_{x^L}(k,n)}\right), \tag{4}$$

where p_{x^R}, p_{x^L} are short-time estimates of the power of each channel. ICLD has a typical level range of

$$-\ell \le \mathrm{ICLD}_k(n) \le \ell. \tag{5}$$

ICTD (in samples) describes the time difference between two channels, and is the time instance at which the maximum value of a short-time estimate of the normalized cross-correlation function has occurred, i.e.,

$$\Phi_{x^R x^L}(k,n) = \frac{p_{x^R x^L}(k,n)}{\sqrt{p_{x^R}(k,n)p_{x^L}(k,n)}}, \tag{6}$$

where $p_{x^R x^L}$ is a cross-power estimate between the two channels. ICTD has a typical time range (samples) of

$$-N \le \mathrm{ICTD}_k(n) \le N. \tag{7}$$

ICC defines the coherence between two channels (x^R and x^L) and can be expressed as

$$\mathrm{ICC}(k,n) = \max_{n=N_0}\{\Phi_{x^R x^L}(k,N_0)\}, \tag{8}$$

considering the maximum value of the instantaneous normalized cross-correlation. ICC has a range of

$$0 \le \mathrm{ICC}(k,n) \le 1, \tag{9}$$

where 1 indicates that x^R and x^L are perfectly coherent. Based on the definitions of Eqs. (4)–(9), the spatial cues are computed within the following psychoacoustically motivated ranges:

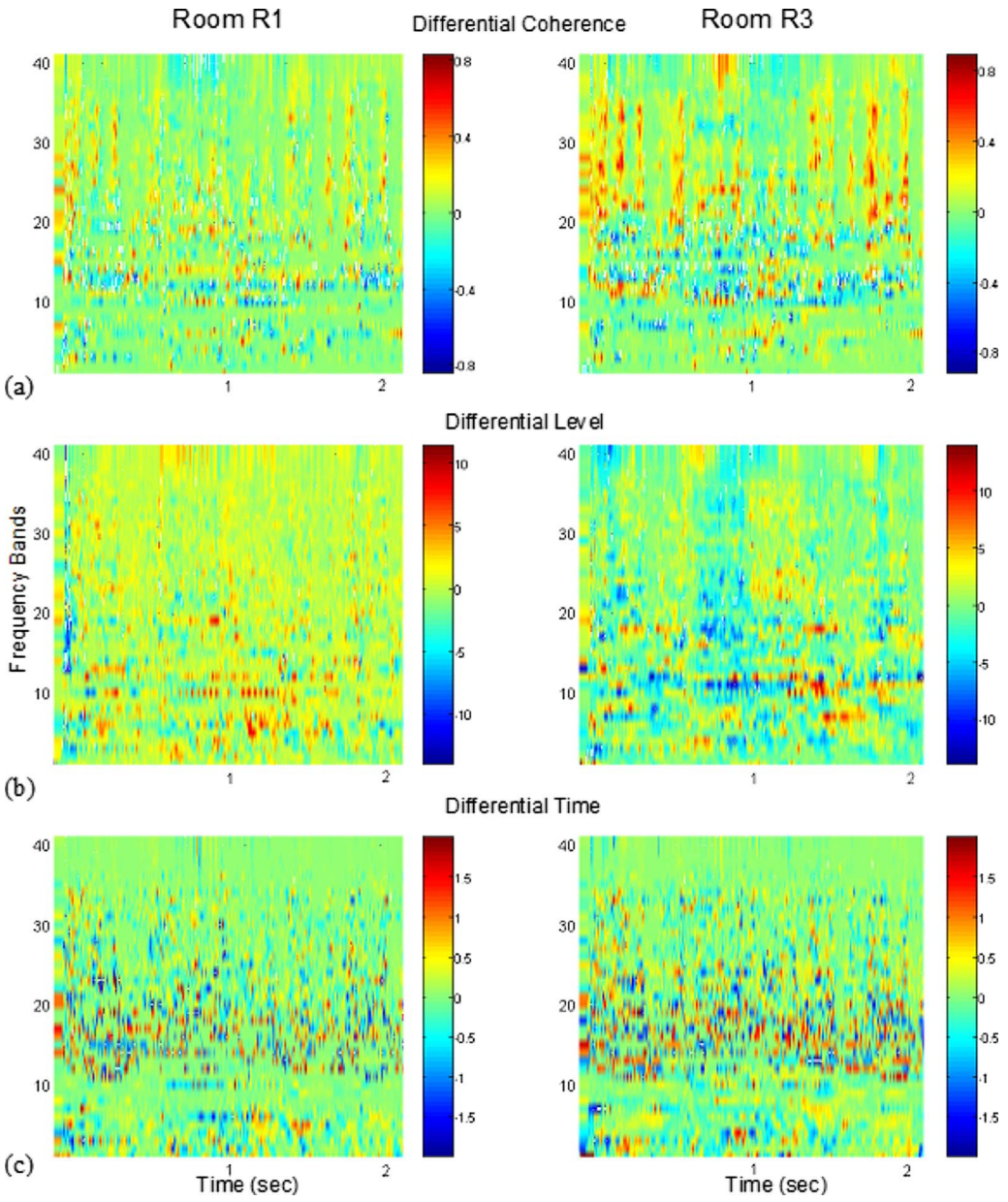


FIG. 7. (Color online) Differential inter-channel cue maps, (a) coherence, (b) level difference (in dB), and (c) time difference (in ms), for rooms R1 and R3 and using a jazz music segment as test signal. Note that darker areas indicate maximum alterations due to reverberation.

$$-\ell_1 \le \mathrm{ICLD}_{k,n} \le \ell_1 \text{ and } -\ell_2 \le \mathrm{I\tilde{C}LD}_{k,n} \le \ell_2,$$

$$-N_1 \le \mathrm{ICTD}_{k,n} \le N_1 \text{ and } -N_2 \le \mathrm{I\tilde{C}TD}_{k,n} \le N_2,$$

$$0 \le \mathrm{ICC}_{k,n} \le 1 \text{ and } 0 \le \mathrm{I\tilde{C}C}_{k,n} \le 1. \quad (10)$$

The $\mathrm{ICLD}_{k,n}$ and $\mathrm{I\tilde{C}LD}_{k,n}$ cues are estimated in the range of $[-7,7]$ dB. Accordingly, the $\mathrm{ICLD}_{k,n}$ and $\mathrm{I\tilde{C}TD}_{k,n}$ cues are estimated in the range of $[-1,1]$ ms, i.e., $(N_1/f_s$ and $\mathrm{N}_2/f_s) \in [-1,1]$ ms for f_s=44 100 Hz.

F. Differential inter-channel metrics

In order to calculate the differences between the spatial cues which correspond to original (source) and received (reverberation-afflicted) signals, differential 2D maps are generated (as T-F presentations). Therefore, the differential map for each cue is defined as

$$D^{\ell}_{k,n} = \mathrm{I\tilde{C}LD}_{k,n} - \mathrm{ICLD}_{k,n},$$

$$D^{t}_{k,n} = \mathrm{I\tilde{C}TD}_{k,n} - \mathrm{ICTD}_{k,n},$$

$$D^{c}_{k,n} = \mathrm{I\tilde{C}C}_{k,n} - \mathrm{ICC}_{k,n}, \quad (11)$$

and based on Eqs. (10) and (11), typical level, time, and coherence range for each differential metric will be

$$-(\ell_1+\ell_2) \le D^{\ell}_{k,n} \le (\ell_1+\ell_2),$$

$$-(N_1+N_2) \le D^{t}_{k,n} \le (N_1+N_2),$$

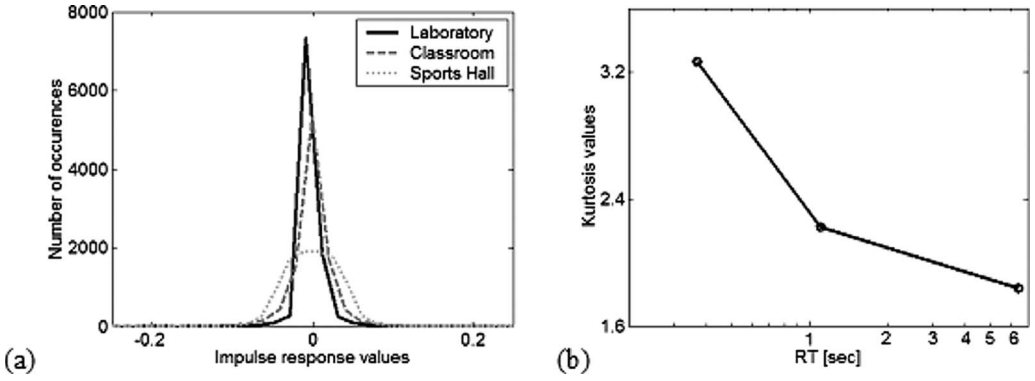


FIG. 8. (a) PDF of RIRs for different rooms. (b) Kurtosis values for corresponding RIRs.

$$-1 \leq D_{k,n}^{c} \leq 1. \tag{12}$$

Typical differential maps, corresponding to the acoustically-treated room (room R1) and the large sports-hall (room R3), can be seen in Fig. 7. It can be observed that for low reverberation and controlled reflections, as is the case of room R1, all differential inter-channel metrics display low dispersion (distortion maps have large T-F regions close to green, i.e., 0 dB for $D_{k,n}^{\ell}$ metric). Furthermore, deviations are low around this value. With heavy reverberation (e.g., room R3), increased variance can be observed for each differential inter-channel metric.

For the RMI ($D_{k,n}^{m}$) and the differential metrics ($D_{k,n}^{c}, D_{k,n}^{\ell}, D_{k,n}^{t}$), the mean values can also be evaluated for each test case. Such overall mean values, (formulated in a logarithmic scale except for the $D_{k,n}^{\ell}$ differential metric), provide an averaged interpretation for the global perceptually-relevant alterations which can be derived from each map, for the complete time interval of the specific acoustic signal reproduction.

III. STATISTICAL ANALYSIS OF SYSTEM- AND SIGNAL-DEPENDENT ALTERATIONS

A. System-dependent statistical analysis

Since more than half century ago room acoustics has been employing statistical analysis for the interpretation of reverberation phenomena.[46–48] Room response frequency and time domain statistics are also related to each other[8] being well-accepted that these can be treated as stochastic processes beyond certain frequency (the RTF Schroeder frequency) and/or time interval (practically corresponding to RIR tail). Nevertheless, for any RIR the autocorrelation function and any histogram of the amplitudes of all samples of the impulse response of a room will vary along time, indicating the relative dominance of the direct signal, the early reflections or the exponentially decaying stochastic reverberation tail.[55] Here, a simplified statistical analysis framework is introduced which evaluates the RIR amplitude histograms treating them as normal distributions with mean value μ and variance σ^2. As it is known, normal distribution has a probability function, i.e.,

$$P(x;\sigma,\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}. \tag{13}$$

Hence, assuming that the amplitude values of the RIRs $h_i(n)$ samples are random variables, an assumption that always applies for the response tail,[8] the corresponding PDFs will be

$$P(h_i(n)) = P(h_i(n);\sigma_i,\mu_i). \tag{14}$$

From this statistical system-dependent perspective, the acoustic properties of different rooms may approach normal distributions represented by different variance. Variations of the general trends in distributions may also register corresponding variations of the acoustic properties (e.g., a RIR distribution with longer tails may indicate higher RT value). Higher order statistics (e.g., kurtosis) may then be employed in order to describe the acoustic properties of any room or even to formulate a reasonable metric of reverberation present in each RIR.

As is known, kurtosis is a measure of whether a data set [e.g., $h_i(n)$] is "peaked" or "flat" relative to a normal distribution and is expressed as the fourth central moment divided by fourth power of the standard deviation, i.e.,

$$K = \frac{E\{(x-\mu)^4\}}{\sigma^4}, \tag{15}$$

where μ is the mean and σ^2 the standard deviation of x, being known that a data set with a normal distribution has a kurtosis of three.[59,60] The assumption of the relationship between room RT and kurtosis was here confirmed by analysis of RIRs corresponding to three spaces with significant variation in acoustic properties, as it is described in Sec. IV. For the test cases considered and from RIR distributions, a running measure of kurtosis was evaluated for successive non-overlapping blocks of 1024 samples of the room's response. Figure 8(a) shows that RIR statistics generally follow normal distributions, but with significantly varying spread of amplitude values from the mean value: the large sports-hall with long RT has a widely spread distribution of non-zero amplitude values, corresponding to the late reverberation energy. Figure 8(b) shows that the kurtosis value when plotted against room RT value, clearly displays these acoustic trends for the three different acoustic spaces. The acoustically-treated laboratory with near-perfect acoustics has a kurtosis

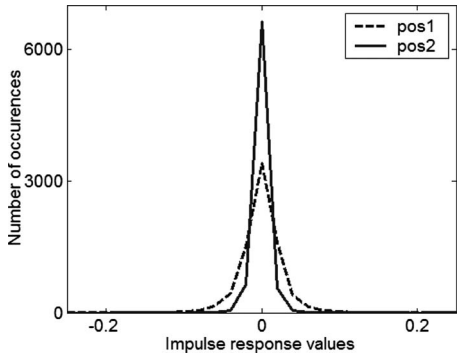


FIG. 9. PDF of RIRs for different source/receiver positions within room R2.

value above 3 (the reference value for data with ideal normal distribution), whereas the corresponding value for the sports-hall is below 2. Hence, a high value of RIR kurtosis will correspond to high direct to reverberant (D/R) ratio in the RIR and low RT value for the room.

Although the kurtosis provides an overall indication of the room's reverberation, as is recorded in the RIR, variations of source/receiver position within any single room will also appear as a smaller variation in kurtosis due to the alterations in the RIR direct to reverberant ratio. This is shown by analysis of RIRs measured in two different positions in room R2 (classroom), as is illustrated by Fig. 9 and displayed in Table I.

B. Signal-dependent statistical analysis

It is now useful to extend this statistical analysis to the perceptually-relevant signal-dependent parameters defined in the previous sections. Considering the definition of the RMI [see Eq. (3)] and noting that it will be always positive, the numerical results of the $D_{k,n}^{m}$ appear to follow a distribution with general characteristics similar to inverse gamma distribution with a PDF:

$$P(x;\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\cdot x^{-\alpha-1}\cdot e^{-\beta/x},\quad x>0, \tag{16}$$

where α is the shape parameter, β is the scale parameter, and $\gamma(.)$ is the gamma function:

$$\Gamma(\kappa)=\int_{0}^{\infty}t^{\kappa-1}e^{-t}dt,\quad \kappa>0. \tag{17}$$

As was discussed in Sec. III A, the $D_{k,n}^{m}$ values may be considered as random variables (with strictly positive values), thus the corresponding probability function is defined according to

TABLE I. RIR kurtosis for different positions within room R2.

Source/receiver distance		Kurtosis
Position 1	6.6 m	2.2251
Position 2	2 m	2.9531

TABLE II. Properties of rooms and RIRs used for the tests.

Room	Dimensions $L\times W\times H$ (m)	T_{60} (s)	D/R (dB)	Description
R1	$7.15\times4.60\times2.90$	0.368	−0.2013	Acoustically-treated laboratory
R2	$10.20\times7.05\times2.65$	1.1	−3.88	Large classroom
R3	$60\times42\times13.8$	6.4	−8.61	Sports-hall
R4	$30\times25\times14$	1.9	−4.39	1000-seat auditorium and concert hall

$$P(D_{k,n}^{m};\alpha,\beta)=\frac{\beta^{n}}{\Gamma(\alpha)}\cdot(D_{k,n}^{m})^{-\alpha-1}\cdot e^{-\beta/D_{k,n}^{m}},\quad D_{k,n}^{m}>0. \tag{18}$$

Similarly, for the differential metrics, the PDF is defined according to Eq. (14), i.e.,

$$P(D_{k,n}^{c,\ell,t})=P(D_{k,n}^{c,\ell,t};\sigma,\mu), \tag{19}$$

where $D_{k,n}^{c,\ell,t}$ corresponds to differential coherence, differential level, and differential time metrics, these also being considered as random variables.

IV. TESTS AND RESULTS

A. Test method

Tests were conducted using typical stereo 16-bit signals sampled at f_s=44 100 Hz. These tests were used as input (reference) audio signals from different categories, typically wave-files of big band jazz ("jazz"), solo classical piano ("piano"), male speech ("speech"), and solo castanets ("percussion"). As second input in the system of Fig. 1 were the corresponding signals recorded under various reverberation conditions in different rooms ranging from a acoustically-treated laboratory to a large sports-hall, with properties given in Table II.

The log-envelope functions of RIR for rooms R1, R2, and R3 (low-passed with a cut-off frequency of 20 Hz) are shown in Fig. 10.

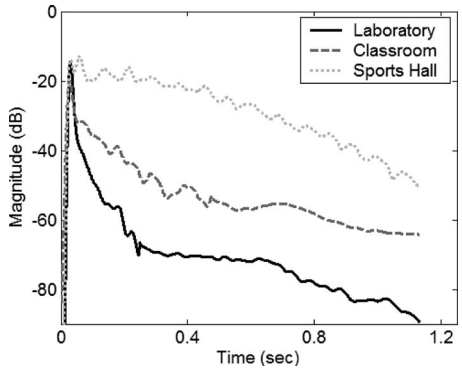


FIG. 10. Low-passed RIR log-envelopes corresponding to rooms R1, R2, and R3.

TABLE III. Monaural RMI $D^m_{k,n}$ (dB) for different real rooms and different audio signals.

Signal	Room R1	R2	R3	R4
Jazz	−12.57	−9.26	−5.03	−6.45
Piano	−7.15	−5.12	−3.93	−4.40
Speech	−3.72	−1.05	0.73	0.04

B. Global signal-dependent alterations

From the set of the 2D maps, derived as was described in Sec. II, the local variations and the overall metrics for each specific test case were evaluated and the results are given in Secs. IV B 1, IV B 2, and IV C.

1. Monaural masking

Table III shows the variation of the RMI $D^m_{k,n}$ (dB) for different audio signals, recorded in the three different rooms. As it is shown, the results for room R3 (large sports-hall) indicate a higher degree of perceived alterations for all types of signal. Furthermore, the estimated value for perceptually-relevant alterations appears to depend also on the input signal. Similarly, within a single room, the evaluated RMI appears to correlate with the source-receiver distance (Table IV), hence detecting features of reverberation-loaded signal due to variation in direct to reverberant energy ratio.

Hence, the proposed masking criterion appears globally to follow well the acoustical properties of the room, to detect the amount of reverberation in the recorded signal, and also to depend on the input signal.

2. Differential inter-channel alterations

Table V shows the variation of the inter-channel differential metrics for the same rooms, using the piano segment as input signal. As it is shown, the variation in the differential inter-channel coherence $D^c_{k,n}$ and the differential inter-channel time $D^t_{k,n}$ metrics between rooms R1 and room R3 is close to 3 dB.

These results indicate that the proposed metrics describe well the variations in the inter-channel cues of the reverberation-afflicted signal and correlate to established acoustical parameters such as the RT. Furthermore, these global metrics seem also to depend on the specific audio signals.

TABLE IV. Monaural RMI $D^m_{k,n}$ (dB) for different positions and different audio signals in room R2.

Signal	Room R2 Position 1 (s/r distance=6.6 m)	Position 2 (s/r distance=2 m)
Jazz	−9.26	−10.41
Piano	−5.12	−7.05
Speech	−1.05	−2.25

TABLE V. Differential metrics for different real enclosures and piano as a test signal.

Differential metric	Room R1	R2	R3
Coherence $D^c_{k,n}$	−32.15	−30.52	−29.30
Level $D^\ell_{k,n}$	0.08	1.64	2.09
Time $D^t_{k,n}$	−26.88	−23.54	−23.99

C. Statistical distributions of signal-dependent metrics

It is now useful to examine the statistics of the proposed metrics. Evaluating histogram distributions for the RMI of Eq. (3) and taking into account the discussion in Sec. III B, the results for different types of audio signals relayed in three different rooms are shown in Figs. 11(a)–11(c).

From these results, it becomes clear that the RMI statistics follow both signal-dependent and system-dependent trend considering the acoustic properties of a room. In the acoustically dry room [room R1, Fig. 11(a)], the alterations remain low-valued and hence are less severe. Given that PDF spread toward higher amplitude values (right-hand side biased) indicates severity of alterations, speech seems to be more sensitive than the music signals, of which, the solo piano signal seems to be more susceptible to such alterations. It is also clearly illustrated that rooms with significant reverberation will increase the occurrence and the severity of alterations, as is shown in Fig. 11(c) and can be also observed in Fig. 5.

From this analysis, it is evident that the proposed metric will display PDFs with right-hand tail mimicking the overall system-dependent RIR envelope decay, as can be observed by comparison of Fig. 11 with Fig. 10. However, it is also clear that the modeled alterations will also greatly depend on the specific audio signal. Position-sensitivity within a single room is also evident in Figs. 11(d)–11(f) for results obtained in room R2 and for a position of measurement close to the source (at 2 m, Position 2), as opposed to a far position (at 6.6 m, Position 1). From these results it becomes clear that the speech signal is far more susceptible to alterations due to reverberation and also that all signals display more severe alterations for the far measuring position. In both cases the characteristics of the corresponding distributions are approximating an inverse-gamma distribution and somehow relate to the overall shape of the RIR reverberation tail (see Fig. 10), as it is masking the reproduced signal. The exact pattern of the distributions appears to be determined by the shape parameter α and scale parameter β [see Eq. (18)]. For example, large values of the scale parameter β denote that the distribution is more spread out (e.g., the speech signal in room R1). Accordingly, small values of β indicate that the distribution is relative peaked (e.g., the jazz signal in room R2).

Similarly, the shape parameter α will allow the distribution to take a variety of shapes. Clearly, for all the test cases

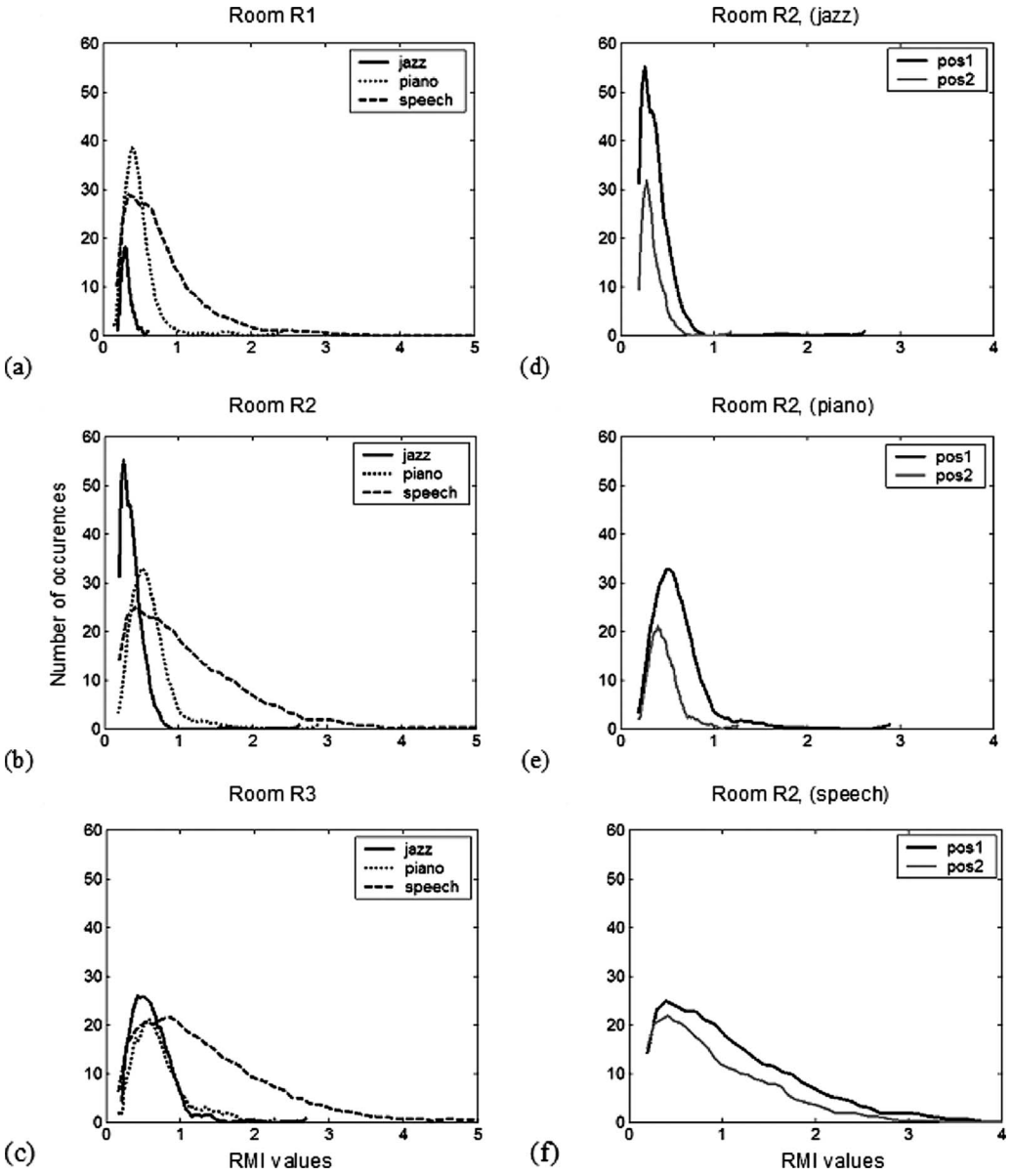


FIG. 11. PDFs of the monaural RMI $D^m_{k,n}$: [(a)–(c)] different signals in rooms R1, R2, and R3; [(e)–(g)] different signals in room R2 for 2 positions (Pos1: s/r distance=6.6 m; Pos2: s/r distance=2 m).

considered, the general trend of the distributions was that of a right-skewed distribution (e.g., relative long right-hand tail, see Fig. 11). Similar trends were also observed for the differential metrics, but unlike the monaural RMI discussed previously, these distributions take both positive and negative values. Figure 12 shows the variation of the interchannel differential metrics for the above enclosures, using the jazz and the percussion as input signals. The variations are plotted as distributions (as described in Sec. III B) and can be interpreted as divergence from a typical normal distribution.

The results shown in Fig. 12 indicate that the statistics of the differential perceptual criteria are also influenced by a combination of room acoustics (see Appendix) and the properties of the specific audio signal used for the test. In most cases, reverberation will bias the differential metrics away from the zero mean value associated with system-dependent acoustic metrics, analyzed in Sec. III A. This describes a systematic shift in the corresponding differential spatial cue. Nevertheless, critical source signals such as the percussion seem to exhibit a histogram that is increasingly diverting from that of a normal distribution. In this case, it is likely that the two distinct distributions observed in Fig. 12(b) are due to the separate contributions of intervals where signal (castanet) is present and for intervals without any signal present (silence and reverberation intervals).

V. DISCUSSION

The present work has described a method for representing perceptually-relevant alterations of audio signals due to reverberation with appropriate 2D maps. The results for the proposed reverberation masking maps and index show that the frequency-averaged masking metric increases during the decay of the reverberation tail of a reverberation-carrying signal, indicating the increase in the signal-induced masking

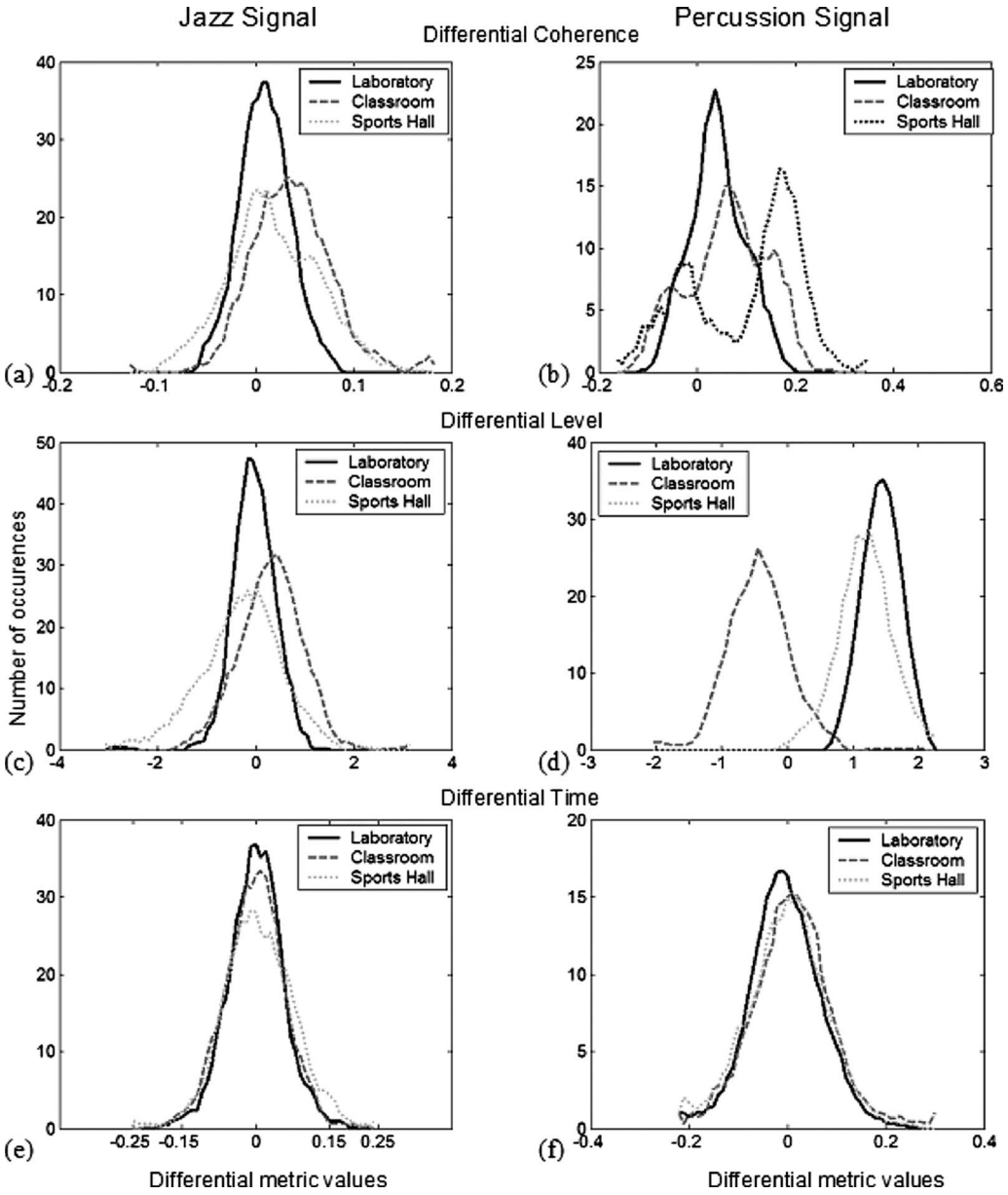


FIG. 12. PDFs for differential metrics $D_{k,n}^{c,\ell,t}$ and for jazz as a test signal [(a), (c), and (e)], or percussion as a test signal [(b), (d), and (f)], in room R1, room R2, and room R3.

due to reverberation tail. The variation of the RMI depends also on room acoustics, the value of the RMI increasing with RT. Hence, as expected, the masking effects of reverberation are more pronounced for the larger rooms and/or the more distant receiver positions. For the proposed differential inter-channel metrics, which describe alterations of the spatial cues, similar overall trends can be observed. For low reverberation, all differential inter-channel metrics display low values. With heavy reverberation, higher differences are observed in these maps, indicating increased severity of the alterations of the original spatial cues in the received stereo signal. When evaluated over a long time interval, these metrics seem to correlate qualitatively well with the degree of acoustical degradation measured via established acoustical criteria.

To assess more formally such trends, the present work has also introduced an efficient and useful statistical analysis of room reverberation and the perceptually-relevant effects. It was here confirmed that from system-dependent perspective, the specific acoustic properties of the RIRs can be described by the corresponding amplitude PDFs and are mani-

TABLE VI. Parameters for the CAMM.

Parameter	Description	Value
c_j	Weighting factors for the realization of $h_{\text{norm}}(n)$ filter	$c_1=0.227$ $c_2=0.278$ $c_3=0.494$
n_j	Time constants involved in $h_{\text{norm}}(n)$ filter	$n_1=2.5\ \text{ms}\cdot f_s$ $n_2=10\ \text{ms}\cdot f_s$ $n_3=75\ \text{ms}\cdot f_s$
g_k	Gain factor inherent in signal dependent compression (SDC) module	$g_k=1/7$

TABLE VII. Mean value μ and variance σ^2 of the PDFs of the differential metrics for jazz as test signal, in room R1, room R2, and room R3.

	Differential coherence			Differential level			Differential time		
Parameter	R1	R2	R3	R1	R2	R3	R1	R2	R3
μ	0.01	0.037	0.018	−0.0391	0.3056	−0.2961	0.001	−0.002	0.0064
σ^2	$6.27 \cdot 10^{-4}$	0.0016	0.0021	0.1771	0.4074	0.6375	0.003	0.0036	0.0050

fested as significantly varying spread from the mean value. It was found that the kurtosis of the corresponding RIRs describes well these acoustic trends; an acoustically-treated laboratory with near-perfect acoustics has a kurtosis value above 3 (the reference value for data with ideal normal distribution), whereas the corresponding value for a highly reverberation-charged sports-hall is below 2.

This statistical analysis of room responses was then extended to the signal-dependent parameters introduced in this work. Evaluating histogram distributions for the monaural RMI and for different audio signals and rooms, it became clear that this follows both signal-dependent and system-dependent trends. In an acoustically near-perfect room the RMI remains low-valued, being different for each class of signals. Source-receiver position-sensitivity within a single room was also evident, since for all signals more severe perceptually-relevant alterations were measured for the far positions, again the speech signal being more susceptible to such effects. Overall, these PDFs are closely resembling inverse gamma histograms, with the distribution's tail somehow mimicking reverberation decay in the RIR reverberation tail. Hence the severity of the reverberation tail, as it affects a specific audio/speech signal, appears to be well-represented by the proposed perceptually-motivated masking metric. The statistics of the differential spatial cues were similarly influenced by a combination of room acoustics and the properties of the audio signals. In most cases, reverberation was found to bias the differential metrics away from the zero mean value associated with system-dependent acoustic metrics, increased reverberation being manifested as lowering the distribution's kurtosis. These shifts in PDFs give a good illustration of the broad alterations in the spatial attributes imposed by the room on any stereo signal.

VI. CONCLUSION

The present work introduces a novel framework describing some perceptually-relevant effects of reverberation on the reproduced acoustic signal within rooms. It is promising that statistical results derived via the proposed metrics seem to follow the trend of the established physical acoustical parameters for the reverberation-charged space. However, unlike existing acoustical measurements, the proposed maps are dynamically varying with each signal's evolution and are dependent on the specific audio signal in a way that seems to be compliant with perceived alterations of the reproduced audio. This detailed detection of signal distortions may enable the introduction of signal processing techniques for treatment or compensation of such effects. Therefore, such maps may help to reconsider the problem of reverberation from a signal-dependent processing perspective that is robust, efficient, closer to perception, and appropriate for each specific signal reproduced inside any room. Future work will consider signal adaptation based on the proposed masking index and differential cue mapping, so that perceptually robust acoustic signal reproduction may be realized within reverberation-charged rooms.

APPENDIX

1. Typical values of critical parameters of CAMM

The values of the critical parameters for the realization of the CAMM used in this work, are presented in Table VI (f_s is the sampling frequency in kilohertz).

The $h_{\text{norm}}(n)$ filter is composed of a summation of L (here $L=3$) exponential functions with time constants n_j and weighting factors c_j, i.e., $h_{\text{norm}}(n)=\Sigma_{j=1}^{L} c_j \cdot e^{n/n_j}$.

2. Relationship between the differential metric statistical parameters and room acoustic

From the statistical analysis presented in Sec. III and the differential metric distributions depicted in Fig. 12, Tables VII and VIII list the corresponding statistical parameters.

From these tables and the measured RT of each room (see Table II), Figures 13(a) and 13(b) show the relationship between differential metric statistics and room acoustics. As it can be observed, for both audio audio test signals the pro-

TABLE VIII. Mean value μ and variance σ^2 of the PDFs of the differential metrics for percussion as test signal, in room R1, room R2, and room R3.

	Differential coherence			Differential level			Differential time		
Parameter	R1	R2	R3	R1	R2	R3	R1	R2	R3
μ	0.0417	0.0616	0.0976	1.4431	−0.4256	1.1798	0.005	0.0094	−0.0024
σ^2	0.003	0.0066	0.0127	0.0721	0.1422	0.1568	0.002	0.0061	0.062

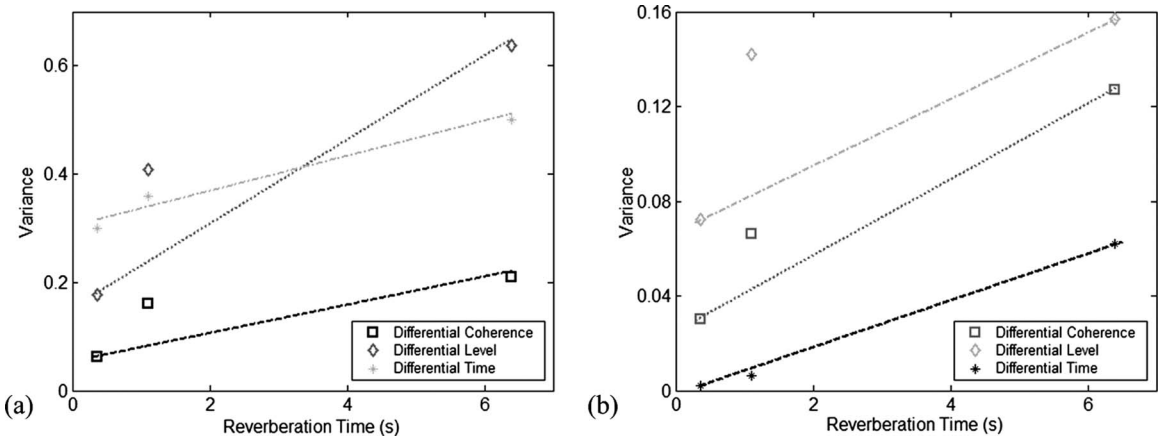


FIG. 13. Variance of differential metrics vs RT, for (a) jazz as a test signal and (b) percussion as test signal. Note that statistical data were scaled for illustrative purposes.

posed metric statistical variances increase with room reverberation. Note that some of the statistical data were appropriately scaled for illustrative purposes.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Dr. J. Buchholz of Ørsted DTU, Copenhagen, in developing the CAMM and Dr. P. Hatziantoniou, University of Patras, Greece, for the RIRs measurements employed in this work. They also acknowledge the anonymous reviewers for their constructive comments and helpful suggestions.

[1]M. R. Schroeder and B. F. Logan, "Colorless artificial reverberation," J. Audio Eng. Soc. **9**, 192–197 (1961).
[2]B. S. Atal, M. R. Schroeder, G. M. Sessler, and J. E. West, "Evaluation of acoustic properties of enclosures by means of digital computers," J. Acoust. Soc. Am. **40**, pp. 428–440 (1966).
[3]R. Wyber, "The application of digital processing to acoustic testing," IEEE Trans. Acoust., Speech, Signal Process. **22**, 66–72 (1974).
[4]J. Blauert, *Spatial Hearing: The Psychophysics of Human Localization* (MIT, Cambridge, 1997).
[5]T. Okano, L. L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E) and apparent source width (ASW) in concert halls," J. Acoust. Soc. Am. **104**, 255–265 (1998).
[6]J. Blauert and W. Lindemann, "Auditory spaciousness: Some further psychoacoustic analyses," J. Acoust. Soc. Am. **80**, 533–542 (1986).
[7]H. Kuttruff, *Room Acoustics* 2nd ed. (Applied Science, London, 1979).
[8]J. M. Jot and A. Chaigne, "Analysis and synthesis of room reverberation based on statistical time-frequency model," in *Proceedings of the AES 103rd International Convention*, New York (1997).
[9]J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal processing technique to remove room reverberation from speech signals," J. Acoust. Soc. Am. **62**, 912–915 (1977).
[10]M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," IEEE Trans. Acoust., Speech, Signal Process. **36**, 145–152 (1988).
[11]Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalisation of room transfer functions by using common acoustical poles," IEEE Trans. Speech Audio Process. **5**, 325–333 (1997).
[12]M. Karjalainen, T. Paatero, J. Mourjopoulos, and P. D. Hatziantoniou, "About room response equalization and dereverberation," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE WASPAA'05)*, New York (2005).
[13]J. L. Flanagan and R. C. Lummis, "Signal processing to reduce multipath distortions in small rooms," J. Acoust. Soc. Am. **47**, 1475–1481 (1970).
[14]J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," J. Sound Vib. **102**, 217–228 (1985).
[15]R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," J. Acoust. Soc. Am. **21**, 577–580 (1949).
[16]J. M. Buchholz, J. Mourjopoulos, and J. Blauert, "Room masking: Understanding and modeling the masking of room reflections," presented at the 110th Convention of the Audio Engineering Society, Amsterdam (2001), Preprint 5312.
[17]F. E. Toole, "Loudspeakers and rooms for sound reproduction—A scientific review," J. Audio Eng. Soc. **54**, 451–476 (2006).
[18]T. Djelani and J. Blauert, "Investigations into the build-up and breakdown of the precedence effect," Acta. Acust. Acust. **87**, 253–261 (2001).
[19]R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654 (1999).
[20]C. Faller, "Parametric multichannel audio coding: Synthesis of coherence cues," IEEE Trans. Audio, Speech, Lang. Process. **14**, 299–310 (2006).
[21]C. Avendano and J. M. Jot, "Ambience extraction and synthesis from stereo signals for multichannel audio upmix," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Orlando, FL (2002).
[22]F. Baumgarte and C. Faller, "Binaural cue coding—Part I: Psychoacoustic fundamentals and design principles," IEEE Trans. Speech Audio Process. **11**, 509–519 (2003).
[23]C. Faller and F. Baumgarte, "Binaural cue coding—Part II: Schemes and applications," IEEE Trans. Speech Audio Process. **11**, 520–531 (2003).
[24]V. Pulkki, "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc. **55**, 503–516 (2007).
[25]J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," EURASIP J. Appl. Signal Process. **2005**, 1305–1322 (2005).
[26]C. Faller, "Parametric joint-coding of audio sources," presented at the 120th Convention of the Audio Engineering Society, Paris, France (2006), Preprint 6752.
[27]D. Griesinger, "Objective measures of spaciousness and envelopment," presented at the 16th Audio Engineering Society International Conference: Spatial Sound Reproduction, Finland (1999), Paper No. 16-003.
[28]D. Griesinger, "Measures of spatial impression and reverberance based on the physiology of human hearing," presented at the 11th Audio Engineering Society International Conference: Test and Measurement, Portland, OR (1992), Paper No. 11-016.
[29]J. S. Bradley, "Comparison of concert hall measurements of spatial impression," J. Acoust. Soc. Am. **96**, 3525–3535 (1994).
[30]K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality," J. Acoust. Soc. Am. **74**, 1726–1733 (1983).
[31]C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089 (2004).
[32]K. Lehn, "Fuzzy temporal-cluster analysis of monaural and interaural cues as a model of auditory scene analysis," Ph.D. thesis, Ruhr-Univ. Bochum (2000).
[33]S. Harding, J. Barker, and G. J. Brown, "Mask estimation for missing data speech recognition based on statistics of binaural interaction," IEEE Trans. Audio, Speech, Lang. Process. **14**, 58–67 (2006).
[34]G. J. Brown, S. Harding, and J. P. Barker, "Speech separation on the statistics of binaural auditory features,"in *Proceedings of ICASSP*, Toulouse, France (2006).

[35]J. M. Buchholz and J. Mourjopoulos, "A computational auditory masking model based on signal-dependent compression. I. Model description and performance analysis," Acta. Acust. Acust. **90**, 873–886 (2004).
[36]J. M. Buchholz and J. Mourjopoulos, "A computational auditory masking model based on signal-dependent compression. II. Model simulations and analytical approximations," Acta. Acust. Acust. **90**, 887–900 (2004).
[37]R. Mason and F. Rumsey, "Interaural time difference fluctuations: Their measurement, subjective perceptual effect and application in sound reproduction," presented at the 19th Audio Engineering Society International Conference: Surround Sound-Techniques, Technology, and Perception, Schloss Elmau, Germany (2001), Paper No. 1894.
[38]T. Zarouchas, J. Mourjopoulos, J. Buchholz, and P. Hatziantoniou, "A perceptual measure for assessing and removing reverberation from audio signals," presented at the 120th Convention of the Audio Engineering Society, Paris, France (2006), Preprint 6702.
[39]T. Zarouchas and J. Mourjopoulos, "Perceptual distortion maps for room reverberation," presented at the 122nd Convention of the Audio Engineering Society, Vienna, Austria (2007), Preprint 7093.
[40]J. Herre, E. Eberlein, H. Schott, and K. Brandenburg, "Advanced audio measurement system using psychoacoustic properties," presented at the 92nd Convention of the Audio Engineering Society, Vienna, Austria (1992), Preprint 3321.
[41]W. Rix, J. G. Beerends, D. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality—Technology and applications," IEEE Trans. Audio, Speech, Lang. Process. **14**, 1890–1901 (2006).
[42]J. H. Plasberg and W. B. Kleijn, "The sensitivity matrix: Using advanced auditory models in speech and audio processing," IEEE Trans. Audio, Speech, Lang. Process. **15**, 310–319 (2007).
[43]J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," IEEE Trans. Inf. Theory **45**, 1082–1091 (1999).
[44]M. R. Schroeder, "Modulation transfer functions: Definition and measurement," Acustica **49**, 179–182 (1981).
[45]http://www.wcl.ee.upatras.gr/audiogroup/Subjective/audio_demos.html (Last viewed November, 2008).
[46]M. R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," J. Audio Eng. Soc. **35**, 299–306 (1987).
[47]R. V. Waterhouse, "Statistical properties of reverberant sound fields," J. Acoust. Soc. Am. **43**, 1436–1444 (1968).
[48]D. Lubman, "Fluctuations of sound with position in a reverberant room," J. Acoust. Soc. Am. **44**, 1491–1502 (1968).
[49]R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," IEEE Trans. Speech Audio Process. **13**, 845–856 (2005).
[50]B. W. Gillespie, H. S. Malvar, and D. A. F. Florencio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," in *Proceedings of ICASSP*, Salt Lake City, UT, (2001).
[51]B. Yegnanarayana and P. S. Murthy, "Enhancement of reverberant speech using LP residual signal," IEEE Trans. Speech Audio Process. **8**, 267–281 (2000).
[52]M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," IEEE Trans. Audio, Speech, Lang. Process. **14**, 774–784 (2006).
[53]D. T. Fee, C. F. N. Cowan, S. Bilbao, and I. Ozcelik, "Predictive deconvolution and kurtosis maximization for speech dereverberation," presented at the 14th European Signal Processing Conference, EUSIPCO, Florence, Italy (2006).
[54]K. Furuya, S. Sakauchi, and A. Kataoka, "Speech dereverberation by combining MINT-based blind deconvolution and modified spectral subtraction," in *Proceedings of the ICASSP*, Toulouse, France (2006).
[55]D. Extra, U. Simmer, S. Fischer, and J. Bitzer, "Artificial reverberation: Comparing algorithms by using monaural analysis tools," presented at the 121st AES Convention, San Francisco (2006), preprint 6298.
[56]V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," Acta. Acust. Acust. **88**, 433–442 (2002).
[57]T. Dau, B. Kollmeier, and A. Kohlrausch, "Modeling auditory processing of amplitude modulation. II Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919 (1997).
[58]S. E. Olive and F. E. Toole, "The detection of reflections in typical rooms," J. Audio Eng. Soc. **37**, 539–553 (1989).
[59]M. Kendall, A. Stuart, and J. K. Ord, *Advanced Theory of Statistics: Distribution Theory*, 6th ed. (Hodder Arnold, London, 1994).
[60]J. Usher and J. Benesty, "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer," IEEE Trans. Audio, Speech, Lang. Process. **15**, 2141–2150 (2007).

Finite element modeling of sound transmission with perforations of tympanic membrane

Rong Z. Gan,[a] Tao Cheng,[b] Chenkai Dai,[c] and Fan Yang
School of Aerospace and Mechanical Engineering and Bioengineering Center, University of Oklahoma, Norman, Oklahoma 73019

Mark W. Wood
Hough Ear Institute, 3400 N.W. 56th Street, Oklahoma City, Oklahoma 73112

(Received 8 December 2008; revised 11 April 2009; accepted 13 April 2009)

A three-dimensional finite element (FE) model of human ear with structures of the external ear canal, middle ear, and cochlea has been developed recently. In this paper, the FE model was used to predict the effect of tympanic membrane (TM) perforations on sound transmission through the middle ear. Two perforations were made in the posterior-inferior quadrant and inferior site of the TM in the model with areas of 1.33 and 0.82 mm^2, respectively. These perforations were also created in human temporal bones with the same size and location. The vibrations of the TM (umbo) and stapes footplate were calculated from the model and measured from the temporal bones using laser Doppler vibrometers. The sound pressure in the middle ear cavity was derived from the model and measured from the bones. The results demonstrate that the TM perforations can be simulated in the FE model with geometrical visualization. The FE model provides reasonable predictions on effects of perforation size and location on middle ear transfer function. The middle ear structure-function relationship can be revealed with multi-field coupled FE analysis.
 [DOI: 10.1121/1.3129129]

PACS number(s): 43.64.Ha, 43.64.Bt [BLM] Pages: 243–253

I. INTRODUCTION

The human middle ear includes the eardrum or tympanic membrane (TM) and three ossicles (malleus, incus, and stapes) that are suspended in an air-filled cavity by suspensory ligaments and muscles and connected by two joints (incudo-mallear and incudo-stapedial joints). Sound waves collected in the ear canal are passed to the middle ear through the vibration of the TM, which initiates acoustic-mechanical transmission in the ear. The output of the middle ear system is the movement of the stapes footplate, which sits in the oval window and transmits the ossicular vibration into cochlear fluid. Therefore, transfer function of the middle ear is commonly described as the relationship between vibrations of the TM and stapes footplate in response to sound stimuli in the ear canal (e.g., Zwislocki, 1962; Merchant *et al.*, 1996; Aibara *et al.*, 2001; Gan *et al.*, 2004b).

Finite element (FE) method as a general numerical procedure has distinct advantages in modeling a complex biological system such as the ear. Since the first FE model of the cat TM was reported in 1978 (Funnell and Laszlo, 1978), FE modeling of static and dynamic behaviors of the middle ear has become a fast growing research area in the study of ear mechanics (e.g., Wada and Metoki, 1992; Koike and Wada, 2002; Sun *et al.*, 2002; Kelly *et al.*, 2003; Gan *et al.*, 2004a; Gan and Wang, 2007). Using the technologies of three-dimensional (3D) reconstruction and multi-field FE coupled analysis, we have recently developed a comprehensive FE model of the human ear including the external ear canal, middle ear, and uncoiled cochlea with two straight fluid channels separated by the basilar membrane (BM) (Gan *et al.*, 2007). The acoustic-structure-fluid coupled FE analysis or "three-chamber" multi-field FE modeling, including the air in the ear canal and middle ear cavity, the TM and middle ear ossicular structures, and the fluid in the cochlea, has been developed with the model. With the complete middle ear cavity connected to cochlea through the oval window and round window, modeling of middle ear transfer function in the ear with TM perforation becomes possible. In this study, we simulate TM perforations in the FE model to predict the perforation-induced change of middle ear function. This study evaluates a complex combination of two sound conduction routes: the mechanical route through the ossicular chain and the acoustic route through the air in the middle ear cavity.

Perforation of the TM is frequently caused by otitis media and Eustachian tube dysfunction, head trauma, and blast exposure. The reduction in energy transfer efficiency caused by TM perforation has been measured in human temporal bones and simulated in a simple circuit model by Voss *et al.* (2001a; 2001b; 2001c; 2007) with a series of publications. It has been tested in animals (Bigelow *et al.*, 1996; Santa Maria *et al.*, 2007) and clinical studies (Ahmad and Ramani, 1979; Mehta *et al.*, 2006). However, the circuit model does not have structural geometry, and the parameters of the model are determined by fitting model results with experimental

[a]Author to whom correspondence should be addressed. Electronic mail: rgan@ou.edu
[b]Present address: Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA 02114.
[c]Present address: Department of Otolaryngology/Head and Neck Surgery, Johns Hopkins University, Baltimore, MD 21218.

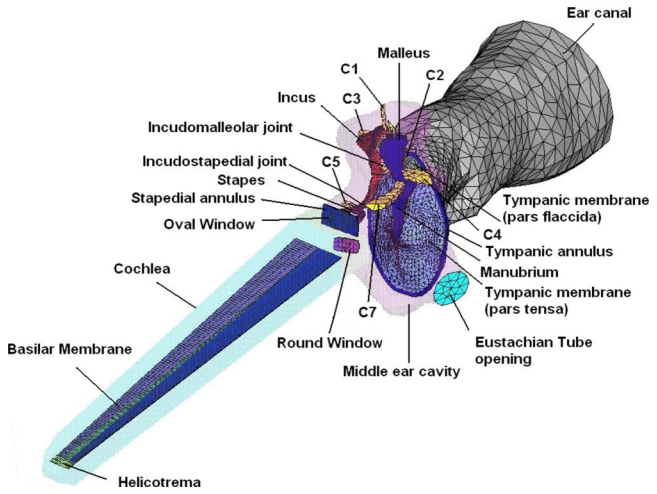


FIG. 1. (Color online) FE model of human left ear including the external ear canal, the middle ear [TM three ossicles (malleus, incus, and stapes), two joints and manubrium, ligaments and muscle tendons, tympanic annulus, stapedial annular ligament, and middle ear cavity], and the uncoiled cochlea in anterior-medial view. The middle ear cavity and cochlear chambers were assumed transparent. Here, C1, C2, C3, C4, C5, and C7 stand for the superior mallear ligament, lateral mallear ligament, posterior incus ligament, anterior mallear ligament, stapedial tendon, and tensor tympani tendon, respectively.

measurements in each individual temporal bone. In this paper, we will demonstrate that the FE model can be used to simulate various perforations with different sizes and different locations and to predict the effect of TM perforation on sound transmission from the ear canal to cochlea.

Perforations were created in the model as well as in human cadaver temporal bones. Two laser Doppler vibrometers were used to measure simultaneously the TM and stapes footplate movements in temporal bones. The FE model-derived results were compared with the measurements obtained from the bones. The study reported here is considered as a step toward the potential clinical applications of the FE model on prediction of structure alteration-induced hearing loss.

II. METHODS

A. 3D finite element model

1. Brief description of the FE model

A 3D FE model of the human left ear with structures of the external ear canal, middle ear, and cochlea was developed recently by Gan *et al.* (2007). The model has complete geometry of the external ear canal and middle ear, including the TM, ossicles, suspensory ligaments and muscle tendons, and the middle ear cavity (Fig. 1). An uncoiled cochlear model was connected to the middle ear. The geometry of the cochlea was based on published dimensions of human cochlea in the literature. The scala vestibuli was connected to the stapes at the oval window and the scala tympani was contacted to the middle ear cavity air at the round window. These two chambers were separated by the BM and filled with an incompressible viscous fluid (perilymph).

The ossicles, ligaments, and tendons were assumed as isotropic materials while the TM was assumed as orthotropic material. The mechanical properties of the TM, ossicles, joints, and manubrium in the model are listed in Table 1 of the paper of Gan *et al.* (2006). Poisson's ratio was assumed to be 0.3 for all materials of the system. The Rayleigh damping parameters α and β for the middle ear system were assumed to be 0 s^{-1} and 0.75×10^{-4} s, respectively. The human middle ear model was described as a linear acoustic-mechanical transmission system for sound energy within the normal hearing range.

The boundaries of the TM and middle ear ossicular chain include the tympanic annulus, middle ear suspensory ligaments or muscle tendons, stapedial annular ligament, and cochlea. Young's moduli of suspensory ligaments/tendons, tympanic annulus, and stapedial annular ligament were the same as that used in our previous analysis (Gan *et al.*, 2007). Material properties used for cochlear structures, including the density and Young's modulus of the oval and round windows, the inner and outer supports of the BM along the cochlear partition, and the stiffness of the BM, can be found from Gan *et al.* (2007). The fluid inside the scala vestibuli and scala tympani in cochlea was assumed as a viscous fluid with a density of 1000 kg/m^3 and a viscosity of 0.001 Ns/m^2 or 1 cP. The damping coefficient β for the fluid was assumed as 1.0×10^{-4} s. Since the ear canal was open to the atmosphere in temporal bone experiment setup in this study, the boundary condition at the ear canal entrance of the FE model was set free. The Eustachian tube was modeled as being blocked similar to the experimental condition of the temporal bones.

2. TM perforations simulated in the FE model

Two perforations were made in the posterior-inferior quadrant and the inferior site of the TM in the model, as shown in Fig. 2A. The perforation size for Hole 1 were approximately 1.3 mm in diameter with a surface area of 1.33 mm^2 and for Hole 2 was 1.0 mm in diameter with an area of 0.82 mm^2. It is noted that the edge of the hole in the model was not smooth because of the element size. Similar holes were created in the temporal bones, in the same locations. The size of the hole created in the bone was approximately equal to the size of the hole in the model. In addition to Hole 1 and Hole 2, two other perforations were made in the model at the same location as Hole 1 (the posterior-inferior site). The size of Hole 3 [Fig. 2B] was approximately 1.65 mm in diameter with a surface area equal to the sum of Holes 1 and 2 (2.15 mm^2). The area of Hole 4 [Fig. 3C] was approximately double that of Hole 3 at 4.34 mm^2 with an approximate diameter of 2.35 mm. The purpose for Hole 3 was an attempt to investigate the effect of multiple perforations (Holes 1 and 2) versus a single perforation (Hole 3) with the same perforation surface area. However, comparison of Holes 1 and 2 with Hole 3 may be also affected by different location. The design for Hole 4 was to study the effect of perforation size (from Hole 1 to Hole 3 and Hole 4) in the same location on middle ear function for sound transmission. It is noted that the size and shape of holes made in the model and temporal bones was approximately similar and the boundary or edge of the holes was not smooth, as shown in

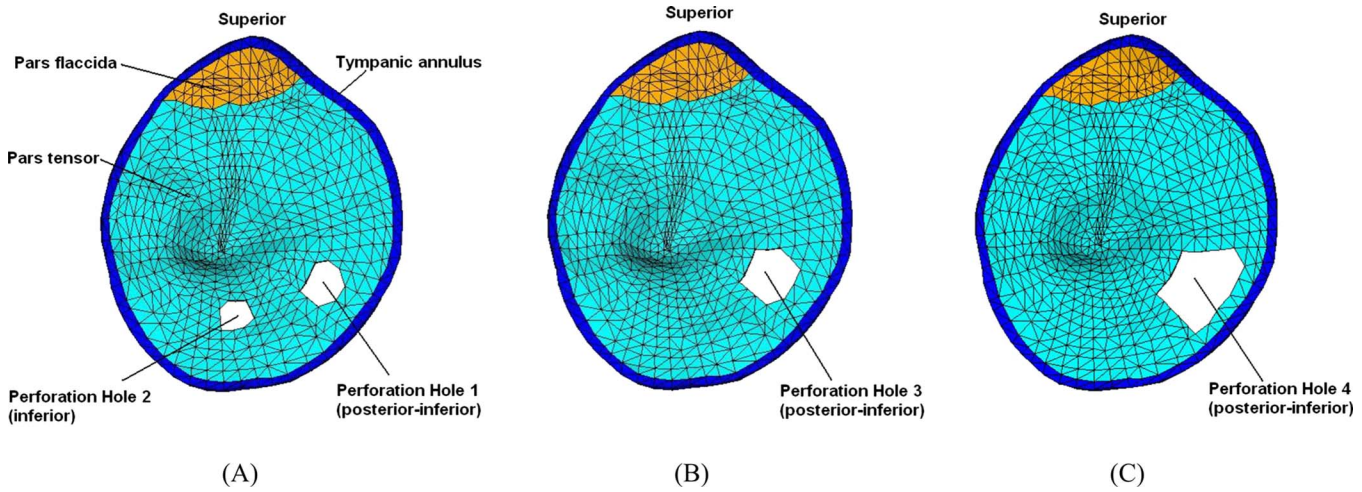


FIG. 2. (Color online) Lateral view of the TM with perforations. The sizes of perforations are given in the text. (A) Two perforations: Hole 1—the perforation located in the posterior-inferior quadrant of the TM; Hole 2—the perforation located in the inferior part. (B) Single perforation Hole 3 in the posterior-inferior site. (C) Single perforation Hole 4 in the posterior-inferior site.

Fig. 2. These limitations were due to the element size of the TM model and the tool used for creating the holes in the temporal bones.

B. Multi-field coupled FE analysis

In this study, the multi-field coupled analysis, including the air in the ear canal and middle ear cavity, the structures of the TM, middle ear ossicles and cochlea, and the fluid inside cochlea, was conducted on the FE model. The air in the ear canal and middle ear cavity was modeled as acoustic elements and governed by the simplified acoustic wave equation under the assumptions that the fluid (air) is compressible and inviscid with uniform mean density and pressure:

$$\frac{\partial^2 P}{\partial t^2} - c^2 \nabla^2 P = 0, \tag{1}$$

where P is acoustic pressure, c is speed of sound and $c = \sqrt{k/\rho_o}$ in fluid medium, ρ_o is mean fluid density, k is bulk modulus of fluid, and t is time. The speed of sound and density of the air were assumed as 343 ms^{-1} and 1.21 kgm^{-3}, respectively.

The surface of acoustic elements (air) next to a fixed structure, such as the canal and middle ear cavity bony walls, was defined as "impedance surface" and assigned with a specified acoustic absorption coefficient μ (Gan *et al.*, 2007). The surface of acoustic elements next to the movable structure, such as the TM (the lateral and medial surfaces of the TM and the surface around the holes on the TM), ossicles, and suspensory ligaments, was defined as a fluid-structural interface (FSI) where the acoustic pressure distribution was coupled into structural analysis as the force input in ANSYS (ANSYS Inc., Canonsburg, PA). The TM (both pars tensa and pars flaccida) had two FSIs on its lateral and medial sides, respectively. The round window membrane had two FSIs, one on its middle ear cavity side coupled with the air elements and one on the scala tympani side coupled with viscous fluid. On both sides of the BM (i.e., scala vestibuli and scala tympani), fluid that interfaced with solid structure nodes was kept distinct but coincident. These nodes were coupled together to form a no-slip condition for the fluid and to create the FSIs. The motion of fluid is described by the equation,

$$P = -k \nabla \cdot \boldsymbol{u}, \tag{2}$$

where **u** is displacement vector, k is bulk modulus of fluid and $k = c^2\rho$. In this study, k was assumed as 220 GPa for perilymphatic fluid in the cochlea.

The harmonic analysis over the auditory frequency range of 100 Hz–10 kHz was conducted in the model using ANSYS v. 11. A sound pressure of 90 dB sound pressure level

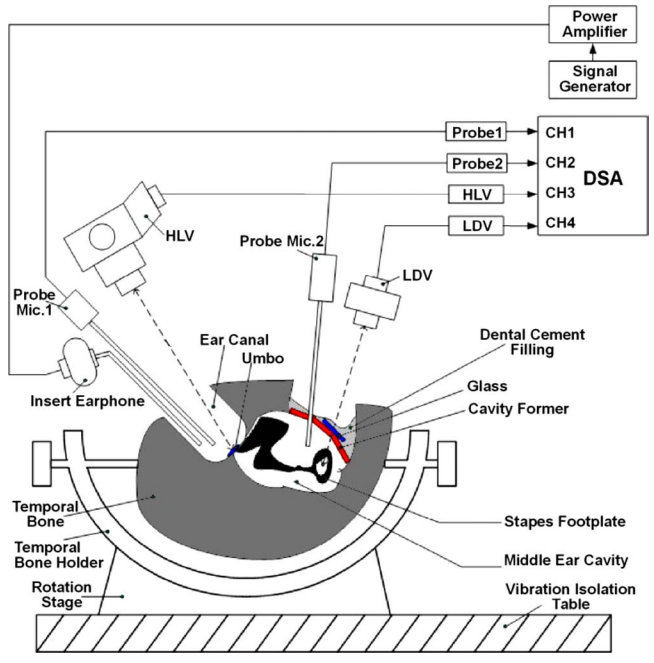


FIG. 3. (Color online) Schematic diagram of the experimental setup in human temporal bone with two laser vibrometers for measuring vibrations at the TM (umbo) and stapes footplate simultaneously. Two probe microphones were used for monitoring sound pressures in the ear canal and middle ear cavity.

(SPL) (0.632 Pa or N/m^2, rms value) was applied at the nodes of acoustic elements (Fluid 30) in the ear canal at 2 mm away from the TM at the umbo. This is the same situation as the experimental setup for human temporal bones in our laboratory. Our previous study on acoustic pressure distributions in the ear canal of the FE model (Gan *et al.*, 2006) demonstrates that the pressure distributions reflect superposition of the incident and reflected sound waves from the TM and canal wall in the canal. The superposition is closely related to the frequency and location of the input sound source. The coupled FE analysis resulted in displacement movements of the TM (umbo) and stapes footplate as well as acoustic pressure inside the middle ear cavity.

C. Measurement of middle ear transfer function on human temporal bones

Five fresh-frozen, cadaveric temporal bones (two male and three female) obtained through the University of Oklahoma Health Sciences Center were included in this study. The donors' average age was 73.8. The preparation of temporal bone specimen was the same as reported in our previous papers (Gan *et al.*, 2004a, 2004b). The experimental setup for this study is schematically shown in Fig. 3. Briefly, after performing a simple mastoidectomy and extended facial recess approach on the bone, 90 dB SPL pure tone sound (P1) from a function generator (Model 193, Wavetek, San Diego, CA) was delivered to the TM by an insert earphone (Model ER-2, Etymotic Research, Elk Grove Village, IL) and monitored by two probe microphones (Model ER-7, Etymotic Research). The first probe tube was placed in the canal approximately 2 mm from the umbo. The second probe was placed between the round window and stapes for measuring the middle ear sound pressure (P2). The middle ear cavity was then covered by a glass sheet and sealed by filling dental cement (Reprosil, DDI Inc., Milford, DL). This glass sheet removed most of the mastoid cavity from the experimental volume.

Two laser vibrometers (Polytec, HLV-1000 and LDV-OFV 501) were used to measure vibrations of the TM and stapes simultaneously with one laser focused on the umbo and the other focused on the stapes footplate. The HLV laser beam was directed through the ear canal at the reflective tape on the umbo. Deviation of this laser beam varied 0° to 35° with respect to the direction of the stapes piston-like movement (i.e., the direction perpendicular to the plane of the footplate). In our experiments, the average angle was about 30°. The LDV laser was aimed on the reflective tape placed at the center of the stapes footplate. The deviation of the laser beam with respect to the piston-like direction of the stapes movement varied 30° to 55° and an average angle of 50° was used for this study. The deviation angles of laser beams were used as cosine correction factor to obtain the umbo and footplate displacements along the stapes piston-like direction.

Control experiments were performed first to measure the normal middle ear function with the intact TM. After control data were collected, two perforations were created [see Fig. 2A] using the cauterizer (Fine Science Tools, Inc., Foster, CA). It is noted that the edge of the hole made by cauterizer was not as smooth as that made by argon laser used by Voss *et al.* (2001a). The laser measurements of the umbo and footplate were conducted with Hole 1 first. The perforation Hole 2 was then made and Hole 1 was patched with cigarette paper (Voss *et al.*, 2001a). To assure a patch on Hole 1 was effectively blocking the sound, the laser measurement on the umbo and stapes footplate was conducted before Hole 2 was made. The results showed that the difference between the intact TM and Hole 1 patched with the paper was within 1 dB. Finally, the paper was removed and the measurement was made with a combination of both Hole 1 and Hole 2.

III. RESULTS

The FE model of a normal ear with the intact TM was first validated by comparison of the displacement curves of the TM (at the umbo) and stapes footplate derived from the model with that measured from human temporal bones. The magnitudes of the TM and footplate displacements were calculated from the model and projected to the stapes piston-like direction, the same direction as the data reported from temporal bones using laser vibrometers. Figure 4 shows the model-derived frequency response curves of the TM and footplate displacements (thick broken lines) in comparison with the umbo and stapes footplate displacement curves measured from five temporal bones with intact TMs when 90 dB SPL sound was applied in the ear canal. Panels (A) and (B) display the magnitudes of displacement data and panels (C) and (D) show the phase angles. The experimental data were consistent with these reported in our early papers (Gan *et al.*, 2004a, 2004b).

As can be seen in Fig. 4, the FE model-predicted TM (umbo) displacement curves fall well into the range of five temporal bone experimental curves across the frequencies of 200–8000 Hz. However, the model did not show the peak at 1000 Hz that was usually seen experimentally in most ears. There are some discrepancies between the model and experimental results on stapes footplate, as shown in Fig. 4B. The displacement magnitudes from modeling at high frequencies ($f > 1000$ Hz) are below the experimental curves. The phase curves of the TM and footplate vibrations from modeling are slightly lower than the experimental results. The causes of these discrepancies could be due to the simplification of the real ear geometry and the accuracy of mechanical properties of ear tissues used for the model. However, in general, the simulations from the model show patterns that are similar to the temporal bone data.

The effect of TM perforations [as shown in Fig. 2A] on sound transmission through the middle ear was investigated in temporal bones and FE model. Figures 5–7 illustrate the frequency response curves of the TM (umbo), stapes footplate, and middle ear pressure obtained from the temporal bones and the FE model, side-by-side for comparison. Panels (A) and (C) in Fig. 5 show the mean displacement curves with standard error bars of the magnitude and phase angle measured at the umbo from five bones under control, Hole 1, Hole 2, and combined Holes 1 and 2. The TM displacement curves predicted by the model are shown in panels (B) and (D). The experimental curves with perforations [Fig. 5A]

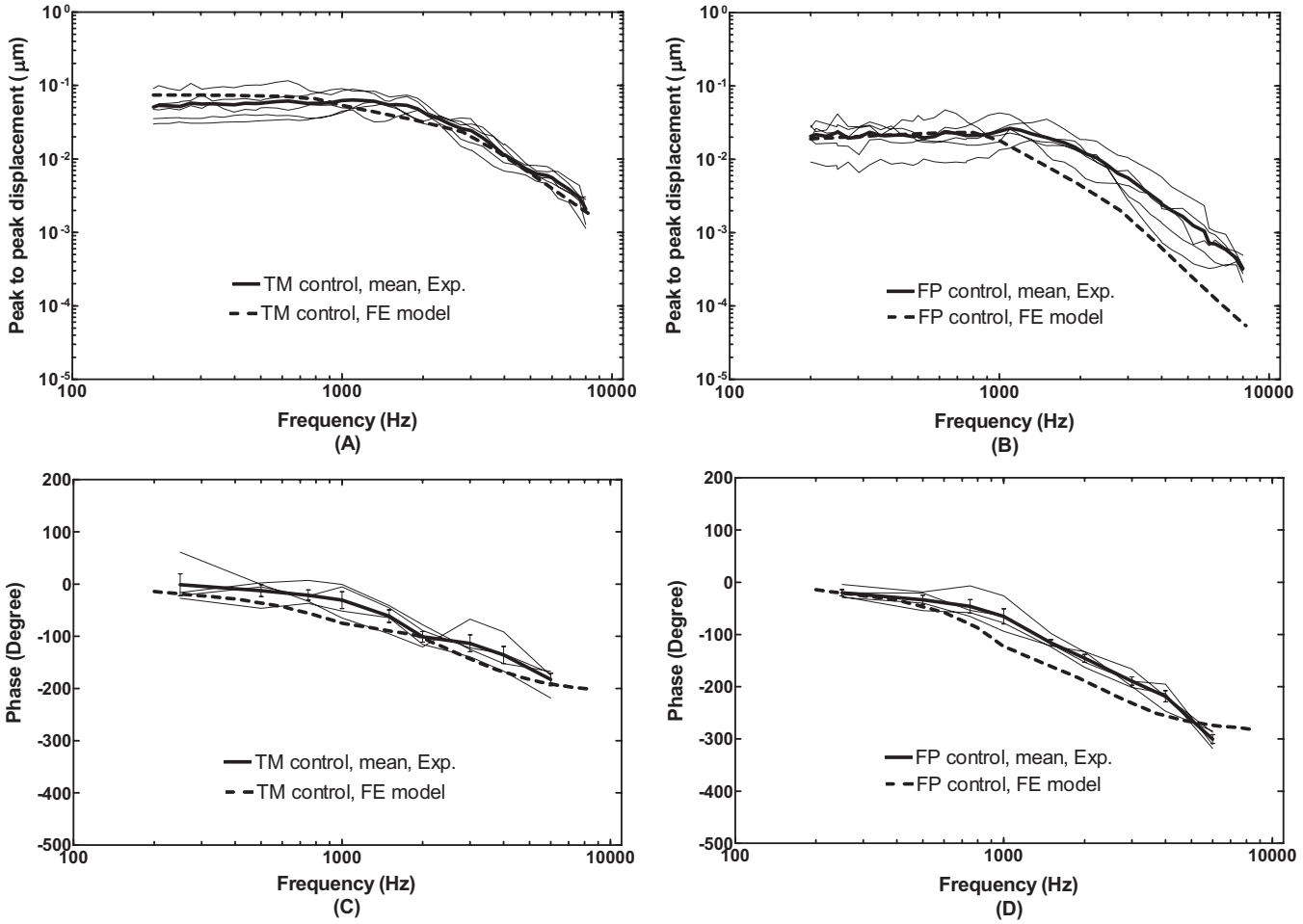


FIG. 4. Comparison of the FE model-derived TM (at the umbo) displacement and stapes footplate displacement with the measurements from five human temporal bones with intact TM's in magnitude and phase angle. The input sound pressure level was 90 dB at 2 mm from the umbo in the ear canal. The thick broken lines represent the model results. The thin solid lines represent the curves obtained from the individual temporal bones with the mean curves (thick solid lines). [(A) and (C)] TM displacement; [(B) and (D)] footplate displacement.

show a maximum reduction of 30–33 dB of the TM displacement at 200 Hz. As frequency increased from 200 to 3000 Hz, the TM displacement increased rapidly and surpassed the control curves at 1–1.5 kHz. The curves flatten at the lowest frequencies ($f < 300$ Hz) which was probably due to noise effect. The similar phenomenon was observed in the stapes footplate curves at the lowest frequencies [Fig. 6A].

The model results [Fig. 5B] show a maximum reduction of 28–31 dB of the TM displacement at 200 Hz caused by perforations and then the displacement increased slowly as frequency increased to 3 kHz. The model curves are much less steep than the experimental curves at frequency below 1 kHz. The reduced displacement values caused by perforations decreased to zero and then increased for the frequencies over 3 kHz. The TM displacement values at 3 kHz had the effect of downplaying the differences between experimental and model results. The experimental phase curves [Fig. 5C] at three perforation cases are higher than that of the control curve from 200 to 4000 Hz, which are also shown in the model phase curves. Figure 5 demonstrates that Hole 1 resulted in more reduction in the TM displacement than Hole 2, and combined Holes 1 and 2 resulted in more reduction in the TM movement than Hole 1 or Hole 2 only. This was observed from both the model and bone experiments, which demonstrated the effect of perforation size. It is noted that the difference of combined Holes 1 and 2 from a single hole, or the difference of Hole 1 from Hole 2, may be also affected by different locations. However, for small perforations (hole area less than 5% of the TM surface area), the effect of hole location on sound transmission through the middle ear may not be important as reported by Voss *et al.* (2001a) and Ahmad and Ramani (1979).

Figure 6 displays the frequency response curves of the stapes footplate displacement measured from the bones and derived from the FE model. Panels (A) and (C) of Fig. 6 show the mean magnitude and phase angle data of stapes footplate displacement with standard error bars from five bones for control, Hole 1, Hole 2, and combined Holes 1 and 2. The model-derived footplate displacement curves are displayed in panels (B) and (D). As can be seen in Fig. 6A, the footplate displacement had a maximum reduction of 23–25 dB at 200 Hz and 18–20 dB reduction at 500 Hz measured from temporal bones. The displacement increased as frequency increased to 1.5 kHz and surpassed the control curves at 1.5–3 kHz with a value less than 2 dB. Voss *et al.* (2001a) reported the reduction in the stapes velocity at frequencies below 1.5 kHz measured in temporal bones. The

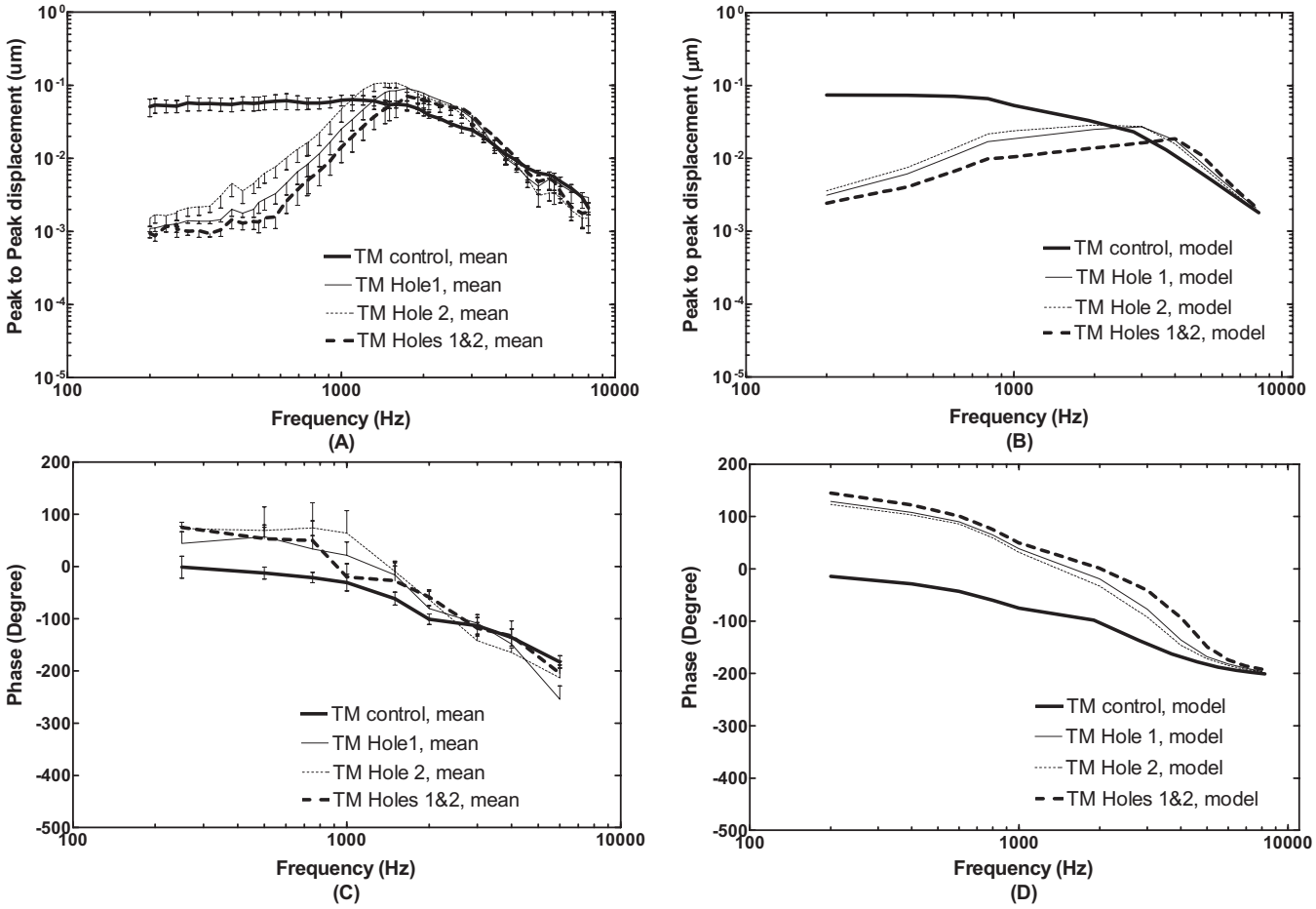


FIG. 5. Mean displacements with standard errors ($N=5$) measured at the TM (umbo) and calculated from the FE model in control or intact TM (thick solid lines), perforations Hole 1 (thin solid lines), Hole 2 (thin dashed lines), and combined Holes 1 and 2 (thick broken lines). The input sound pressure level was 90 dB at 2 mm from the umbo in the ear canal. [(A) and (C)] Bone experiments; [(B) and (D)] FE model.

maximum reduction with a perforated TM at the hole size of 1.2 mm of diameter was reached at 200 Hz over the tested frequency range, and the reduction decreased as frequency increased to 1.5 kHz (Voss *et al.*, 2001a). The results from temporal bones in this study are generally consistent with results of Voss *et al.* (2001a). However, there are some differences between this study and results of Voss *et al.* (2001a) at lowest frequencies. In the measurements of Voss *et al.*, (2001a) the stapes displacement with holes continues to decrease as frequency decreases to 200 Hz, which is not seen here. The low frequency flattening of displacement curves observed in this study may be due to the noise level effect in our experiments.

The FE model shows a maximum of 25 dB reduction in footplate displacement for three perforations at 200 Hz and 20 dB reduction at 500 Hz in Fig. 6B. The displacement increased gradually and the reduction decreased to 7–13 dB at 1 kHz. Then, the footplate displacement with holes started to meet the control curve and surpassed the control at frequencies greater than 2.5 kHz with a value less than 2 dB. Both phase data from the model and experimental measurements with TM perforations increased compared with control curves.

In Fig. 7, the FE model-predicted frequency response curves of acoustic pressure P2 in the middle ear cavity (between the round window and stapes) were compared with the experimental data obtained in temporal bones. Panels (A) and (C) display the mean pressure curves (magnitude and phase) with standard errors measured from five temporal bones for control and three perforations. Panels (B) and (D) display pressure curves derived from the FE model with control and three perforations similar to the experiments in temporal bones. Sound pressure differences across the TM, P1-P2, measured from the temporal bones in control (with intact TM) were 15, 10, 15, and 25 dB at frequencies of 200 Hz, 1 kHz, 2 kHz, and 4 kHz, respectively. The model-derived pressure difference (P1-P2) at the corresponding frequencies of 200 Hz, 1 kHz, 2 kHz, and 4 kHz was 16, 8, 7, and 14 dB, respectively.

When the TM was perforated, the pressure difference across the TM vanished at $f<1$ kHz, which was observed from the measurements in temporal bones as well as the results calculated from the model. However, there are some discrepancies between the measurements and FE model at high frequencies ($f>1$ kHz). Experimental data [Fig. 7A] show that the P2 was greater than P1 at the frequency between 1 and 2 kHz for single hole (peak around 1.4 kHz) and between 1 and 2.5 kHz for two holes (peak around 2 kHz). Then, P2 decreased to less than P1 at the frequency

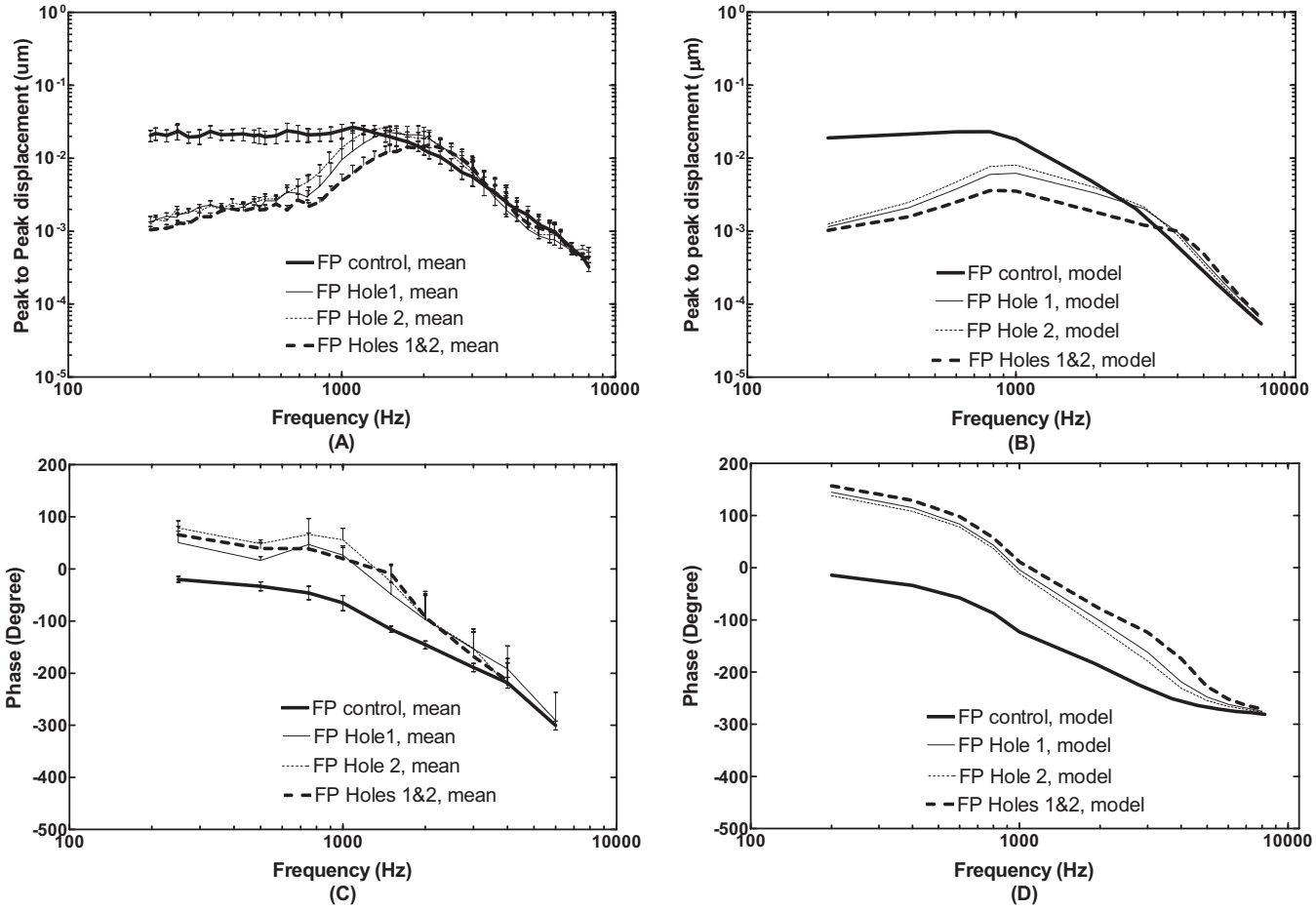


FIG. 6. Mean displacements with standard errors (N=5) measured at the stapes footplate and calculated from the FE model in control or intact TM (thick solid lines), perforations Hole 1 (thin solid lines), Hole 2 (thin dashed lines), and combined Holes 1 and 2 (thick broken lines). The input sound pressure level was 90 dB at 2 mm from the umbo in the ear canal. [(A) and (C)] Bone experiments; [(B) and (D)] FE model.

greater than 2.5 kHz with some variations as frequency increases. For model results [Fig. 7B], P2 increased over P1 at the frequency between 1 and 3 kHz for single hole and between 1 and 5.5 kHz for two holes. Compared with experimental curves, peaks of the model curves were shifted to 3 kHz for single hole and 4 kHz for two holes. To explain the discrepancies of frequency behavior of P2 between the model and measurements after TM perforation needs further study on acoustic properties of the FE model.

IV. DISCUSSION

A. FE model prediction of middle ear transfer function change by perforation

In this paper, the structural alterations of the ear (i.e., TM perforations) were visually created in our 3D model of human ear and the multi-field coupled FE analysis was conducted in the model. The TM and footplate displacements derived from the model were compared with the data measured from the temporal bones under normal and perforated conditions. The FE model predicted that the TM perforations affect the TM and stapes footplate displacements at low frequencies, which are consistent with the observations reported by Voss *et al.* (2001a, 2001b; 2007), Bigelow *et al.* (1996), and Ahmad and Ramani (1979). The value of displacement reduction was related to the perforation size, location, and multiplicity. The maximum reduction was reached at lowest frequency such as 200 Hz tested in this study. These results are consistent with our measurements in temporal bones and those reported by Voss *et al.*(2001a). The reduction in displacement decreased as the frequency increased and finally approximated zero. The high frequency end point for perforation effect was sensitive to the number of holes on the TM. For a single hole such as Hole 1 or Hole 2, the displacement curves of the TM and footplate were not shifted to a high frequency as predicted by the model and measured from temporal bones (Figs. 5 and 6). For multiple holes such as combined Hole 1 and Hole 2, the displacement curves were shifted toward the high frequency.

To further investigate the effects of perforation size and location on sound transmission through the middle ear, we created two additional perforation cases in the model which are displayed in Figs. 2B and 2C. Hole 3 in Fig. 2B has the size equal to the size of Hole 1 plus Hole 2 and is located in the posterior-inferior site, the same location as Hole 1. The TM and footplate displacement curves with two holes and with a single Hole 3 from the model are shown in Figs. 8A and 8B, respectively. As can be seen in each figure, the displacement curve obtained from the multiple holes was different from that of a single hole even though the perforation areas were equal. Multiple holes caused 1–3 dB more reduc-

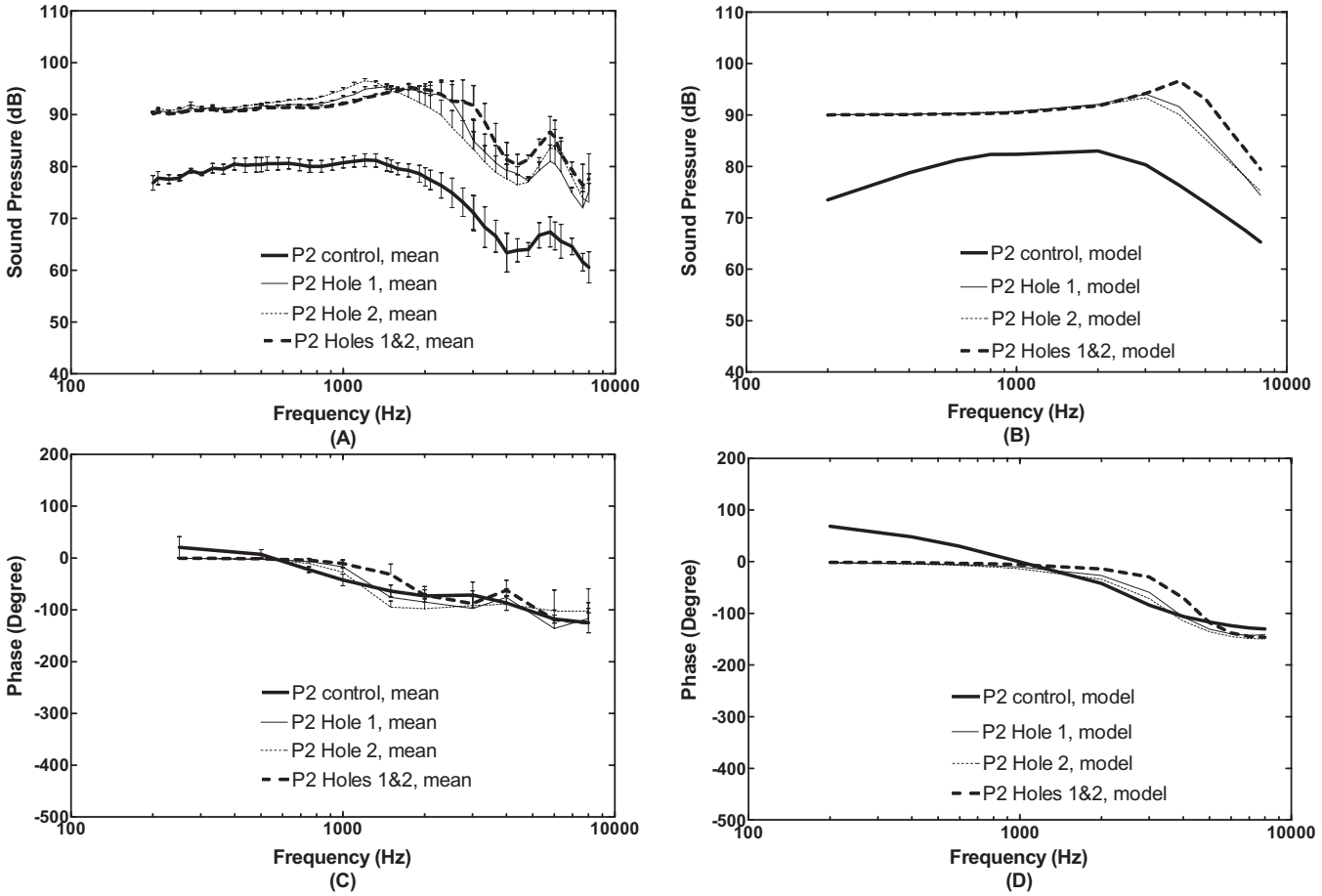


FIG. 7. Mean sound pressure level in the middle ear cavity (P2) with standard errors measured from five temporal bones and calculated from the FE model in control or intact TM (thick solid lines), perforations Hole 1 (thin solid lines), Hole 2 (thin dashed lines), and combined Holes 1 and 2 (thick broken lines). The input sound pressure P1 was 90 dB at 2 mm from the umbo in the ear canal. [(A) and (C)] Bone experiments; [(B) and (D)] FE model.

tion in the TM or footplate displacement than that caused by a single hole at frequencies below 4 kHz. This effect is mainly due to multiplicity. We did increase the size of Hole 2 to the same as Hole 1 in the FE model and calculated the TM and footplate displacements. The results (not shown here) indicated that the effect of different location on TM or stapes footplate displacement curve was limited within 1 dB.

Hole 4 in Fig. 2C has the size twice of Hole 3 and is also located in the posterior-inferior site, the same as Hole 3 and Hole 1. Comparison of the TM and footplate displacement curves obtained from three single perforations: Hole 1, Hole 3, and Hole 4 with the surface areas of 1.33, 2.15, and 4.34 mm^2, respectively, demonstrate the sole effect of perforation size on middle ear transfer function. The results are shown in Figs. 8C and 8D. As can be seen in these figures, the TM or footplate displacement decreased proportionally with the increase in hole size at frequencies below 4 kHz. The curve was not shifted to a high frequency when the hole size changed. These findings suggest that the perforation size plays an important role for reducing the TM and stapes movements or causing conductive hearing loss.

It is noted that the case of multiple perforations is worse than a single perforation even though the perforation size is same [Figs. 8A and 8B]. When there is more than one hole on the TM, the location of the hole also plays a role on changes of middle ear transfer function. The transmission of acoustic energy from the ear canal to the cochlea is affected by locations of the TM perforations as observed in multiple perforations. However, the mechanism of multiple perforations on sound wave transmission through the middle ear needs future study.

In summary, the perforation size and location are two main factors affecting perforation-induced reduction in middle ear transfer function or perforation-induced conductive hearing loss. These two factors interact with each other to influence the acoustic-mechanical transmission through the ear. For a better understanding of the effect of perforation on middle ear function, further studies such as different combinations of perforation size and location are needed. Moreover, it is noticed that the frequency response curves of the displacement magnitude and phase at the TM (umbo) and footplate with perforations from the model were not always similar to the experimental curves obtained in temporal bones (Figs. 5 and 6). Three speculated reasons could be counted for these discrepancies: first, the hole size and hole edge smoothness in FE model were not the same as those holes made in the temporal bones; second, the possible change of local material properties of the TM due to the process of creating the hole (e.g., cauterization) as well as the change of TM tension were not simulated in the model; third, the current model does not include the mastoid cavity even though the mastoid cavity was partially removed

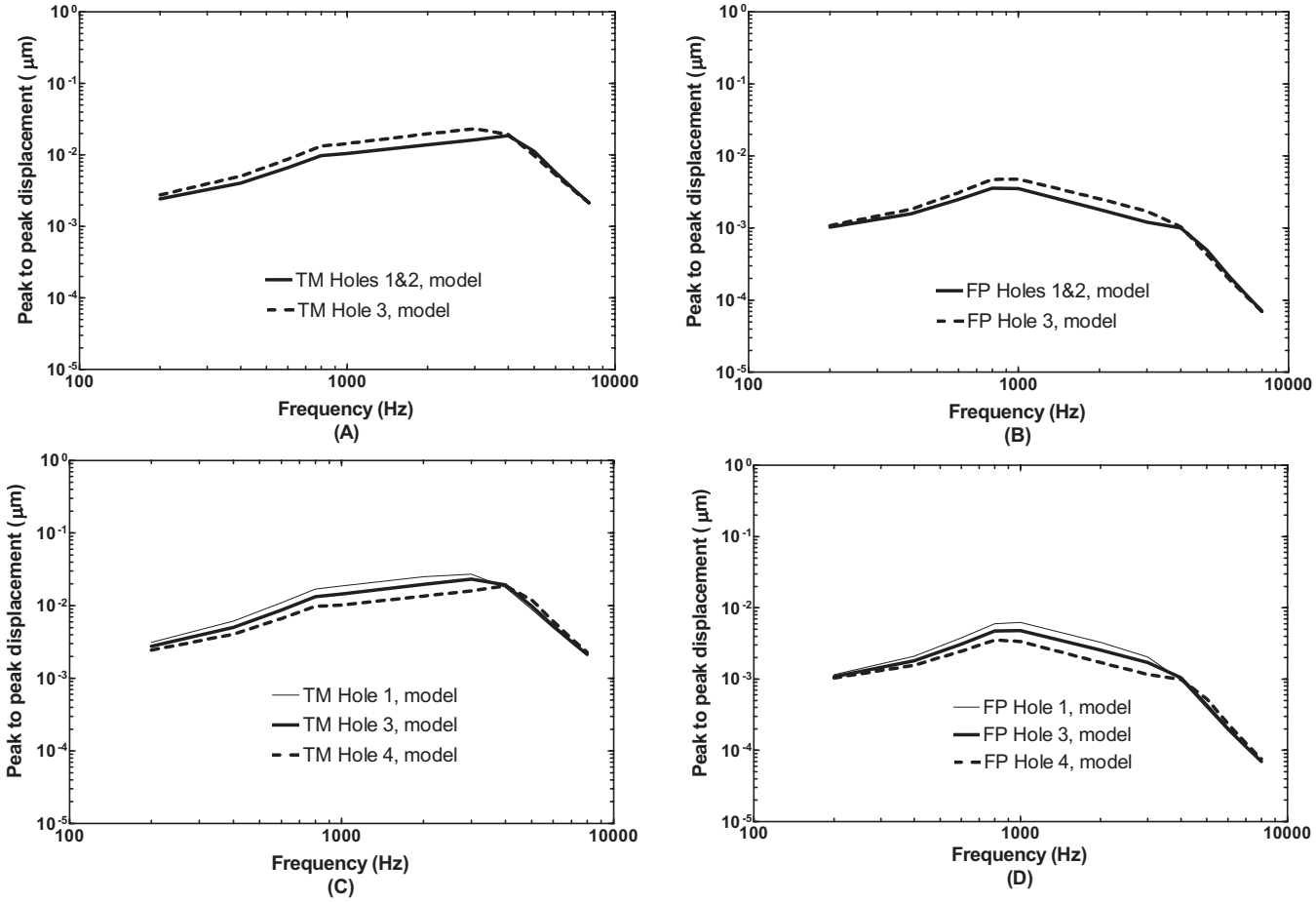


FIG. 8. Comparison of the effects of perforation size and location on TM (umbo) and footplate displacements. (A) TM displacements obtained from two holes (Hole 1 and Hole 2) and a single Hole 3 with the same perforation area. (B) Footplate displacements obtained from two holes (Hole 1 and Hole 2) and a single Hole 3. (C) TM displacement obtained from single hole with different sizes. (D) Footplate displacement obtained from single hole with different sizes.

through facial recess surgical preparation. These factors may need further consideration in our model. However, the most critical factor affecting the model accuracy is the material properties of ear tissues which should be improved by including non-linear viscoelastic parameters in frequency domain in FE analysis.

B. FE model prediction of middle ear cavity pressure transfer function

Our previous study on sound pressure distribution in middle ear cavity (Gan *et al.*, 2006) has demonstrated that there is no significant difference of acoustic pressure at different locations inside the cavity, particularly, at frequencies below 4000 Hz. In this study, we calculated the sound pressure P2 at a location between the round window and stapes in the model and monitored P2 at the same location in temporal bones. A constant sound pressure P1 of 90 dB SPL was applied at 2 mm from the umbo in the ear canal. The measurements are shown in Figs. 7A and 7C and the modeling results are in Figs. 7B and 7C . With the TM perforated, the pressure difference P1-P2 was zero at low frequencies for all perforation sizes. Voss *et al.* (2007) reported similar results for the perforation sizes they tested.

The ratio of middle ear cavity pressure P2 to the pressure at the ear canal P1 predicted by the model and measured in temporal bones from this study and from Voss *et al.* (2007) are shown in Fig. 9. The ratio P2/P1 in decibels represents the middle ear cavity pressure transfer function. Figure 9A displays that the frequency response curves of the ratio P2/P1 obtained from five bones (mean data) in this study are very close to the curve reported by Voss *et al.* (2007) on one bone at frequencies below 3 kHz with a 1.2 mm TM perforation. As frequency increased above 3 kHz, a peak at 4 kHz was observed by Voss *et al.* (2007), but not in this study. The modeling results show a peak at 3 kHz and the P2/P1 ratio difference between the FE model and measurements in temporal bones were 2–3 dB at frequencies below 3 kHz. The discrepancies of the P2/P1 ratio between the model and measurements at the frequencies of 3–6 kHz were mainly due to the peak of P2 curves. The P2 curve was shifted to a high frequency with a TM perforation shown in Fig. 7B. The phase of the ratio P2/P1 obtained from the model and experimental measurements on temporal bones are in general consistent, as shown in Fig. 9B. However, there are some discrepancies at high frequencies (f =1.5–3 kHz) and an approximate 90° difference between the model and data of Voss *et al.* (2007) is observed.

In summary, the TM is driven by pressure difference across the TM or the ratio P2/P1. The change of P2/P1 ratio in the case of a perforation is correlated to the TM and stapes

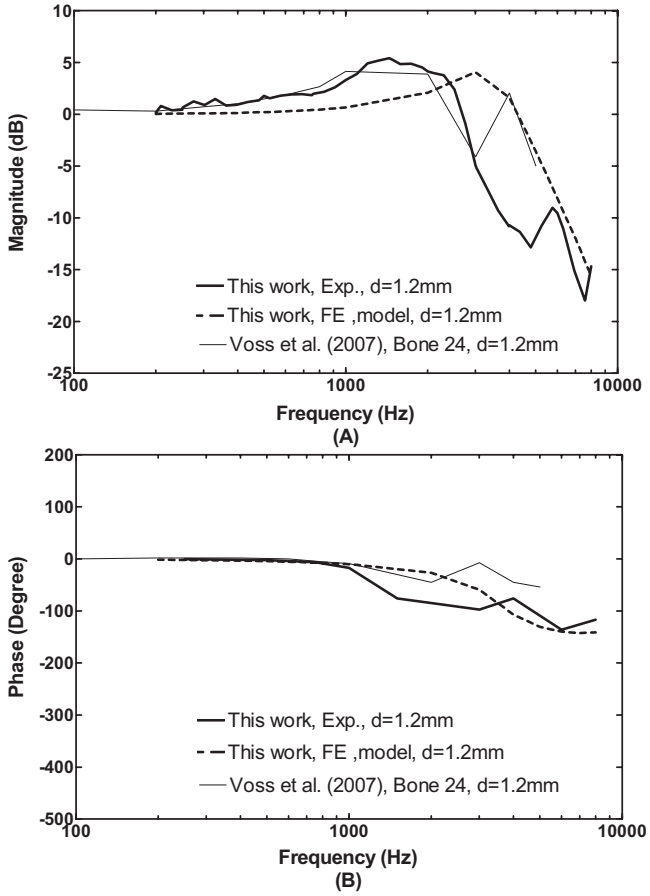


FIG. 9. The ratio of middle ear cavity pressure P2 to pressure in the ear canal P1, or middle ear cavity pressure transfer function P2/P1, measured from temporal bones and calculated from the FE model with perforations in comparison with the results reported by Voss *et al.* (2007). (A) Magnitude; (B) phase angle.

footplate movements. This correlation is shown in the experimental measurements and FE modeling results in this study. The peaks of P2 pressure and P2/P1 ratio of the measurements with Hole 1 in Figs. 7 and 9 around 1.5 kHz reflect the peaks of TM and footplate displacement measurement curves in Figs. 5 and 6. However, the peak of P2 or P2/P1 ratio curve from the model with Hole 1 is at 3 kHz in Figs. 7 and 9, which represents that the resonance of the middle ear cavity in the model with hole is different from the real temporal bone experimental setup because of the cavity volume difference. Thus, discrepancies of the TM and footplate displacements curves at high frequencies between the model and measurements observed in Figs. 5 and 6 are probably due to the resonance difference between the model and temporal bone.

V. CONCLUSION

A 3D FE model of human ear including structures of the external ear canal, middle ear, and cochlea has been used to predict the effect of TM perforations on sound transmission through the middle ear. The displacement curves of the TM (at the umbo) and the stapes footplate as well as the sound pressure inside the middle ear cavity, or the middle ear cavity pressure transfer function, were derived from the model over the frequency range of 200–8000 Hz and compared with the data measured from temporal bones. These comparisons between the model results and measurements (Figs. 4–7) are important to verify the validity of the model. We noticed that there were discrepancies of the TM, stapes footplate, and middle ear pressure P2 between the model and experimental measurements in temporal bones, particularly at higher frequencies. The dynamic behavior of the model for acoustic energy transmission through the middle ear needs to be improved over the auditory frequency range. However, this study indicates that the FE model of human ear equipped with multi-field coupled analysis capabilities can be used as a tool to investigate the relationship between middle ear structural changes and middle ear function alteration with geometric visualization.

The model predicts that the TM perforations mainly affect the middle ear transfer function at low frequencies, which is similar to the measurements from temporal bones observed in this study and reported in the literature. Using the model, we further investigated the effects of perforation size and location on TM and stapes footplate displacements. The results indicate that increase in perforation size reduces the TM and footplate movements and results in conductive hearing loss. Compared with a single perforation, the multiple perforations not only reduce the magnitude of the TM and footplate displacement but also move the displacement curve toward the high frequency. Therefore, the location of TM perforation may have some effect on hearing loss.

The effect of TM perforation on middle ear function for sound transmission has been thoroughly studied by Voss *et al.* (2001a, 2001b, 2001c) through a series of publications since 2001. They used a simple circuit model or lumped model to simulate different perforation-induced changes in stapes velocity and middle ear input impedance. The model results were compared with the measurements obtained from temporal bones (Voss *et al.*, 2001c, 2007). Compared with the circuit model used by Voss *et al.* (2001, 2001c, 2007), the 3D FE model presented in this study is much more complicated because the structural geometry, mechanical properties, and boundary conditions of the ear are all involved in the model. The TM and stapes movements depend on all parameters of the model as well as the FE analysis method. Thus, it requires a great effort to improve the model by validating it with experimental measurements in temporal bones. The present study suggests that the FE model needs to be improved through our understanding of the middle ear mechanical system in both mechanical properties of middle ear components and dynamic behavior of the system.

We also realize that TM perforations may result in a complex combination of changes in sound transmission through two routes: the mechanical route through the ossicular chain and the acoustic route through the air in the middle ear cavity. The contributions of the mechanical route and acoustic route to the change of sound conduction from the ear canal to oval window, round window, and cochlea will be related to perforation size and location. Some of our future investigations will focus on identifying the mechanism and distribution of sound transmission through these two routes in the presence of TM perforations.

ACKNOWLEDGMENTS

The authors thank Don Nakmali, BSEE, at Hough Ear Institute for his expert technical assist in the temporal bone experiments. NIH/NIDCD Grant No. R01DC006632 and NSF/CMS Grant No. 0510563 supported this work.

Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (**2001**). "Human middle-ear sound transfer function and cochlear impedance," Hear. Res. **152**, 100–109.

Ahmad, S. W., and Ramani, G. V. (**1979**). "Hearing loss in perforations of the tympanic membrane," J. Laryngol. Otol. **93**, 1091–1098.

Bigelow, D. C., Swanson, P. B., and Saunders, J. M. (**1996**). "The effect of tympanic membrane perforation size on umbo velocity in the rat," Laryngoscope **106**, 71–76.

Funnell, W. R. J., and Laszlo, C. A. (**1978**). "Modeling of the cat eardrum as a thin shell using the finite-element method," J. Acoust. Soc. Am. **63**, 1461–1467.

Gan, R. Z., and Wang, X. (**2007**). "Multi-field finite element analysis for sound transmission in otitis with effusion," J. Acoust. Soc. Am. **122**, 3527–3538.

Gan, R. Z., Feng, B., and Sun, Q. (**2004a**). "Three-dimensional finite element modeling of human ear for sound transmission," Ann. Biomed. Eng. **32**, 847–859.

Gan, R. Z., Wood, M. W., and Dormer, K. J. (**2004b**). "Human middle ear transfer function measured by double laser interferometry system," Otol. Neurotol. **25**, 423–435.

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 distributions," Med. Eng. Phys. **28**, 395–404.

Gan, R. Z., Reeves, B. P., and Wang, X. (**2007**). "Modeling of sound transmission from ear canal to cochlea," Ann. Biomed. Eng. **35**, 2180–2195.

Kelly, D. J., Prendergast, P. J., and Blayney, A. W. (**2003**). "The effect of prosthesis design on vibration of the reconstructed analysis of four prostheses," Otol. Neurotol. **24**, 11–19.

Koike, T., and Wada, H. (**2002**). "Modeling of the human middle ear using the finite-element method," J. Acoust. Soc. Am. **111**, 1306–1317.

Mehta, R. P., Rosowski, J. J., Voss, S. E., O'Neil, E., and Merchant, S. N. (**2006**). "Determinants of hearing loss in perforations of the tympanic membrane," Otol. Neurotol. **27**, 136–143.

Merchant, S. N., Ravicz, M. E., Rosowski, J. J. (**1996**). "Acoustic input impedance of the stapes and cochlea in human temporal bones," Hear. Res., **97**, 30–45

Santa Maria, P. L., Atlas, M. D., and Ghassemifar, R. (**2007**). "Chronic tympanic membrane perforation: a better animal model is needed," Wound Repair Regen **15**, 450–458.

Sun, Q., Gan, R. Z., Chang, K. H., and Dormer, K. J. (**2002**). "Computer-integrated finite element modeling of human middle ear," Biomech. Model. Mechanobiol. **1**, 109–122.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001a**). "Middle-ear function with tympanic-membrane perforations. I. Measurements and mechanisms," J. Acoust. Soc. Am. **110**, 1432–1444.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001b**). "How do tympanic-membrane perforations affect human middle-ear sound transmission?" Acta Oto-Laryngol. **121**, 169–173.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001c**). "Middle-ear function with tympanic-membrane perforations. II. A simple model," J. Acoust. Soc. Am. **110**, 1445–1452.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2007**). "Non-ossicular signal transmission in human middle ears: Experimental assessment of the 'acoustic route' with perforated tympanic membranes," J. Acoust. Soc. Am. **122**, 2135–2153.

Wada, H., and Metoki, T. (**1992**). "Analysis of dynamic behavior of human middle ear using a finite method," J. Acoust. Soc. Am. **92**, 3157–3168.

Zwislocki, J. (**1962**). "Analysis of the middle ear function. Part I. Input impedance," J. Acoust. Soc. Am. **34**, 1514–1523.

Optimal electrode selection for multi-channel electroencephalogram based detection of auditory steady-state responses

Bram Van Dun[a] and Jan Wouters
ExpORL, Katholieke Universiteit Leuven, Herestraat 49/721, B-3000 Leuven, Belgium

Marc Moonen
ESAT-SCD/SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

(Received 15 October 2008; revised 21 April 2009; accepted 23 April 2009)

Auditory steady-state responses (ASSRs) are used for hearing threshold estimation at audiometric frequencies. Hearing impaired newborns, in particular, benefit from this technique as it allows for a more precise diagnosis than traditional techniques, and a hearing aid can be better fitted at an early age. However, measurement duration of current single-channel techniques is still too long for clinical widespread use. This paper evaluates the practical performance of a multi-channel electroencephalogram (EEG) processing strategy based on a detection theory approach. A minimum electrode set is determined for ASSRs with frequencies between 80 and 110 Hz using eight-channel EEG measurements of ten normal-hearing adults. This set provides a near-optimal hearing threshold estimate for all subjects and improves response detection significantly for EEG data with numerous artifacts. Multi-channel processing does not significantly improve response detection for EEG data with few artifacts. In this case, best response detection is obtained when noise-weighted averaging is applied on single-channel data. The same test setup (eight channels, ten normal-hearing subjects) is also used to determine a minimum electrode setup for 10-Hz ASSRs. This configuration allows to record near-optimal signal-to-noise ratios for 80% of subjects.
 [DOI: 10.1121/1.3133872]

PACS number(s): 43.64.Ri, 43.66.Yw, 43.66.Sr [BLM] Pages: 254–268

I. INTRODUCTION

Universal detection of infants with hearing loss is recommended before the age of 3 months (Joint Committee on Infant Hearing, 1995). As the number of hearing screened newborns grows worldwide, the need for objective audiometric techniques has greatly increased over the past decade. An appropriate clinical response to the need for rehabilitation of hearing problems is necessary. An effective treatment may consist of the use of a hearing aid. However, fitting of this device requires a frequency specific estimation of the newborn's hearing thresholds. These estimates can be provided using auditory steady-state response (ASSR) detection long before first behavioral thresholds can be obtained. Unfortunately, a reliable ASSR-based hearing threshold estimation procedure takes 45 min with adults and can last several hours with newborns (Luts *et al.*, 2006). The measured signals are often corrupted by artifacts due to the newborn's movements during measurement. The general aim is thus to increase robustness against artifacts and to reduce measurement time. Before actual studies can be carried out on infants, the proposed techniques are evaluated on adults. Behavioral hearing thresholds can easily be determined with adults and hence a reference is obtained for application in the infant population.

The ASSR is an electrophysiological response of the brain evoked by acoustic stimuli. These stimuli are generally based on amplitude modulated (AM) and/or frequency modulated (FM) sinusoidal carriers. The carriers stimulate specific parts of the cochlea, while the modulations activate certain parts of the auditory system. The appearance of the modulation frequency in the electroencephalogram (EEG) can be used as a label for the carrier frequency. The first significant ASSR study was conducted by Galambos *et al.* (1981) with modulation frequencies of 40 Hz. Subsequently, extensive research has shown that the ASSR can indeed provide an objective and frequency specific way to determine reliable hearing thresholds, both applicable to adults and infants [we refer to Picton *et al.* (2003) for an extensive overview].

Many research studies have focused on optimizing detection of ASSRs. First, different stimulus types have been investigated. AM and/or FM stimuli are most widely accepted (John *et al.*, 2001b). Other stimuli are stimuli with exponential modulation envelopes (John *et al.*, 2002), broadband and band-limited noise (John *et al.*, 2003), stimuli composed of several carriers modulated with the same modulation frequency (Stürzebecher *et al.*, 2001), and complex stimuli with broader frequency spectra (Riquelme *et al.*, 2006; Stürzebecher *et al.*, 2006). In general, responses are more difficult to detect when stimuli become more frequency specific. Second, normal averaging has been compared with weighted averaging and artifact rejection (John *et al.*, 2001a). Third, the possibility to increase ASSR detection ef-

[a]Also at ESAT-SCD/SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. Author to whom correspondence should be addressed. Electronic mail: bram.vandun@med.kuleuven.be

ficiency by means of multi-channel EEG recordings has been evaluated (van der Reijden *et al.*, 2004; Van Dun *et al.*, 2007a, 2007b). Last, several statistical decision criteria have been studied: tests that use spectral phase information, spectral amplitude information, or both; tests that evaluate only the first harmonic of the response (Cebulla *et al.*, 2001; Dobie and Wilson, 1996) and recently, tests that include higher harmonics for response detection (Cebulla *et al.*, 2006).

Most studies are based on EEG data obtained from subjects who are instructed to relax or sleep. In practice, however, and especially with newborns, numerous artifacts prevent a successful measurement unless one takes appropriate measures against these artifacts. This can be achieved using artifact rejection or weighted averaging (John *et al.*, 2001a). Data blocks severely corrupted with artifacts are entirely discarded or are not allowed to contribute much to the final result. The disadvantage of these techniques is measurement time. Therefore, other techniques have been researched that are more robust against artifacts. These techniques should allow detection of significantly more responses in the presence of artifactual energy in the EEG without extending measurement time.

A well-chosen electrode derivation makes response registration easier. The Cz-Oz or Cz-neck derivations are the most commonly used derivations for adults because mean ASSR thresholds are smaller for these derivations (John *et al.*, 2001a; Lins and Picton, 1995; Luts and Wouters, 2005). This is confirmed for Cz-Oz derivations by van der Reijden *et al.* (2004).

A multi-channel setup is used to search for derivations offering the largest signal-to-noise ratio (SNR) values and thus the smallest ASSR thresholds for adults. This Cz-Oz derivation does not always guarantee the largest SNR and sometimes other derivations (for example, Cz-mastoid, Cz-Pz, and Cz-neck) offer larger SNRs. This is supported by observations that ASSR thresholds (and their SNRs) show large variations across subjects (John *et al.*, 2001a; Lins and Picton, 1995; Luts and Wouters, 2005). However, the derivation with the largest SNR is subject-dependent and cannot be predicted beforehand. Hence, EEG data recorded using multiple derivations simultaneously, combined with appropriate multi-channel processing, are required to select the "best" channel (or combination of channels) for as many subjects as possible.

The first part of this introduction focuses on hearing threshold estimation using ASSRs with modulation frequencies between 80 and 110 Hz. These responses are generated mainly in the brainstem. The problem statement is not restricted to this frequency range however. It would be interesting to focus on multi-channel EEG data containing ASSRs with lower modulation frequencies, mainly generated in the auditory cortex. The amplitude of these responses varies more with subject, modulation frequency, and state of arousal. Moreover, generally more noise is obscuring the response in lower-frequency regions due to other brain processes. ASSRs to lower frequencies may be more closely related to actual processes of speech understanding. Speech is a complex acoustic waveform containing different spectral and temporal cues for intelligibility. Lower modulation frequencies consisting of envelope information (with modulation frequencies below 20 Hz) appear critical for speech intelligibility (Drullman *et al.*, 1994a, 1994b; Saberi and Perrott, 1999; Shannon *et al.*, 1995). Earlier publications mention ASSRs at frequencies between 2 and 40 Hz [for example, Cohen *et al.* (1991); Galambos *et al.* (1981); Picton *et al.* (1987)]. Recent studies introduce envelope-following responses to slow rates using AM noise (Purcell *et al.*, 2004) and spoken sentences (Aiken and Picton, 2008). Alaerts *et al.* (2009) linked cortical ASSRs at low frequencies to behavioral outcomes of speech processing. ASSRs are significantly correlated with speech reception thresholds for phonemes and sentences in noise, independent of intensity level. Trends of increased responses around 10 and/or 20 Hz are found. Therefore, it is decided in this study to use 10 Hz as a representative for cortical ASSRs.

This paper evaluates practical performance of a multi-channel processing strategy based on a detection theory approach when applied to adults. It is based on eight-channel EEG measurements of ten normal-hearing adults. The results serve as an intermediate step for application of multi-channel processing to infants. First, the EEG derivation with best estimation of behavioral hearing threshold for ASSRs originating mainly from the brainstem is determined (modulation frequencies between 80 and 110 Hz in this paper). For ASSRs originating mainly from the auditory cortex (a modulation frequency of 10 Hz in this paper), the EEG derivation providing largest SNR is determined. Then, the minimum EEG-channel combination is determined for brainstem ASSRs with a near-optimal estimation of behavioral hearing thresholds for as many subjects as possible. As a practical application of this result, robustness of this minimum set to artifacts is investigated. Similarly, for auditory cortex ASSRs, the minimum EEG-channel combination is determined that provides near-optimal SNRs for a maximum number of subjects.

II. THEORETICAL BACKGROUND

ASSRs are difficult to detect due to noise originating from sources both inside and outside the skull. As a result, clinical measurements can last an unacceptably long period of time. Specifically designed signal processing techniques may reduce this measurement time. These techniques may increase the robustness against energetic artifacts that could suddenly emerge. These artifacts appear frequently in ASSR measurements.

This section describes a multi-channel signal processing technique appropriate for ASSR detection. Its rationale is based on a concept from detection theory, the *sufficient statistic*. This sufficient statistic can exploit spatio-temporal information present in the EEG measurements. A more extensive description can be found in Van Dun *et al.* (2009).

A. ASSR data model

A general ASSR data model can be given as

$$\boldsymbol{Z} = \alpha \boldsymbol{S}\boldsymbol{D}^T + \boldsymbol{N}, \tag{1}$$

with $\boldsymbol{Z}^{n\times m}$ an observation matrix, α a scalar representing the ASSR source amplitude proportional to the applied stimulus intensity, $\boldsymbol{S}^{n\times 2}$ a desired signal matrix with columns representing the ASSR (sinusoid and cosinusoid, oscillating at the known modulation frequency), $\boldsymbol{D}^{m\times 2}$ a steering matrix, $\boldsymbol{N}^{n\times m}$ an additive noise matrix, n the number of data points, and m the number of measured EEG-channels.

In Eq. (1), given only $\boldsymbol{Z}$ and $\boldsymbol{S}$, $\alpha\boldsymbol{D}$ and $\boldsymbol{N}$ can be estimated using a QR-factorization:

$$\overbrace{[\boldsymbol{s}_1^{n\times 1} \quad \boldsymbol{s}_s^{n\times 1}}^{\boldsymbol{S}} \quad \boldsymbol{Z}^{n\times m}] = \boldsymbol{Q}^{n\times(m+2)}\boldsymbol{R}^{(m+2)\times(m+2)}, \tag{2}$$

with

$$\boldsymbol{Q} = [\boldsymbol{s}_1 \quad \boldsymbol{s}_2 \quad \boldsymbol{Q}_*^{n\times m}], \quad \boldsymbol{R} = \begin{bmatrix} 1 & 0 & \hat{\boldsymbol{d}}_1^T \\ 0 & 1 & \hat{\boldsymbol{d}}_2^T \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_*^{m\times m} \end{bmatrix}. \tag{3}$$

$\boldsymbol{Z}$ can then be written as

$$\boldsymbol{Z} = \overbrace{[\boldsymbol{s}_1 \quad \boldsymbol{s}_2]}^{\boldsymbol{S}} \overbrace{\begin{bmatrix} \hat{\boldsymbol{d}}_1^T \\ \hat{\boldsymbol{d}}_2^T \end{bmatrix}}^{\hat{\boldsymbol{D}}^T} + \overbrace{\boldsymbol{Q}_*\boldsymbol{R}_*}^{\hat{\boldsymbol{N}}}, \tag{4}$$

where $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{N}}$ are the estimates provided for $\alpha\boldsymbol{D}$ and $\boldsymbol{N}$.

A physical interpretation is given here. The ASSR generator has an unknown source amplitude α, which depends on stimulus level. After propagation through the skull, the distribution of the recorded ASSR on the electrodes can be described by steering matrix $\boldsymbol{D}$. This steering matrix $\boldsymbol{D}$, unlike the additive EEG noise, is usually assumed to be stationary as it is only a representation of source position, directivity pattern, electrode positions, and propagation attenuation from source to electrode. Physically, no measurable delay occurs between ASSR source and electrodes (order of nanoseconds). Therefore the ASSR delay difference and hence the ASSR phase difference between electrodes are assumed to be zero. On the other hand, the delay between ASSR stimulus and response is measurable (order of milliseconds), but unknown to the observer as it depends on several physical parameters of the subject. The ASSR phase φ at the electrodes is thus unknown but equal in all channels.

Equation (1) can then be rewritten as

$$\boldsymbol{Z} = \alpha\boldsymbol{S}\overbrace{\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}}^{\boldsymbol{D}^T}\boldsymbol{d}^T + \boldsymbol{N} \tag{5}$$

$$= \alpha\boldsymbol{s}\boldsymbol{d}^T + \boldsymbol{N}, \tag{6}$$

where now $\boldsymbol{s}$ and $\boldsymbol{d}$ are vectors, and φ corresponds to the ASSR phase.

The phase φ can be estimated based on Eq. (4),

$$\hat{\varphi} = \arg\left\{ \max_{\varphi} [\cos\varphi \ \sin\varphi]\hat{\boldsymbol{D}}^T(\hat{\boldsymbol{N}}^T\hat{\boldsymbol{N}})^{-1}\hat{\boldsymbol{D}}\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix} \right\}. \tag{7}$$

The estimated phase $\hat{\varphi}$ expresses the direction that maximizes the projection of the observation matrix $\boldsymbol{Z}$ onto the desired signal $\boldsymbol{S}[\cos\varphi \ \sin\varphi]^T$ as $\hat{\boldsymbol{D}} = \boldsymbol{Z}^T\boldsymbol{S}$. Based on $\hat{\varphi}$, a rotation transformation is applied to $\hat{\boldsymbol{D}}$,

$$\hat{\boldsymbol{d}} = \hat{\boldsymbol{D}}\begin{bmatrix} \cos\hat{\varphi} \\ \sin\hat{\varphi} \end{bmatrix}, \tag{8}$$

such that the estimated steering matrix $\hat{\boldsymbol{D}}$ is transformed into an estimated steering vector $\hat{\boldsymbol{d}}$.

A spatio-temporal noise covariance matrix $\boldsymbol{K}^{mn\times mn}$ can be defined as

$$\boldsymbol{K} = \mathcal{E}\{\boldsymbol{n}\boldsymbol{n}^T\}, \tag{9}$$

with $\boldsymbol{n} = \mathrm{vec}(\boldsymbol{N})$. Here $\mathcal{E}\{\cdot\}$ is the expected value operator and the vec$(\cdot)$ operator stacks the columns of a matrix $\boldsymbol{X}$ into one column vector $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$.

If spatial and temporal correlations are separable, as will be observed here, the spatio-temporal noise covariance matrix $\boldsymbol{K}$ can be written as (Johnson and Dudgeon, 1993)

$$\boldsymbol{K} = \boldsymbol{K}_{\mathrm{spat}} \otimes \boldsymbol{K}_{\mathrm{temp}}, \tag{10}$$

where $\otimes$ represents the Kronecker product, with a spatial noise covariance matrix $\boldsymbol{K}_{\mathrm{spat}}^{m\times m}$, representing noise correlation across the different channels,

$$\boldsymbol{K}_{\mathrm{spat}} = \mathcal{E}\{\boldsymbol{N}^T\boldsymbol{N}\}, \tag{11}$$

and a temporal noise covariance matrix $\boldsymbol{K}_{\mathrm{temp}}^{n\times n}$, representing correlation of noise samples within the same channel,

$$\boldsymbol{K}_{\mathrm{temp}} = \mathcal{E}\{\boldsymbol{N}\boldsymbol{N}^T\}. \tag{12}$$

Based on Eq. (4) the noise covariance matrices can be estimated as

$$\hat{\boldsymbol{K}}_{\mathrm{spat}} = \hat{\boldsymbol{N}}^T\hat{\boldsymbol{N}}, \quad \hat{\boldsymbol{K}}_{\mathrm{temp}} = \hat{\boldsymbol{N}}\hat{\boldsymbol{N}}^T. \tag{13}$$

B. Detection theory: Detecting signals with unknown amplitude in noise

Detection theory is a means to quantify the ability to discern between signal and noise (Green and Swets, 1966). Here, a sufficient statistic $\Upsilon(z)$ is defined as

$$\Upsilon(z) = z^T\boldsymbol{K}^{-1}\tilde{\boldsymbol{s}}, \tag{14}$$

with a target signal $\tilde{\boldsymbol{s}}$ having an unknown amplitude α, and $\boldsymbol{K}$ the covariance matrix of the noise $\boldsymbol{n}$ that corrupts the observation vector z,

$$z = \alpha\tilde{\boldsymbol{s}} + \boldsymbol{n}. \tag{15}$$

A sufficient statistic theoretically captures best the amount of useful target signal $\tilde{\boldsymbol{s}}$ in the observation vector z. A decision threshold η for $\Upsilon(z)$ can be determined by specifying a false-alarm probability. This threshold η decides whether the observation vector z contains the target signal $\tilde{\boldsymbol{s}}$ or not.

C. A detection theory framework for ASSR processing

The ASSR signal model of Eq. (6) can be reformulated as

$$z = \alpha\overbrace{\boldsymbol{d}\otimes\boldsymbol{s}}^{\tilde{\boldsymbol{s}}} + \boldsymbol{n}, \tag{16}$$

with $z=\mathrm{vec}(\boldsymbol{Z})$ and $\boldsymbol{n}=\mathrm{vec}(\boldsymbol{N})$, respectively.

The sufficient statistic [Eq. (14)] can be applied here by identifying Eq. (15) with Eq. (16),

$$\Upsilon_A(z)=z^T\boldsymbol{K}^{-1}\tilde{\boldsymbol{s}}. \tag{17}$$

Replacing $\boldsymbol{K}$ by using Eq. (10), and $\tilde{\boldsymbol{s}}$ by using Eq. (16), Eq. (17) becomes

$$\Upsilon_A(z)=z^T(\boldsymbol{K}_{\mathrm{spat}}\otimes\boldsymbol{K}_{\mathrm{temp}})^{-1}(\boldsymbol{d}\otimes\boldsymbol{s}) \tag{18}$$

$$=\boldsymbol{s}^T\boldsymbol{K}_{\mathrm{temp}}^{-T}\boldsymbol{Z}\boldsymbol{K}_{\mathrm{spat}}^{-1}\boldsymbol{d}. \tag{19}$$

Substituting the estimates based on Eqs. (2)–(4), (7), (8), and (13) leads to a sufficient statistic $\hat{\Upsilon}_A(z)$ suitable for ASSR detection,

$$\hat{\Upsilon}_A(z)=[\cos\hat{\varphi}\ \sin\hat{\varphi}]\boldsymbol{S}^T\hat{\boldsymbol{K}}_{\mathrm{temp}}^{-T}\boldsymbol{Z}\hat{\boldsymbol{K}}_{\mathrm{spat}}^{-1}\hat{\boldsymbol{D}}\begin{bmatrix}\cos\hat{\varphi}\\ \sin\hat{\varphi}\end{bmatrix}. \tag{20}$$

When the EEG (noise) is stationary over all channels throughout the measurement, the covariance matrix $\boldsymbol{K}$ is constant too. When noise characteristics change over time, $\boldsymbol{K}$ cannot be considered constant anymore and Eq. (20) needs to be modified. If noise in $\boldsymbol{Z}$ is stationary only in blocks of T samples (rows) and uncorrelated between such blocks, then $\Upsilon_A(z)$ can be calculated as the sum of the sufficient statistics $\Upsilon_{A,i}(z_i)$ for the individual blocks $\boldsymbol{Z}_i$, i.e.,

$$\Upsilon_A(z)=\sum_{i=1}^{n/T}\boldsymbol{s}_i^T\boldsymbol{K}_{\mathrm{temp},i}^{-T}\boldsymbol{Z}_i\boldsymbol{K}_{\mathrm{spat},i}^{-1}\boldsymbol{d} \tag{21}$$

$$=\sum_{i=1}^{n/T}\Upsilon_{A,i}(z_i). \tag{22}$$

For each $\Upsilon_{A,i}(z_i)$ an approximation as in Eq. (20) can be used.

D. Detection using fast Fourier transformation

In many ASSR studies an objective procedure is developed where a fast Fourier transformation (FFT) analysis is carried out prior to response detection (John and Picton, 2000; Lins and Picton, 1995; Luts *et al.*, 2006; Valdes *et al.*, 1997). Detection is based on the ratio between response power P_r at the modulation frequency and mean noise power σ^2 in M neighboring frequency bins at each side,

$$\frac{P_r}{\sigma^2}=\frac{a_{\mathrm{response}}^2}{\frac{1}{M}\sum_{p=1}^{M}a_{\mathrm{noise},p}^2}, \tag{23}$$

with a_{response} the amplitude at the modulation frequency and $a_{\mathrm{noise},p}$ the noise amplitude in the pth adjacent frequency bin (John and Picton, 2000).

For brainstem responses (with modulation frequencies between 80 and 110 Hz), M is taken equal to 120 (approximately 3.7 Hz on each side). For auditory cortex responses (10 Hz), M is taken equal to 60 (approximately 1.85 Hz on each side).

III. METHODS

Two studies are described here: one based on ASSRs mainly evoked in the brainstem (modulation frequencies between 80 and 110 Hz) and one on ASSRs mainly evoked in the auditory cortex (10 Hz).

The 80–110-Hz range is the frequency region of interest for hearing threshold determination (Picton *et al.*, 2003). Several studies focus on the sources of these ASSRs. Herdman *et al.* (2002) suggested that 88-Hz ASSRs are mainly generated in the brainstem. Kuwada *et al.* (2002) indicated that ASSRs with modulation frequencies below 80 Hz mainly originate in the auditory cortex. Above 80 Hz, there appear to be at least two generators that are likely subcortical. This is confirmed by Purcell *et al.* (2004) who used response latencies to derive source locations. They concluded that for modulation frequencies above 75 Hz, most of the ASSR is generated in the brainstem. Above 95 Hz, the source lies entirely in the brainstem. As a result, in this paper responses to stimuli modulated with frequencies between 80 and 110 Hz are referred to as "responses mainly evoked in the brainstem."

The modulation frequency of 10 Hz is chosen because of its relevance for speech envelope modulations and speech perception (Drullman *et al.*, 1994a, 1994b; Saberi and Perrott, 1999; Shannon *et al.*, 1995). When lowering the modulation frequency of the stimulus, the main source of the ASSR shifts to the auditory cortex. According to Herdman *et al.* (2002), the main source can be found in the auditory cortex, but a source originating from the brainstem may still be present. This is supported by Purcell *et al.* (2004), showing that the brainstem still contributes at lower modulation frequencies, but the main source of the low-frequency ASSR is located in the auditory cortex. Responses to 10 Hz-modulated stimuli are in this paper referred to as "responses mainly evoked in the auditory cortex."

For both studies, 10 normal-hearing subjects (8 male, 2 female) with mean age 28.2 years (range 22–32 years) have been selected. Their hearing thresholds do not exceed 20 dB hearing level (HL) on the octave audiometric frequencies. All experiments have been carried out in a double-walled soundproof room with Faraday cage. Subjects have been instructed to lie down on a bed and relax or sleep for the brainstem study in Sec. III A and to stay awake while watching a movie for the auditory cortex study in Sec. III B. Lights have been switched off. This procedure has been repeated a number of weeks later to collect both test and retest data.

Kendall jelly snap electrodes have been placed on the positions described in Table I adhering to the international 10-20 system (Malmivuo and Plonsey, 1995). This configuration has been chosen similar to van der Reijden *et al.* (2004) with some extra channels added to ensure a symmetrical configuration of all electrodes. They have been placed on the subject's scalp after the skin has been abraded with Nuprep abrasive skin prepping gel. A conductive paste has been used to keep electrodes in place and to avoid that inter-electrode impedances exceed 5 kΩ at 30 Hz. Electrodes have been connected to a low-noise eight-channel Jaeger–Toennies amplifier. Each EEG-channel has been am-

TABLE I. Recording electrode positions for an eight-channel setup according to the international 10–20 system (Malmivuo and Plonsey, 1995). All channels are referenced to the reference electrode on top of the head (Cz). A graphical representation is shown in Fig. 2. The term "side" indicates on which hemisphere of the head the electrode is placed in the case it is not positioned on the middle line from front to back. In general, the sides can be "left" or "right." However, for ASSRs from the auditory cortex, being described in Sec. V (with stimuli referenced to the right ear), one refers to the sides as "ipsilateral" and "contralateral." The ipsilateral side is the side of stimulation.

Channel	Position	Side
1	Occiput (Oz)	
2	P4	R/I
3	P3	L/C
4	Right mastoid (M2)	R/I
5	Left mastoid (M1)	L/C
6	F4	R/I
7	F3	L/C
8	Forehead (Fpz)	
Reference	Vertex (Cz)	
Ground	Right clavicle	

plified ($\times 50\,000$) and bandpass filtered between 70 and 170 Hz (6 dB/octave) for the brainstem study in Sec. III A and between 1 and 30 Hz (6 dB/octave) for the auditory cortex study in Sec. III B. The sampling rate initially has been set to 1000 Hz, and afterward to 250 Hz using the *resample* function from MATLAB. The "accuracy" option has been set to $N=100$, which corresponds to an anti-aliasing (lowpass) finite impulse response filter with 800 taps. The MATLAB function compensates for filter delay. No artifact rejection was applied during testing, but a threshold has been determined offline that rejects around 10% of the recorded data blocks ("epochs") that exceed this artifact rejection threshold. All separate acoustic stimuli have been calibrated at 70-dB sound pressure level (SPL), using a Brüel & Kjær Sound Level Meter 2260 in combination with a 2-cc coupler DB0138. All stimuli have been presented to the subject, and amplified EEG signals have been recorded using the SOMA program from Van Dun *et al.* (2008) and an RME multichannel soundcard (RME, Germany).

It is important to note that all possible measures have been taken to avoid false responses not originating from statistical noise effects (the relative number of false responses originating from statistical noise is allowed to be 5%). Measures to avoid additional false responses include appropriate shielding of stimulus devices and electrodes, filtering prior to downsampling, and adequate separation of stimulus and electrode cables. Absence of false responses has been regularly checked by applying the stimuli to a simulated "deaf" subject (by blocking the ear canal of the subject using ear plugs).

A. Brainstem stimulation

Two combined stimuli with four 100% AM and 20% FM carrier frequencies each have been applied to each ear. The carrier frequencies are the same for both ears, namely, 0.5, 1, 2, and 4 kHz. Modulation frequencies have been chosen in the vicinity of 82, 90, 98, and 106 Hz for the left ear and 86, 94, 102, and 110 Hz for the right ear. These modulation frequencies have been adjusted to ensure that a non-fractional number of modulation cycles fits into one data block (epoch) of 1.024 s. For example, 82 Hz is converted to round(82×1.024)/1.024 Hz (John and Picton, 2000). Throughout the rest of this paper, the non-adjusted frequencies are referred to for conciseness.

For Secs. IV A and IV B, stimuli have been applied at an initial intensity of 60-dB SPL and lowered by 10 dB after a trial of approximately 10 min per intensity (36 sweeps, each sweep lasting 16.384 s). 10-dB SPL is the lowest intensity that has been applied. After EEG data collection, each separate channel, or each combination of channels, has been reduced to 32 sweeps, using an artifact rejection threshold that removes exactly 4 sweeps per channel (or per combination of channels). Artifact rejection for multi-channel data (i.e., a combination of channels) implies the removal of all simultaneous epochs over different channels to preserve correlation over simultaneous channels.

For the analysis of robustness against artifacts in Sec. IV C, EEG data from the previous paragraph are combined with extra measurements that emphasize occurrence of artifacts. Subjects have been instructed to be seated on a chair and to carry out a repeated series of movements in cycles of approximately 6 s. Meanwhile, the two acoustic stimuli described above have been applied at an intensity of 30-dB SPL. Measurements are 32 sweeps long. Movements have been carried out in the following order: turn the head up, down, left, and right. This procedure serves as a controlled generator of artifacts on all channels due to muscle activity and electrode cable movement. No artifact rejection has been applied as artifact rejection applied to single-channel artifact-rich EEG data would result in a significant reduction in withheld data and detected responses. When applied to multi-channel data, the number of withheld sweeps even would reduce to close to zero. A solution for excessive data rejection could be a less strict rejection threshold. This would negatively influence the detection performance for EEG data with few artifacts however. Also, the EEG quality is not known beforehand, which makes the choice of a fixed detection threshold difficult.

B. Auditory cortex stimulation

Contrary to brainstem stimulation, where multiple stimuli are applied simultaneously (Dimitrijevic *et al.*, 2002; Herdman and Stapells, 2001; Luts and Wouters, 2005), the use of such combined stimuli with modulation frequencies below 40 Hz has not been thoroughly documented yet (Armstrong and Stapells, 2007). Therefore, only one carrier frequency is modulated (instead of eight as in Sec. III A), which implies that only one ear is stimulated.

Speech-weighted Leuven intelligibility sentences test (LIST) noise from van Wieringen and Wouters (2008) has been 100% AM with an (adjusted) modulation frequency close to 10 Hz. The resulting stimulus has been applied with intensities of 50 and of 70-dB SPL to the ear with the smallest pure-tone average (PTA). Multi-channel EEG recordings

TABLE II. For each channel from Table I, mean SNRs (and standard deviations between brackets) of modulation frequencies (82, 90, 98, 106 Hz left, and 86, 94, 102, 110 Hz right) modulating the corresponding carrier frequency (500, 1000, 2000, and 4000 Hz, left and right) at a stimulus level of 30-dB SPL are displayed. Recording lengths are 32 sweeps after artifact rejection. Mean rms values (and standard deviations) of noise from each channel are shown at the right of the table.

Channel No.	Left (dB SNR)				Right (dB SNR)				rms (nV)
	500 Hz	1000 Hz	2000 Hz	4000 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	
1 (Oz)	5.1 (6.5)	9.3 (2.8)	8.3 (4.0)	5.9 (6.3)	7.1 (5.2)	9.5 (7.0)	10.2 (3.9)	6.8 (5.6)	4.0 (1.3)
2 (P4)	4.2 (5.9)	5.8 (5.4)	4.9 (4.6)	3.1 (5.6)	5.9 (6.7)	9.7 (7.1)	10.1 (3.1)	6.0 (5.1)	2.5 (0.6)
3 (P3)	5.5 (4.7)	8.5 (4.6)	7.9 (4.1)	4.7 (4.9)	4.1 (5.3)	5.6 (6.3)	2.1 (5.4)	−0.4 (4.5)	2.7 (0.6)
4 (M2)	3.8 (7.6)	7.7 (5.0)	4.6 (6.6)	3.6 (7.0)	7.4 (7.7)	12.0 (5.5)	9.9 (2.5)	6.8 (4.7)	7.1 (3.1)
5 (M1)	4.5 (7.6)	10.2 (4.8)	8.1 (6.8)	5.1 (5.7)	4.5 (6.1)	7.2 (6.5)	8.7 (4.4)	3.4 (5.1)	6.1 (1.9)
6 (F4)	0.9 (2.2)	1.8 (5.0)	−0.5 (2.5)	−0.3 (4.7)	2.5 (3.1)	4.6 (4.5)	5.4 (3.0)	0.7 (3.2)	4.9 (3.0)
7 (F3)	2.5 (3.0)	5.0 (5.7)	1.9 (8.6)	0.5 (6.4)	3.0 (4.6)	3.9 (5.5)	2.9 (6.8)	0.6 (5.5)	3.7 (1.8)
8 (Fpz)	0.1 (4.8)	2.7 (5.1)	0.9 (5.1)	−1.3 (5.0)	1.8 (6.9)	3.5 (4.2)	3.3 (4.8)	0.7 (3.0)	4.9 (2.6)

have a length of 23 sweeps, reduced to 20 sweeps by removing epochs that exceed the calculated artifact rejection threshold (per channel or per combination of channels), similar to Sec. III A.

IV. RESULTS FOR BRAINSTEM ASSRS

This section addresses the following questions:

- What is the most appropriate single EEG-channel to record brainstem ASSRs?
- What is the most appropriate combination of multiple EEG-channels to record brainstem ASSRs?
- Are above conclusions influenced by the side of stimulation (left or right)?

All results in this section are based on the concept of *difference scores*. A difference score is the difference (in dB) between a ASSR threshold (in dB SPL) and a behavioral threshold (in dB SPL). The difference score is used in various ASSR studies (Dimitrijevic *et al.*, 2002; Herdman and Stapells, 2001; Luts and Wouters, 2005). In this study, it allows the available SNR data at six different intensities to be combined into one quantitative value which averages out the variation in behavioral thresholds between subjects and test sessions within the same subject.

The *ASSR threshold* is defined as the lowest intensity that still produces a ASSR at the corresponding modulation frequency. In the current study this definition is stricter to avoid incorrect threshold estimations. A threshold (in dB SPL) is only accepted if more than half of the intensity levels above this threshold (in steps of 10-dB SPL up to 60-dB SPL) generate a significant response. For example, when stimulus levels at 30- and 40-dB SPL show a response, but those at 50- and 60-dB SPL do not, one does not define a threshold at 30-dB SPL. However, if a response is found at 20-, 30-, and 40-dB SPL, the ASSR threshold is defined at 20-dB SPL. If no threshold can be defined at 60-dB SPL, the ASSR threshold is set to 70-dB SPL.

First, results are shown using SNRs. Next, the justification to convert SNRs to difference scores is presented. The rest of the section presents results using difference scores.

A. Separate channels

For each EEG-channel, Table II displays mean SNRs (and standard deviations) of eight modulation frequencies. Also, root-mean-square (rms) values of the noise are shown for a frequency range between 77 and 115 Hz (modulation frequencies omitted). Intensity of stimulation is 30-dB SPL. Recording length is 32 sweeps after artifact rejection. The table shows means over ten subjects, with test and retest data averaged for each subject (test-retest not significant over all intensities, $p=0.404$, no interactions). For conciseness, values at other intensities (60-, 50-, 40-, 20-, and 10-dB SPL) are not shown.

When comparing 160 difference score values obtained from all eight frequencies and all ten subjects (both test and retest) with 160 corresponding SNRs obtained from all frequencies and all subjects at a specific intensity, the following Spearman correlations r_s are obtained ($p<0.001$): $r_s=-0.110$ (10-dB SPL), $r_s=-0.378$ (20-dB SPL), $r_s=-0.676$ (30-dB SPL), $r_s=-0.707$ (40-dB SPL), $r_s=-0.678$ (50-dB SPL), and $r_s=-0.573$ (60-dB SPL). The strong correlations just above ASSR hearing threshold justify the condensation of six SNR values into a single-valued difference score. This simplifies analyses further on and allows difference scores to be used instead of objectively measurable SNRs. The test-retest statistic for difference scores is not significant ($p=0.439$, no interactions), same for the test-retest statistic of SNRs. As such, difference scores obtained from the same subject can be averaged.

Figure 1 shows mean difference scores (and standard deviations) per EEG-channel for all eight modulated carrier frequencies. Difference scores vary significantly over channels ($p=0.004$). Comparing pairwise (with Bonferroni correction, $p<0.05$), channel 1 has significantly better difference scores than channels 3, 6, 7, and 8; channel 3 has significantly better difference scores than channel 6; channel 4 has significantly better difference scores than channels 7 and 8. There is an interaction between subjects and channels ($p=0.028$). This indicates that a significant difference is present between the best channel of one subject compared to

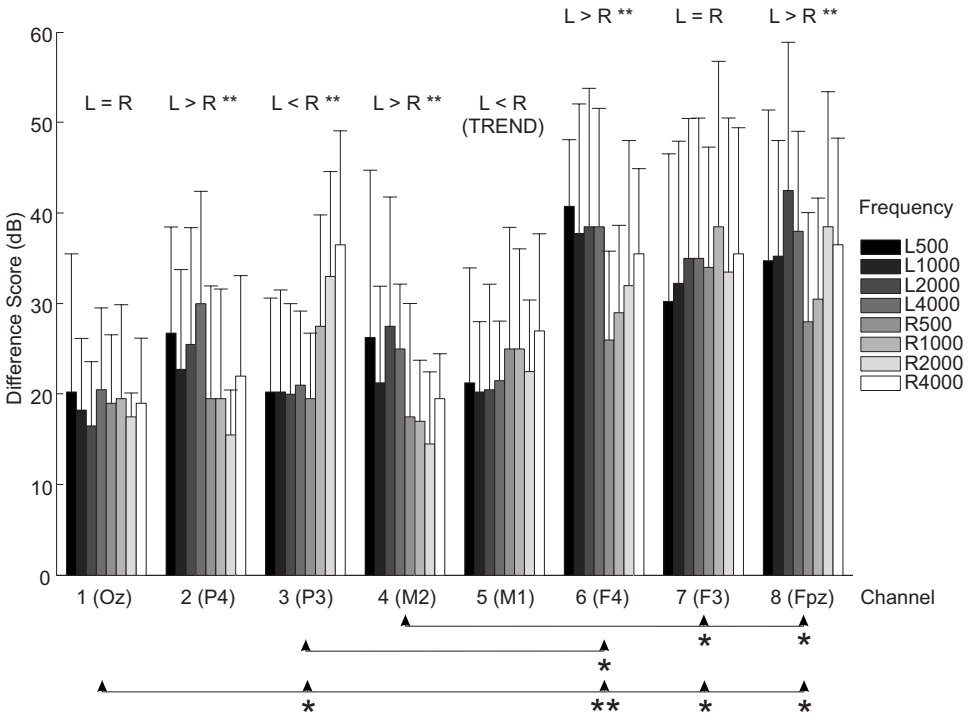


FIG. 1. For each channel from Table I, mean difference scores (and standard deviations) are displayed for all modulated carriers (500, 1000, 2000, and 4000 Hz, left and right). Above each cluster of bars, the relation between carriers applied to the left ear (L) and the right ear (R) is shown. Under each cluster of bars, a comparison between channels is displayed. If a statistical difference is present, this is indicated with * ($0.05 > p \geq 0.01$) and ** ($0.01 > p \geq 0.001$). A trend is considered present if $0.1 > p \geq 0.05$.

the best channel of another subject. Using one particular predefined EEG-channel to optimize recording results cannot be generalized across subjects.

There is no significant difference ($p \geq 0.05$) between frequencies and left-right presentation over all channels. However, there is an interaction effect between channels and left-right application ($p = 0.006$). Figure 1 shows which channels present a significant difference between application to the left ear and application to the right ear. Electrodes placed on the right hemisphere will produce significantly smaller difference scores for carrier frequencies that are applied to the right ear. A similar but somewhat smaller effect can be observed for the left hemisphere. These effects are significantly present on channels 2–4 and 6 and are confirmed on all these channels with six out of ten subjects. Nine out of ten subjects show these effects in at least three out of four significant channels.

If only three out of ten electrodes are available for placement, the best position for the active and reference electrode would be vertex (Cz) and occiput (Oz), respectively, which corresponds to channel 1. This channel returns significantly smaller difference scores than four other channels, including all channels with electrodes on the front side of the head. The side of stimulation is not of significant importance for this channel. This conclusion can also be derived from Table II, where in general higher SNRs are recorded using channel 1.

B. Combination of channels

Section IV A presents data of all individual channels in Table I. This results in a proposal for a channel that records the smallest difference scores. This section discusses the optimal combination of these channels for maximum detection performance (in terms of smallest difference scores), valid for as many subjects as possible and using a minimum number of electrodes. Channels are combined using the detection theory based multi-channel processing technique described in Sec. II. Basically this technique is weighting each channel optimally such that the combined result has the theoretically largest SNR for the ASSR one is looking for (Van Dun *et al.*, 2007b, 2009).

All 255 ($2^8 - 1$) combinations of all eight EEG-channels in Table I are subject to analysis. Channels are combined using Eq. (20). Adding extra channels increases the number of detections. The number of extra channels that can be added is limited however as the number of both true and false detections increases. To keep under control the number of false detections, i.e., detections originating from noise only (which is kept at 5% in this study), detection decision thresholds need to be tightened. In theory, a Bonferroni correction should be applied. In practice, however, this correction proofs too strict. A detection threshold η is calculated using "noise" frequencies with no response present. These noise frequencies are selected close to the modulation frequencies that could contain a response (for example, 1 Hz smaller). To calculate the noise estimate, frequency bins with a response are removed. By selecting the 95 percentile of this distribution of $\Upsilon_A(z)$ values that are calculated on noise frequencies, the practical detection threshold η is found. This is done for all 255 combinations.

For all subjects (both test and retest), difference scores of eight modulated carrier frequencies are calculated for all 255 combinations. No significant difference between test and retest is found ($p = 0.754$). Therefore, for each subject a difference score average is available per frequency and per combination ($10 \times 255 \times 8$ difference scores in total).

Difference scores of most combinations are not normally distributed across subjects. Multivariate analysis requires data with a normal distribution. We follow the nonparametric analysis described in van der Reijden *et al.* (2005), in which

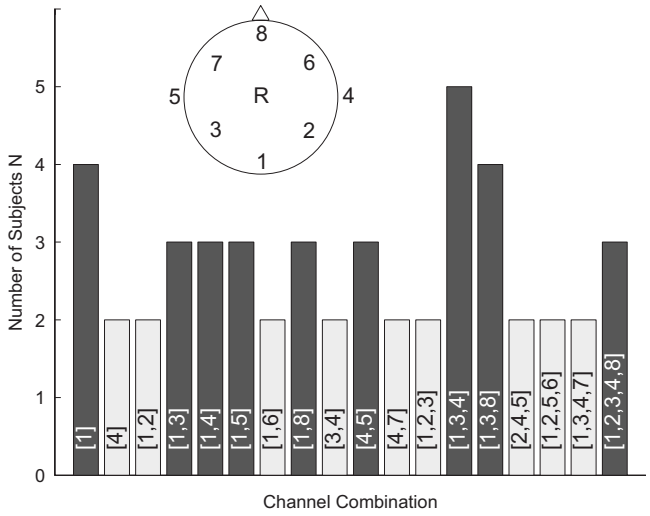


FIG. 2. In Sec. IV B 1, 255 channel combinations per subject are ranked based on each combination's difference score (lowest is best). Difference scores are determined using recorded brainstem responses evoked by left and right stimulus presentations (cf. fourth column of Table III). The top-ten of each list is of interest here. Dark bars represent channel combinations that appear in more than two top-ten lists. Channel combinations appearing only twice in these lists are displayed using light bars. 56 combinations appear only once and are not shown here. Channels are combined using Eq. (20). Only channel combinations represented by dark bars are withheld. These exceed the upper threshold $U=2$, determined using Monte Carlo simulations. The inset represents a scalp viewed from above showing electrode positions from Table I.

outliers are not an issue and equal weights are given to the combinations from each subject regardless of his/her average difference scores.

1. Construction of a top-ten list of combinations

For each subject, a top-ten list is made of combinations with the smallest frequency-averaged difference scores. Five ties are included, limited to the same number of combined channels. For example, when channel combination [1,4,8] is ranked 10th, channel combination [3,6,7] with the same difference score is still included, while channel combination [2,3,4,5] (also with the same difference score) is not, as this combination requires one extra electrode which would be less practical. In this particular case, the top-ten list thus contains 105 combinations. The number of times a particular combination appears in the lists is summed over subjects. This is summarized in Fig. 2. The histogram is subject to further analysis identifying any high ranking combinations that are dominant in all subjects and that will therefore suffice to record the ASSR efficiently.

The average chance level $\bar{d}$ for a specific combination to appear in a subject's top-ten list is $\bar{d}=105/255=0.412$. Monte Carlo simulations are used to find the upper limit U of the 95% confidence interval of the 95 percentile of this distribution d (Press *et al.*, 1992). If this value U can be found, a specific combination only appears more than U times in the top-ten list by chance in less than 5% of the cases. Moreover, this statement has a reliability of 95%. To practically calculate upper limit U, 1000 distributions d_i are constructed. From each distribution d_i, the 95 percentile P_i^{95} is determined. This returns 1000 values P_i^{95}. A 95% confidence interval (and its upper limit U) can be calculated of its mean value $\bar{P}^{95}$.

Each of 1000 distributions d_i, containing (an arbitrarily chosen number of) 5000 instances N_j, is constructed as follows. Out of 255 possible combinations (represented by values between 1 and 255, each corresponding to a particular combination), a random selection S_j of 105 combinations is made. An arbitrary combination c_j is chosen, and the number of appearances N_j of this combination c_j among selection S_j is counted. Instance N_j is one of 5000 instances that distribution d_i consists of. As a validation for this approach, it is confirmed that mean values $\bar{d}_i$ are indeed all close to 105/255. This results in an upper limit U equal to 2. All values above $U=2$ in the histogram of Fig. 2 (depicted by the dark bars) are particular combinations not appearing by chance ($p=0.05$). This is a statement made with a confidence of 95%.

The channel combinations in Fig. 2 that appear more than $U=2$ times in the top-ten list with the smallest difference scores over all subjects are [1] [1,3] [1,4] [1,5] [1,8] [4,5] [1,3,4] [1,3,8] [1,2,3,4,8]. This series is also displayed in the fourth column of Table III. This result seems promising as these combinations generally appear to contain channels with the smallest difference scores (Fig. 1). However, a comparison of test and retest data shows that test-retest reproducibility of combinations in the top-ten list is low. For test data the withheld combinations are [2] [1,2] [1,3] [1,5] [1,6] [1,8] [4,7] [1,2,4] [1,3,7] [1,2,3,4]. For retest data these are [1] [1,3] [1,4] [4,5] [1,3,4]. This non-reproducibility effect is mainly caused by the fact that a large number (i.e., 255) of combinations are candidates for the top-ten list. As a result, even a "promising" combination (one that has a small difference score) might not end up in the top-ten list of both test and retest results due to variations in experiment conditions. By focusing on individual channels appearing in the combinations, this test-retest difference may be reduced, as now only eight channels need to be ranked.

2. Relative contribution of each channel to a top-ten list

The relative contribution of each individual channel to the top-ten lists described in Sec. IV B 1 is an indicator of the importance of each individual channel. For each subject, the number of appearances of a particular channel in the subject's top-ten list is counted and divided by the number of channels in the top-ten list. Test and retest are not significantly different ($p=0.509$, no interactions). Test-retest combined data are displayed per subject in Fig. 3, together with the total relative contribution per channel over all subjects. Relative contribution figures can also be found in the fourth column of Table III. An interesting example shows that "unusable" channels are discarded immediately. During recording of data in the first (test) session, channel 4 from both subjects 2 and 4 suffered severely from muscle artifacts at the right mastoid. No contribution is allowed for channel 4 in these subjects.

Based on the total relative contribution over all subjects,

TABLE III. Step-by-step derivation of most optimal electrode combination for responses originating mainly from the brainstem (Sec. IV B) and mainly from the auditory cortex (Sec. V B). The second row shows all combinations appearing more than $U=2$ times in the top-ten lists for a specific type of stimulation (cf. Fig. 2). The third row depicts the relative contribution and ranking of individual channels to these top-ten lists (cf. Fig. 3). The fourth row indicates the minimum channel set $\mathcal{M}$ needed to cover a certain number of subjects. By adding more channels to a minimal set, more subjects are covered. Here, the minimum channel set $\mathcal{M}_{80}$ covering at least 80% of all subjects (8 out of 10) is depicted in bold and shown separately in the fifth row. This channel set is recommended for practical use.

	Brainstem (80–110 Hz)			
	Left	Right	Combined	Cortex(10 Hz)
Selected combinations ($U>2$)	**[1][1,3]**[1,4] [1,5][1,8][4,5] [1,3,4][1,3,8]	**[1][2][4]** **[1,2][2,4]**[3,4] [4,7]**[1,2,4]**	**[1][1,3][1,4]**[1,5] [1,8][4,5]**[1,3,4]** [1,3,8][1,2,3,4,8]	**[5]**[1,5][4,6]**[5,7]** [2,5,7][4,5,7]**[5,6,7]** **[5,7,8]**[2,5,7,8]**[5,6,7,8]** [4,5,6,7,8][2,4,6,7,8]
Relative contribution (%)	22 11 17 16 13 6 7 8	19 16 12 18 9 11 7 8	23 13 15 16 10 9 6 9	6 10 6 12 25 12 18 11
Channel ranking	1 3 4 5 2 8 7 6	1 4 2 3 6 5 8 7	1 4 3 2 5 6 8 7	5 7 6 4 8 2 3 1
Minimum sets $\mathcal{M}$	{1}–5 **{1,3}–9** {1,3,8}–10	{1}–3 {1,4}–7 **{1,4,2}–8** {1,4,2,3}–9 {1,4,2,3,6}–10	{1}–4 {1,4}–7 **{1,4,3}–10**	{5}–5 {5,7}–6 {5,7,6}–7 **{5,7,6,8}–9** {5,7,6,8,2}–10
Recommended electrodes based on this dataset	{1,3}	{1,4,2}	{1,4,3}	{5,7,6,8}

a ranking can be produced for all eight channels. For the current analysis the channel ranking is 1 4 3 2 5 6 8 7 (cf. Table III). Guided by this ranking, the minimum number of channels can be determined that covers as many top-ten lists as possible. The step-by-step construction of this minimum channel set $\mathcal{M}$ is displayed in Table III. The minimum channel set $\mathcal{M}_{80}$ covering at least 80% of all subjects is depicted in bold. In the fourth column of Table III, at least one of seven combinations that can be made from the set $\mathcal{M}_{80}=\{1,4,3\}$ will appear in ten out of ten (i.e., all) top-ten lists.

When minimum channel set $\mathcal{M}_{80}=\{1,4,3\}$ is determined, the list of selected combinations (with an appearance $U>2$) from the fourth column of Table III can be reduced. Channel combinations from this list that can be formed by $\mathcal{M}_{80}$ are [1] [1,3] [1,4] [1,3,4]. Mean difference scores from these channel combinations range between 18.5 and 19.1 dB ($\pm 1.2-1.5$ dB). Difference scores of all these combinations are not significantly different and are significantly ($p<0.05$) pairwise correlated. As a result, any channel selection out of [1] [1,3] [1,4] or [1,3,4] will result statistically in a minimum mean difference score. To be able to select a channel combination that guarantees a difference score that is always in the top-ten of each subject, three active electrodes should be placed at position 1 (Oz), position 3 (P3), and position 4 (right mastoid), a selection corresponding to minimum set $\mathcal{M}_{80}$. Table IV illustrates this by showing difference scores of each of these four combinations for each subject, together with their ranking compared with all other combinations within that subject.

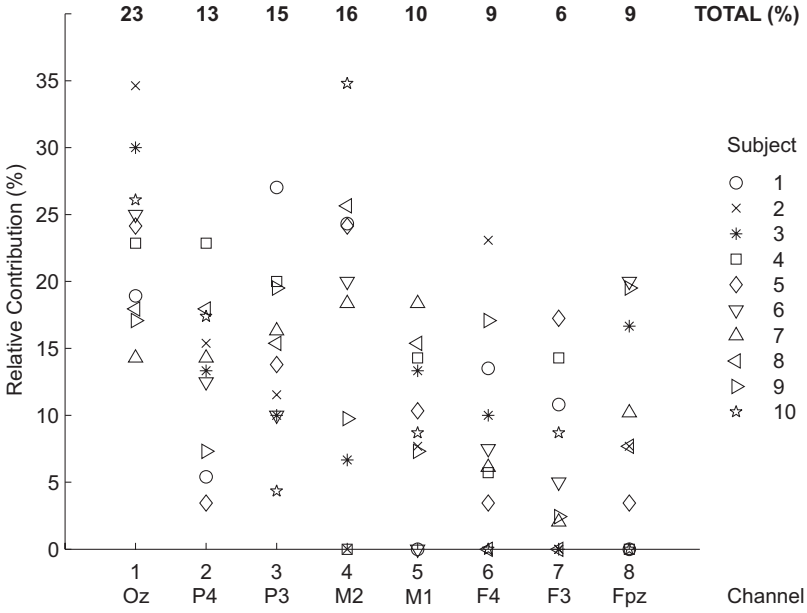


FIG. 3. In Sec. IV B 1, 255 channel combinations per subject are ranked based on each combination's difference score (lowest is best). Difference scores are determined using recorded brainstem responses evoked by left and right stimulus presentations (cf. fourth column of Table III). The top-ten of each list is of interest here. The relative contribution of each of eight channels to each top-ten list per subject is displayed. Additionally, the relative contribution per channel over all subjects is shown.

TABLE IV. Difference scores (dB) are averaged over all eight frequencies for six relevant channel combinations. The rank of each channel combination, relative to all other combinations within the same subject, is given between parentheses. The six channel combinations displayed here are the best combination with the smallest difference score, the worst combination with the largest difference score, and the four combinations selected in Sec. IV B 2. Channel combinations that are present in a subjects's top-ten are highlighted in bold.

Subject	1	2	3	4	5	6	7	8	9	10
Best	14.4 (1)	20.9 (1)	18.1 (1)	13.1 (1)	13.1 (1)	12.5 (1)	16.9 (1)	20.3 (1)	18.4 (1)	10.0 (1)
[1]	21.9 (154)	**22.8 (9)**	**18.1 (1)**	17.5 (58)	14.4 (11)	17.5 (44)	19.3 (83)	25.3 (194)	**20.9 (10)**	**10.0 (1)**
[1,3]	**14.4 (1)**	23.4 (15)	**19.4 (3)**	**13.8 (3)**	15.6 (52)	19.4 (114)	20.0 (134)	23.4 (91)	22.2 (35)	13.8 (23)
[1,4]	20.0 (83)	22.8 (12)	20.6 (13)	20.0 (145)	14.4 (12)	**15.0 (6)**	**18.1 (9)**	22.2 (27)	24.1 (108)	**11.3 (4)**
[1,3,4]	**16.3 (6)**	25.9 (90)	**20.0 (6)**	18.8 (103)	**13.8 (6)**	17.5 (47)	18.1 (15)	**20.9 (3)**	27.2 (188)	**12.5 (10)**
Worst	41.3 (255)	37.2 (255)	53.8 (255)	30.0 (255)	29.4 (255)	33.8 (255)	41.9 (255)	45.9 (255)	52.2 (255)	42.5 (255)

3. Splitting data in left and right stimulus application

To check whether there is a specific reason for the asymmetry in the previous setup, the available data are split according to the two sides of stimulation (left or right). This way, the possible importance of side of stimulation for electrode placement is investigated.

Responses originating from stimuli applied to the left ear are considered separately from those resulting from stimuli applied to the right ear. The procedures from Secs. IV B 1 and IV B 2 are repeated identically.

As shown in the second column of Table III, the minimum set $\mathcal{M}_{80}^{L}$ for stimuli applied to the left ear is $\mathcal{M}_{80}^{L}=\{1,3\}$. This corresponds to active electrodes placed at positions Oz and P3. $\mathcal{M}_{80}^{R}=\{1,4,2\}$ is taken as a minimum set $\mathcal{M}_{80}^{R}$ for stimuli applied to the right ear. Active electrodes should be placed at positions Oz, P4, and the right mastoid. In seven of ten subjects, electrodes on the side of stimulation contribute relatively more to the top-ten lists than the contralateral electrodes. In only one subject, this effect is inverted.

The resulting minimum set $\mathcal{M}_{80}=\{1,4,3\}$ of simultaneous stimuli left and right is a combination of minimum sets $\mathcal{M}_{80}^{L}$ and $\mathcal{M}_{80}^{R}$ of stimuli applied to left and right ears separately. No apparent reason for the asymmetry in $\mathcal{M}_{80}$ can be found however. The asymmetry can be mainly addressed to variability of data across different experiments. When compared with symmetric sets {1,2,3} and {1,4,5}, the sum of relative contributions of single channels from asymmetric set {1,4,3} to top-ten lists is largest in five of ten subjects. Only in one subject, this sum is the smallest. In individual subjects, this asymmetry is thus also present.

C. Artifact robustness

Four different processing methods are considered, two single-channels and two multi-channels. Channel 1 is taken from Table I as the reference channel for single-channel brainstem ASSR measurements with adults (Luts and Wouters, 2005; van der Reijden *et al.*, 2001). Channels 1, 3, and 4 are presented as the minimum set in Sec. IV B 2 and are combined using the multi-channel processing scheme from Sec. II.

The processing schemes are applied on two multi-channel EEG datasets obtained at an intensity of 30-dB SPL. First, they are administered on an EEG dataset with few artifacts ("clean EEG") that is identical to the EEG in Sec. IV A. Second, they are applied on an EEG dataset with a considerable number of artifacts ("dirty EEG"), generated as described in Sec. III A. To calculate a detection threshold η, each of four processing methods below is applied on frequencies without a response (for example, 1 Hz below the modulation frequencies that could contain a response) of artifact-free multi-channel EEG data from Sec. IV A on all intensities. Then, the 95 percentile of this noise distribution is defined as the detection threshold. This way, four different detection thresholds are obtained for each processing method.

1. *Channel 1, normal averaging*. Channel 1 is divided into 32 sweeps (data blocks of 16.384 s) and averaged. Detection is managed using the method described in Sec. II D.
2. *Channel 1, noise-weighted averaging*. Each unfiltered epoch (data blocks of 1.024 seconds) is transformed to the frequency domain using a FFT. Average power P_i between 77 and 115 Hz is computed after removing the power at eight frequencies where responses occur. The time domain epoch is weighted with the average power P_i and concatenated with preceding epochs to form sweeps. Each epoch of the final summed sweep is divided by the sum of the weights $P=\Sigma P_i^{-1}$ of the epochs that have been combined to form that particular epoch [adapted from John *et al.* (2001a)]. Detection is managed using the method described in Sec. II D.
3. Channels 1, 3, and 4 are combined using Eq. (20) with $\boldsymbol{K}_{\text{spat}}$ fixed and $\boldsymbol{K}_{\text{temp}}=\sigma\boldsymbol{I}$ fixed. Mean noise power σ^2 is calculated using Eq. (23) on combined channels 1, 3, and 4.
4. Channels 1, 3, and 4 are combined using Eq. (22) with $\boldsymbol{K}_{\text{spat}}$ fixed. $\boldsymbol{K}_{\text{temp},i}=P_iP\sigma_n\boldsymbol{I}$ is variable and is recalculated for each block i of 8.192 s (this block size returns best results). $\boldsymbol{K}_{\text{temp},i}$ is calculated based on the rationale of noise-weighted averaging. P_i is the average power between 77 and 115 Hz of data block i. P is the sum of the reciprocals of P_i, i.e., $P=\Sigma P_i^{-1}$. Mean noise power σ_n^2 is calculated using Eq. (23) on a noise-weighted average of combined channels 1, 3, and 4.

It is assumed that all responses to the applied stimuli are present in the EEG as the stimulus intensity is 30-dB SPL and hence above the subject's hearing thresholds. This assumption does not guarantee the presence of ASSRs in the EEG however. Adult ASSR thresholds are assumed to lie at

TABLE V. The number of detections (with an assumed maximum of 160) is shown for each processing method described in Sec. IV C. Each processing method is applied to EEG data with few artifacts ("clean EEG") and EEG data with a significant number of artifacts ("dirty EEG"). For methods 1–4, comparisons have been conducted pairwise with Bonferroni correction. Methods 3′, 3″, 4′, and 4″ process two symmetric channel sets but are not used for pairwise comparisons. They are only placed in this table for illustrative purposes.

Method	Description	Clean EEG	Dirty EEG
1	Normal averaging channel 1	117	17
2	Weighted averaging channel 1	122	35
3	$\boldsymbol{K}_{\text{temp}}=\sigma\boldsymbol{I}$ fixed; [1,3,4]	117	57[a]
4	$\boldsymbol{K}_{\text{temp},i}=\boldsymbol{P}_i\boldsymbol{P}\sigma_n\boldsymbol{I}$ var; [1,3,4]	**122**	**72**[b]
3′	$\boldsymbol{K}_{\text{temp}}=\sigma\boldsymbol{I}$ fixed; [1,2,3]	113	31
4′	$\boldsymbol{K}_{\text{temp},i}=\boldsymbol{P}_i\boldsymbol{P}\sigma_n\boldsymbol{I}$ var; [1,2,3]	122	50
3″	$\boldsymbol{K}_{\text{temp}}=\sigma\boldsymbol{I}$ fixed; [1,4,5]	108	44
4″	$\boldsymbol{K}_{\text{temp},i}=\boldsymbol{P}_i\boldsymbol{P}\sigma_n\boldsymbol{I}$ var; [1,4,5]	112	66

[a]Significantly better than approach 1 ($p<0.01$).
[b]Significantly better than approach 1 ($p<0.001$), approach 2 ($p<0.01$), and approach 3 ($p<0.05$).

12 ± 9 dB above hearing threshold for 10-min measurements (Luts and Wouters, 2004). As a result, even for acoustic stimuli that are, for example, 20 dB, above hearing threshold, ASSR presence cannot be guaranteed. However, the different processing schemes are evaluated using identical data sets, eventually eliminating this problem of uncertainty. Hence it is assumed that there are 16 responses available for detection for each subject (test and retest). The maximum number of detections is assumed to be 160. Table V shows the number of detections for each approach: EEG data with few artifacts (clean EEG) and EEG data with a significant number of artifacts (dirty EEG). Test and retest data are not significantly different for both datasets ($p\geq0.67$).

For EEG data with few artifacts, best results are obtained when a noise-weighted averaging approach is used. The improvement relative to normal averaging is not significant however. It does not make a difference if multiple channels are used for processing or just one (reference) channel.

Response detection in EEG data with numerous artifacts is improved significantly when using more than one channel for processing. This improvement can increase significantly again when using a noise-weighted averaging based approach.

In Sec. VI, it is discussed whether a symmetric minimum set would give a more logical electrode placement compared with an asymmetric minimum set. Therefore, for illustrative purposes, Table V adds four extra lines showing the multi-channel processing results of symmetric minimum sets {1,2,3} and {1,4,5}.

V. RESULTS FOR AUDITORY CORTEX ASSRs

This section addresses the following questions:

- What is the most appropriate single EEG-channel to record auditory cortex ASSRs?
- What is the most appropriate combination of multiple EEG-channels to record auditory cortex ASSRs?
- Are above conclusions influenced by the side of stimulation (left or right)?

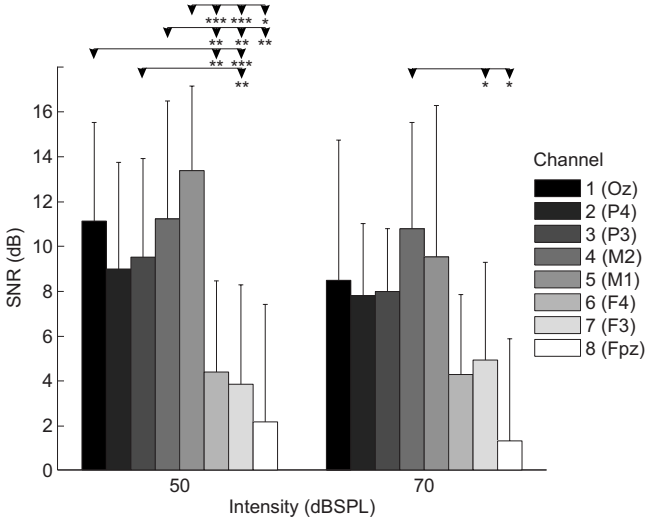


FIG. 4. The mean SNR (and standard deviation) of the 10-Hz response is displayed per channel from Table I. The stimulus is applied to the right ear (next to channel 4) at intensities of 50- and 70-dB SPL. Above each intensity, a comparison between channels is made. If a statistical difference is present, this is indicated with * ($0.05>p\geq0.001$), ** ($0.01>p\geq0.001$), and *** ($0.001>p$).

This section presents results from stimulation at only one ear. Six out of ten subjects are stimulated at the right ear. To preserve symmetry, all electrode positions of the other four subjects are mirrored as if stimulation occurred at the right ear. This swapping operation is allowed as no significant difference ($p=0.904$) in SNR between groups with left ear stimulation and right ear stimulation is found. As a result, Table I refers for this section to the (corrected) side of stimulation as "ipsilateral" and the opposite side as "contralateral."

A. Separate channels

Figure 4 shows mean SNRs (and standard deviations) of the 10-Hz response in all eight channels from Table I for intensities of 50- and 70-dB SPL. Test and retest data are averaged per subject ($p=0.314$, no interactions). There is no significant difference between both intensities ($p=0.113$, no interactions). A main effect between channels is present ($p=0.003$, no interactions). When comparing pairwise (Bonferroni correction) in Fig. 4 at 50-dB SPL, SNRs of channels 4 and 5 are significantly larger than those of channels 6–8. Channel 1 is significantly different from channels 6 and 7. Channel 3 only is significantly different from channel 7. At 70-dB SPL, only SNRs of channel 4 are significantly larger than those of channels 7 and 8. Between subjects an interaction is present with EEG-channels ($p=0.044$). This indicates that the EEG-channel with the largest SNR for one subject is not necessarily the channel with the largest SNR for another subject. The best EEG-channel is subject-dependent.

Due to high variability of SNRs, it is difficult to identify an optimal electrode placement from Table I when only three electrodes are available for placement. From the pairwise comparisons above, channels 4 and 5 (both mastoids) appear to be good choices.

B. Combination of channels

Although no significant difference is present between both intensities in Sec. V A, only the results obtained at 50-dB SPL are focused on. The structure of this section is similar to the structure of Sec. IV. Results at 70-dB SPL are similar and will only be described briefly where different from those at 50-dB SPL.

In Sec. IV, difference scores between ASSR thresholds and behavioral thresholds are used to quantify performance. As no behavioral data are available, only SNRs are used in the analysis, with a minor modification. For example, a response of 12 dB resulting from a five-channel combination is not as convincing as a 12-dB response from a three-channel combination. As a result, the concept of corrected SNR is applicable. The *corrected SNR* is the difference between the SNR of the response and a threshold SNR. The latter increases when more channels are added for combination due to the statistical multiple testing problem. The *threshold SNR* is determined as the SNR that allows 5% false detections when observing a frequency bin without any response present (for example, 1 Hz below the stimulus frequency of 10 Hz). To calculate the noise estimate, frequency bins with a response are removed.

Again, all 255 combinations of all eight EEG-channels in Table I are subject to analysis. Channels are combined using Eq. (20) from Sec. II. No significant difference between corrected SNRs of test and retest data is found ($p=0.805$). Therefore, for each subject an average corrected SNR is available for the 10-Hz response and per combination (10×255 corrected SNRs in total). Corrected SNRs of most combinations do not have a normal distribution across subjects. Therefore, again a top-ten list is made for all subjects. Zero ties are included. The top-ten list thus contains 100 combinations. The number of times a combination appears in this top-ten list is counted (with a mean chance level of $\bar{d}=100/255=0.392$). All combinations appearing more than $U=2$ times in the top-ten list (U as a result of Monte Carlo simulations) are considered not appearing by chance ($p=0.05$), with a confidence of 95%. The withheld combinations for the 10-Hz ASSR are [5] [1,5] [4,6] [5,7] [2,5,7] [4,5,7] [5,6,7] [5,7,8] [2,5,7,8] [5,6,7,8] [4,5,6,7,8] [2,4,5,6,7,8]. They are also displayed in the fifth column of Table III. Only combinations [5] [5,7] [2,5,7] [5,7,8] can also be found in both test and retest results. This nonreproducibility effect is mainly caused by the fact that a large number (i.e., 255) combinations are candidates for the top-ten list. To guarantee test-retest reproducibility, the focus is on the individual channels appearing in these combinations instead. The test-retest difference may be reduced, as now only eight channels need to be ranked.

According to the fifth column of Table III, channels 5 and 7 are by far the channels with highest occurrence in the subjects's top-ten lists. By adding channels 6 and 8, 90% of all subjects are covered. The minimum channel set is thus $\mathcal{M}_{80}=\{5,7,6,8\}$. However, when following the channel ranking of Table III exactly, the minimum set {5,7,6,4} should be the first choice. This minimum set offers the same amount of coverage as {5,7,6,8}, but combination [5,7,6,4] does not appear in the list of withheld combinations however, so channel 8 is chosen instead, and consequently for minimum set {5,7,6,8}.

Based on the minimum channel set $\mathcal{M}_{80}$, only channel combinations [5] [5,7] [5,6,7] [5,7,8] [5,6,7,8] are considered from the list with withheld combinations. Their mean corrected SNRs are between 8.2 and 9.0 dB (±1.1 and 1.2 dB) and not significantly different. They are all significantly pairwise correlated ($p<0.05$). As a result, any channel selection out of these five combinations will result statistically in a maximum mean corrected SNR. To be able to select a channel combination that guarantees a corrected SNR that is always in the top-ten of each subject, however, four active electrodes should be placed at position 5 (contralateral mastoid), position 6 (F4), position 7 (F3), and position 8 (forehead). When data obtained at 70-dB SPL are analyzed, a coverage of 80% is possible with $\mathcal{M}_{80}=\{5,7,6,4\}$, which is similar to the results at 50-dB SPL.

VI. DISCUSSION

All combinations of EEG-channels in this paper are processed using a sufficient statistic that theoretically captures best the amount of useful ASSR target signal present in these EEG-channels. This sufficient statistic originates from the realm of detection theory (Green and Swets, 1966). It can be linked with the multi-channel Wiener filter, an optimal filtering technique which computes an optimal (minimum mean square error) estimate of a reference signal (Scharf, 1991). In the case of ASSR detection, this comes down to a maximization of the SNR of combined channels (Simmer *et al.*, 2001; Van Dun *et al.*, 2007b). This point of view makes it easier to get an idea of what happens when combining EEG data using Eq. (20).

Adding extra data-channels will increase the amount of information that is gathered in the same time period. However, to combine these channels, the use of more advanced signal processing is necessary to truly achieve a performance improvement. If multi-channel data are combined using, for example, simple time domain averaging over channels, the statistical multiple testing penalty will increase the number of false detections proportionally to the number of extra correct detections. By intelligently combining these channels, this multiple testing problem can partly be avoided.

This section discusses results from Sec. V where mainly the brainstem (modulation frequencies between 80 and 110 Hz) and mainly the auditory cortex (10 Hz) are stimulated.

A. Brainstem stimulation

The transition from SNR to difference scores is allowed due to high correlations between both measures. It is assumed that results for one measure can be used for the interpretation of the results of the other measure.

When observing difference scores of reference channel 1 (Cz-Oz), two observations can be made. First, difference scores are slightly elevated compared to other studies. Standard deviations are comparable however (Dimitrijevic *et al.*, 2002; Herdman and Stapells, 2001; Luts and Wouters, 2005;

Picton *et al.*, 2003). The relatively low-noise levels [compared to other studies, like Dimitrijevic *et al.* (2002)] depicted in the rightmost column of Table II are no guarantee for small difference scores. Low-noise levels do not necessarily result in good response detection. Some subjects have very low-noise levels when they sleep, as well as reduced response amplitudes. In these subjects response detection may be easier when they are awake and noise levels are higher (Luts *et al.*, 2008). Therefore, the difference score gap may be caused by the determination method of ASSR hearing thresholds. The determination method in this study appears to be rather strict. However, it is kept identical for the whole study, and final conclusions should not be influenced. Second, although not significantly different, carrier frequencies of 500 and 4000 Hz show a trend of having larger difference scores than 1 and 2 kHz at these intensities, which is also observed in other studies (Dimitrijevic *et al.*, 2002; Herdman and Stapells, 2001; Luts and Wouters, 2004; Picton *et al.*, 1998). Both observations suggest that the data in this study are representative for general brainstem ASSR measurements.

As already indicated in Sec. III, several studies focus on sources of the ASSR with frequencies between 80 and 110 Hz, the frequency region of interest for hearing threshold determination. They concluded that for modulation frequencies above 75 Hz, most of the ASSR is generated in the brainstem. Above 95 Hz, the source lies entirely in the brainstem. These assumptions seem to support the observations made in this paper. When sources are mainly located in the brainstem, smallest difference scores (and largest SNRs consequently) are recorded in electrodes close to the brainstem, on the back of the head and the mastoids. The ipsilateral effect of larger SNRs and smaller difference scores (and thus the optimal combination of electrodes) at the side of stimulus application is partly confirmed by Small and Stapells (2008). They indicated that ASSR amplitudes are significantly smaller at the contralateral side. Difference scores with adults do not differ significantly however. For infants, this ipsilateral effect is reported to be strongly present, both for ASSR amplitudes as ASSR thresholds. This asymmetric effect is also confirmed by van der Reijden *et al.* (2005). This could signify that the orientation of the ASSR sources is stimulus side dependent for adults and especially stimulus dependent for infants. However, the asymmetrical electrode positioning for binaural stimulation cannot be confirmed in literature. The authors assume that results in this paper are due to the used dataset and that a symmetrical electrode placement could be preferred. According to Table V, a recommended symmetric minimum set could be {1,4,5}, only slightly reducing the number of detections for clean EEG as well as dirty EEG. To conclude, van der Reijden *et al.* (2004) reported that a small set of three derivations (Cz-Oz combined with the right mastoid-Cz and the left mastoid-Cz, which corresponds to the minimum set {1,4,5}) yields the best SNR in a larger number of adults than would be expected if all derivations were equally efficient. This result is very similar to the one obtained in this study.

Multi-channel processing does not significantly improve detection performance when it is applied to clean EEG. This shows that most single-channel studies using a Cz-Oz or Cz-neck electrode configuration get close to the best obtainable result [for an extensive overview, we refer to Picton *et al.* (2003)]. The EEG data in these studies are mostly obtained from relaxed adults and sedated children. However, studies that specifically incorporate EEG data with numerous artifacts are rare. Unfortunately, these situations reflect most of the conditions usually observed in a clinical environment. The proposed technique definitely contributes to the ASSR field as it is significantly more robust against artifacts and may improve measurements in these difficult conditions. This improvement can be attributed to the use of specific multi-channel signal processing on multi-channel EEG data. The use of multi-channel EEG measurements is not new (van der Reijden *et al.*, 2004, 2005). However, the application of the multi-channel signal processing strategy described in Sec. II to multi-channel EEG data is a novel approach.

B. Auditory cortex stimulation

As already indicated in Sec. III, when lowering the modulation frequency of the stimulus, the main source of the ASSR shifts more to the auditory cortex, while there is still a source present in the brainstem. This may explain the SNRs being largest on electrodes close to the brainstem (back of the head) and the auditory cortices (both mastoids).

Although no significant differences in SNR are found in Fig. 4 between single electrodes at the ipsilateral and the contralateral mastoid [also reported by Herdman *et al.* (2002)], the dominant channel (contralateral mastoid, channel 5) of the minimum set {5,7,6,8} clearly is located at the opposite side of stimulation. This could be explained by the crossing of the auditory paths beyond the brainstem (Hall, 2007). Extra channels 6–8 of the minimum set {5,7,6,8} lie at the front of the head, contrary to channels with largest SNRs lying at the back of the head, as shown in Fig. 4. This can be explained by the underlying mechanism of the multi-channel processing technique from Sec. II. The combination of weighted channels theoretically has the largest possible SNR (Van Dun *et al.*, 2007b, 2009). For responses originating mainly from the auditory cortex, this result is likely obtained by taking the channel with the largest response (channel 5) and by adding channels that have a small response (channels 6–8) and that have a high-noise correlation with the noise from the channel with the largest response. This observation is in contrast with conclusions in Sec. VI A, which locate the minimum set {1,4,2} in the back of the head for responses originating mainly from the brainstem. The dominant channel for this minimum set is the occiput (channel 1). Any additional channels with small responses (channels 6–8 in front of the head according to Fig. 1) appear to be too remote as noise correlation with these remote channels is presumably too low in the frequency region of interest (80–110 Hz).

VII. CONCLUSION

In this paper, practical performance of a multi-channel EEG processing strategy for ASSR detection based on a detection theory approach is evaluated. Eight-channel measure-

ments of ten normal-hearing adults have been used. First, ASSRs with modulation frequencies between 80 and 110 Hz and mainly originating from the brainstem are considered. It is shown that electrodes should be placed at the back of the head and at mastoids to obtain largest mean SNRs and smallest mean difference scores for individual electrodes. When applying the multi-channel processing strategy, smallest difference scores are found with all subjects when electrodes are placed at five positions: Oz, P3, and right mastoid with Cz as a reference electrode and a ground electrode at, for example, the right clavicle. This combination is significantly more robust against artifacts when compared to a single-channel, three electrode, setup. The number of ASSR detections more than doubles when EEG data with artifacts are considered. Second, ASSRs with a modulation frequency of 10 Hz and mainly originating from the auditory cortex are studied. Again, largest mean SNRs are obtained at single electrodes located at the back of the head and at both mastoids. When applying multi-channel processing, largest mean SNRs for minimum 80% of subjects are obtained when placing electrodes at the contralateral mastoid, F4, F3, and the forehead with Cz as a reference electrode and a ground electrode at, for example, the right clavicle.

ACKNOWLEDGMENTS

This work was supported in part by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen), in part by FWO Project No. G.0504.04 ("Design and analysis of signal processing procedures for objective audiometry in newborns"), and in part by the Concerted Research Action GOA-AMBioRICS.

Aiken, S. J., and Picton, T. W. (**2008**). "Human cortical responses to the speech envelope," Ear Hear. **29**, 139–157.

Alaerts, J., Luts, H., Hofmann, M., and Wouters, J. (**2009**). "Cortical auditory steady-state responses to low modulation rates," Int. J. Audiol. **48**, 1–12.

Armstrong, M., and Stapells, D. R. (**2007**). "Multiple-stimulus interactions in the brainstem (80 Hz) and cortical (14 & 40 Hz) auditory steady-state responses," in Proceedings of the 20th International Evoked Response Audiometry Study Group (IERASG), Bled, Slovenia, p. 148.

Cebulla, M., Stürzebecher, E., and Elbering, E. (**2006**). "Objective detection of auditory steady-state responses: Comparison of one-sample and q-sample tests," J. Am. Acad. Audiol **17**, 93–103.

Cebulla, M., Stürzebecher, E., and Wernecke, K. D. (**2001**). "Objective detection of the amplitude modulation following response (AMFR)," Audiology **40**, 245–252.

Cohen, L. T., Rickards, F. W., and Clark, G. M. (**1991**). "A comparison of steady state evoked potentials to modulated tones in awake and sleeping humans," J. Acoust. Soc. Am. **90**, 2467–2479.

Dimitrijevic, A., John, M. S., Van Roon, P., Purcell, D. W., Adamonis, J., Ostroff, J., Nedzelski, J. M., and Picton, T. W. (**2002**). "Estimating the audiogram using auditory steady-state responses," J. Am. Acad. Audiol **13**, 205–224.

Dobie, R. A., and Wilson, M. J. (**1996**). "A comparison of t test, f test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am. **100**, 2236–2246.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Galambos, R., Makeig, S., and Talmachoff, P. J. (**1981**). "A 40-Hz auditory potential recorded from the human scalp," Proc. Natl. Acad. Sci. U.S.A. **78**, 2643–2647.

Green, D., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hall, J. W. (**2007**). *New Handbook of Auditory Evoked Responses* (Pearson, Boston).

Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., and Picton, T. W. (**2002**). "Intracerebral sources of human auditory steady-state responses," Brain Topogr **15**, 69–86.

Herdman, A. T., and Stapells, D. R. (**2001**). "Thresholds determined using the monotic and dichotic multiple auditory steady-state response technique in normal-hearing subjects," Scand. Audiol. **30**, 41–49.

John, M. S., Dimitrijevic, A., and Picton, T. W. (**2001a**). "Weighted averaging of steady state responses," Clin. Neurophysiol. **122**, 555–562.

John, M. S., Dimitrijevic, A., and Picton, T. W. (**2002**). "Auditory steady-state responses to exponential modulation envelopes," Ear Hear. **23**, 106–117.

John, M. S., Dimitrijevic, A., and Picton, T. W. (**2003**). "Efficient stimuli for evoking auditory steady-state responses," Ear Hear. **24**, 406–423.

John, M. S., Dimitrijevic, A., Van Roon, P., and Picton, T. W. (**2001b**). "Multiple auditory steady-state responses to AM and FM stimuli," Audiol. Neuro-Otol. **6**, 12–27.

John, M. S., and Picton, T. W. (**2000**). "MASTER: A Windows program for recording multiple auditory steady-state responses," Comput. Methods Programs Biomed. **61**, 125–150.

Johnson, D. H., and Dudgeon, D. E. (**1993**). *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, London).

Joint Committee on Infant Hearing (**1995**). "Joint committee on infant hearing 1994 position statement. American academy of pediatrics joint committee on infant hearing," Pediatrics **95**, 152–156.

Kuwada, S., Anderson, J. S., Batra, R., Fitzpatrick, D. C., Teissier, N., and D'Angelo, W. R. (**2002**). "Sources of the scalp recorded amplitude-modulation following response," J. Am. Acad. Audiol **13**, 188–204.

Lins, O. G., and Picton, T. W. (**1995**). "Auditory steady-state responses to multiple simultaneous stimuli," Electroencephalogr. Clin. Neurophysiol. **96**, 420–432.

Luts, H., Desloovere, C., and Wouters, J. (**2006**). "Clinical application of dichotic multiple-stimulus auditory steady-state responses in high-risk newborns and young children," Audiol. Neuro-Otol. **11**, 24–37.

Luts, H., Van Dun, B., Alaerts, J., and Wouters, J. (**2008**). "The influence of the detection paradigm in recording auditory steady-state responses," Ear Hear. **29**, 638–650.

Luts, H., and Wouters, J. (**2004**). "Hearing assessment by recording multiple auditory steady-state responses: The influence of test duration," Int. J. Audiol. **43**, 471–478.

Luts, H., and Wouters, J. (**2005**). "Comparison of MASTER and AUDERA for measurement of auditory steady-state responses," Int. J. Audiol. **44**, 244–253.

Malmivuo, J., and Plonsey, R. (**1995**). *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields* (Oxford University Press, New York).

Picton, T. W., Durieux-Smith, A., Champagne, S. C., Whittingham, J., Moran, L. M., Giguere, C., and Beauregard, Y. (**1998**). "Objective evaluation of aided thresholds using auditory steady-state responses," J. Am. Acad. Audiol **9**, 315–331.

Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (**2003**). "Human auditory steady-state responses," Int. J. Audiol. **42**, 177–219.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., and Maiste, A. C. (**1987**). "Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone," J. Acoust. Soc. Am. **82**, 165–178.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1992**). *Confidence Limits on Estimated Model Parameters in Numerical Recipes in C* (Cambridge University Press, Cambridge), Chap. 15.6, pp. 689–698.

Purcell, D. W., John, M. S., Schneider, B. A., and Picton, T. W. (**2004**). "Human temporal auditory acuity as assessed by envelope following responses," J. Acoust. Soc. Am. **116**, 3581–3593.

Riquelme, R., Kuwada, S., Filipovic, B., Hartung, K., and Leonard, G. (**2006**). "Optimizing the stimuli to evoke the amplitude modulation following response (AMFR) in neonates," Ear Hear. **27**, 104–119.

Saberi, K., and Perrott, D. R. (**1999**). "Cognitive restoration of reversed speech," Nature (London) **398**, 760.

Scharf, L. L. (**1991**). *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*, 1st ed. (Addison-Wesley, 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.
Simmer, K. U., Bitzer, J., and Marro, C. (**2001**). *Post-Filtering Techniques in Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer-Verlag, Berlin), Chap. 3, pp. 39–60.
Small, S. A., and Stapells, D. R. (**2008**). "Normal ipsilateral/contralateral asymmetries in infant multiple auditory steady-state responses to air- and bone-conduction stimuli," Ear Hear. **29**, 185–198.
Stürzebecher, E., Cebulla, M., Elberling, C., and Berger, T. (**2006**). "New efficient stimuli for evoking frequency-specific auditory steady-state responses," J. Am. Acad. Audiol **17**, 448–461.
Stürzebecher, E., Cebulla, M., and Pschirrer, U. (**2001**). "Efficient stimuli for recording of the amplitude modulation following response," Audiology **40**, 63–68.
Valdes, J. L., Perez-Abalo, M. C., Martin, V., Savio, G., Sierra, C., Rodriguez, E., and Lins, O. (**1997**). "Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses," Ear Hear. **18**, 420–429.
van der Reijden, C. S., Mens, L. H. M., and Snik, A. F. M. (**2001**). "Comparing signal-to-noise ratios of amplitude modulation following responses from four EEG derivations in awake normally hearing adults," Audiology **40**, 202–207.
van der Reijden, C. S., Mens, L. H. M., and Snik, A. F. M. (**2004**). "Signal-to-noise ratios of the auditory steady-state response from fifty-five EEG derivations in adults," J. Am. Acad. Audiol **15**, 692–701.
van der Reijden, C. S., Mens, L. H. M., and Snik, A. F. M. (**2005**). "EEG derivations providing auditory steady-state responses with high signal-to-noise ratios in infants," Ear Hear. **26**, 299–309.
Van Dun, B., Rombouts, G., Wouters, J., and Moonen, M. (**2009**). "A procedural framework for auditory steady-state response detection," IEEE Trans. Biomed. Eng. **56**, 1098–1107.
Van Dun, B., Verstraeten, S., Alaerts, J., Luts, H., Moonen, M., and Wouters, J. (**2008**). "A flexible research platform for multi-channel auditory steady-state response measurements," J. Neurosci. Methods **169**, 239–248.
Van Dun, B., Wouters, J., and Moonen, M. (**2007a**). "Improving auditory steady-state response detection using independent component analysis on multi-channel EEG data," IEEE Trans. Biomed. Eng. **54**, 1220–1230.
Van Dun, B., Wouters, J., and Moonen, M. (**2007b**). "Multi-channel Wiener filtering based auditory steady-state response detection," in Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP) Honolulu, HI, Vol. **II**, pp. 929–932.
van Wieringen, A., and Wouters, J. (**2008**). "LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands," Int. J. Audiol. **47**, 348–355.

Masking release for words in amplitude-modulated noise as a function of modulation rate and task

Emily Buss,[a] Lisa N. Whittle, John H. Grose, and Joseph W. Hall III
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 8 May 2008; revised 14 April 2009; accepted 15 April 2009)

For normal-hearing listeners, masked speech recognition can improve with the introduction of masker amplitude modulation. The present experiments tested the hypothesis that this masking release is due in part to an interaction between the temporal distribution of cues necessary to perform the task and the probability of those cues temporally coinciding with masker modulation minima. Stimuli were monosyllabic words masked by speech-shaped noise, and masker modulation was introduced via multiplication with a raised sinusoid of 2.5–40 Hz. Tasks included detection, three-alternative forced-choice identification, and open-set identification. Overall, there was more masking release associated with the closed than the open-set tasks. The best rate of modulation also differed as a function of task; whereas low modulation rates were associated with best performance for the detection and three-alternative identification tasks, performance improved with modulation rate in the open-set task. This task-by-rate interaction was also observed when amplitude-modulated speech was presented in a steady masker, and for low- and high-pass filtered speech presented in modulated noise. These results were interpreted as showing that the optimal rate of amplitude modulation depends on the temporal distribution of speech cues and the information required to perform a particular task. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3129506]

PACS number(s): 43.66.Dc, 43.71.Es [KWG] Pages: 269–280

I. INTRODUCTION

In normal-hearing listeners, masking of a speech signal presented in broadband noise can be reduced by the introduction of masker amplitude modulation (AM). It is widely believed that this masking release can be accounted for in terms of the reduced masker levels associated with modulation minima, providing the listener with brief "glimpses" of the signal at an improved signal-to-noise ratio (SNR) (Miller and Licklider, 1950; Dirks and Bower, 1970; Howard-Jones and Rosen, 1993). This explanation is analogous to looking through a picket fence—an observer can see through the gaps between slats in the fence, and that is often sufficient to build up an accurate impression of the scene behind that fence (Miller and Licklider, 1950). Similarly, with an amplitude-modulated masker, the listener hears brief portions of the signal in the masker minima, and under some conditions that provides enough information to decipher the speech signal.

The rate of AM has an effect on the magnitude of masking release; typically the biggest effects have been reported for relatively slow rates, in the vicinity of 10 Hz or lower (Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Bacon *et al.*, 1998). There is also some evidence that the optimal rate of AM may differ across speech materials. For example, the optimal rate for spondee words was found to be 1 Hz, lower than the optimal rate for other two-syllable words and monosyllabic words (Dirks *et al.*, 1969; Dirks and Bower, 1971). This result was interpreted as reflecting the increased redundancy of spondee words. A slow rate of masker AM provides temporally sparse glimpses of the target speech, but each of those glimpses is of relatively high quality. Because forward masking decays at an approximately constant rate as a function of time (Plomp, 1964), the longer modulation minima associated with lower rates of AM result in less forward masking overall. Because of this reduction in non-simultaneous masking, speech cues coincident with long-duration modulation minima are encoded with greater fidelity. For redundant speech materials, these sparsely distributed high-quality glimpses may be sufficient to identify the target word, whereas less redundant material might require more temporally dispersed glimpses to support identification. The results of Dirks and his colleagues (Dirks *et al.*, 1969; Dirks and Bower, 1971) are consistent with an interaction between modulation rate and cue redundancy. However, the target speech materials used in those studies differed across the redundancy conditions, leaving open the possibility that other factors such as word frequency or acoustic differences across stimulus sets played a role in the pattern of masking release.

One way to influence the redundancy of cues sufficient to perform a speech task is to manipulate the context in which the target material is presented. In an open-set sentence recognition task, the semantic context in which a word is presented can reduce the acoustic cues necessary to identify that word. The speech in noise test employs this approach, with the phrase preceding a target word either strictly limiting the semantically plausible set of final words (high predictability) or allowing for a wide range of final words (low predictability). One drawback to this approach is that it is difficult to precisely quantify the effect of context, with

[a]Author to whom correspondence should be addressed. Electronic mail: ebuss@med.unc.edu

both linguistic and subjective factors playing a role in predictability of the target word. It is also possible to manipulate redundancy in a closed-set task by changing the size of the response set, a procedure which lends itself more easily to parametric manipulations. Miller *et al.* (1951) showed that identification of a target word changes in a comparable way whether cue redundancy is manipulated through semantic context or response set size. The current study used a set-size manipulation of cue redundancy to assess the role of masker AM rate, an approach which has the advantage that a single set of speech stimuli can be used across redundancy conditions.

The current experiments examined the role of cue redundancy in the masking release associated with masker AM in a population of normal-hearing adults. The first experiment estimated masking release as a function of masker AM rate for three tasks: detection, three-alternative forced-choice identification (3AFC-ID), and open-set identification (open-ID). The detection task can be performed based on a very coarse cue, such as an increment in stimulus energy at any frequency associated with speech. The 3AFC-ID task requires more detailed information, such as a phoneme or a stimulus feature that distinguishes the alternatives; in this case the information required to arrive at the correct response could be quite limited, perhaps even based on a single glimpse in time. In contrast, the open-ID requires a relatively complex set of cues, including multiple phonemes distributed over time. Because the masking release associated with masker AM is thought to be limited by temporal resolution, longer duration glimpses associated with lower modulation rates should provide higher quality acoustic information and support accurate performance to the extent that sparsely distributed cues support good performance of the task. This reasoning leads to the hypothesis that the optimal rate of masker AM depends on the task, with best performance for detection being associated with a single glimpse of high quality (i.e., low rate AM) and best performance for the open set requiring a larger number of more widely dispersed glimpses (i.e., high rate AM). Such a result would lend further support to the conclusions of Dirks and Bower (1971), who showed analogous effects using stimuli with inherently high cue redundancy (spondees) or low cue redundancy (non-spondee, two-syllable words).

II. EXPERIMENT 1

A. Methods

1. Observers

Five observers participated in this study (two females) ranging in age from 19.7 to 32.8 years (mean 23.8 years). All observers had pure tone thresholds of 20 dB hearing level (HL) or better at octave frequencies from 250 to 8000 Hz in the test ear (ANSI, 1996), and none reported a history of ear disease. Non-native English speakers were excluded from participation, and all observers spoke with an American accent.

2. Stimuli

Target speech material was a set of 500 consonant-nucleus-consonant (CNC) words (Peterson and Lehiste, 1962), spoken by an adult male with an American accent. These recordings were 444–992 ms, with a mean duration of 744 ms. The sampling rate was 24.4 kHz, and all signals were passed through an 8-kHz second order Butterworth low-pass (LP) filter. Recordings were digitally scaled to equal-rms level across tokens.

The masker was a Gaussian noise that was spectrally shaped to the long-term spectrum of the speech stimuli, referred to as speech-shaped noise. In some conditions the maskers were amplitude modulated via multiplication with a raised sinusoid, with modulation rates of 2.5, 5, 10, 20, or 40 Hz; in these conditions the peak masker level was 75-dB sound pressure level (SPL), for an overall level of 70.8 dB SPL. There were two steady masker conditions. In the *equal-peak* steady masker condition, the masker was presented at 75-dB SPL, matching the peak level of the AM masker. In the *equal-rms* conditions the masker was presented at 70.8 dB SPL, matching the overall level of the AM masker.

3. Procedures

There were three tasks. In the *detection* task there were three listening intervals, visually indicated by lights on a hand-held response box. Listening intervals were 1 s in duration and separated by 500 ms. The observer was asked to select the interval in which a CNC word was presented; a randomly selected speech token was equally likely to be presented in each listening interval, and token onset coincided with the onset of the listening interval. In the 3AFC-ID task the observer was presented with a randomly selected word and then asked to identify the word from three alternatives presented visually after the listening interval, the foils being selected randomly from the remaining 499 tokens. In the open-ID condition the observer was presented with a randomly selected word and asked to repeat that word aloud; at that point the observer was visually presented with the correct response and prompted to score his response as correct or incorrect using buttons displayed on the computer screen.[1] An experimenter monitored each experimental session, including spot checks for correct self-scoring in the open-ID task; in no case did the experimenter have to re-instruct an observer in any of these procedures.

In all conditions the masker was presented continuously. The signal level was adjusted following a one-down, one-up tracking rule estimating 50% correct (Levitt, 1971). At the beginning of a track, signal level was adjusted in 4-dB steps; that stepsize was reduced to 2 dB after the second reversal. Each track continued until a total of 12 reversals had been obtained, and the threshold estimate associated with a track was the average signal level at the last 10 track reversals.

During the first testing session data were collected in one of the two speech identification tasks, selected at random (either 3AFC-ID or open-ID). The second session consisted of data collection for the alternate speech identification task. The detection conditions were completed in the third and final session for all observers. In each of the three sessions,

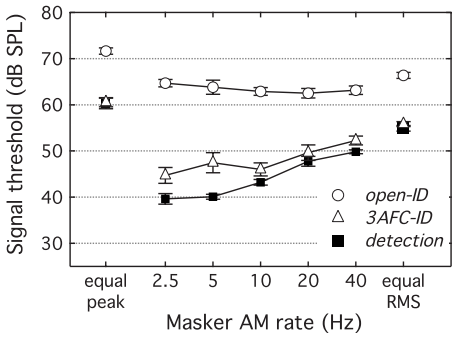


FIG. 1. The mean threshold across observers is plotted as a function of condition, with error bars indicating ± 1 standard error of the mean. Symbols indicate the task: detection (■), 3AFC-ID (△), or open-ID (○). Thresholds for the equal-peak steady masker condition are plotted at the far left and those for the equal-rms condition at the far right of the panel.

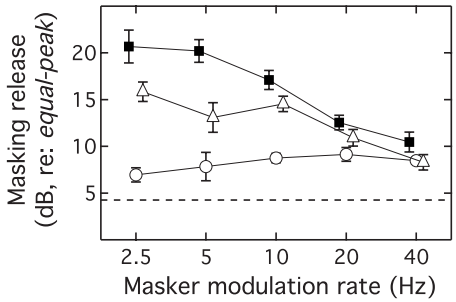


FIG. 2. Masking release is plotted as a function of masker AM rate relative to the threshold obtained in the equal-peak steady masker condition associated with each task. Symbols indicate the three response conditions as in Fig. 1: detection (■), 3AFC-ID (△), or open-ID (○). Error bars indicate ±1 standard error around the mean, and the horizontal dashed line shows the 4.2 dB improvement expected due to the reduction in masker level associated with AM. Data points are slightly offset on the abscissa to aid visual inspection.

observers completed one threshold estimate in each of the seven masker conditions in random order, including five modulation rates and two levels of steady masker. A second estimate was then obtained in all seven conditions in a new random order. As time allowed, a third estimate was obtained in conditions for which the previous two estimates varied widely, at the discretion of the experimenter on a case-by-case basis. All estimates obtained in a condition were averaged to produce a final threshold estimate for each observer. Each listening session lasted for 1 h.

B. Results

Figure 1 shows mean thresholds, averaged across the five observers, plotted as a function of masker condition. Symbols reflect the task, as indicated in the figure legend. These thresholds indicate the signal level required to achieve 50% correct in a fixed-level masker, so low thresholds represent good performance. Results in the baseline, steady masker conditions will be considered first. For the equal-peak masker conditions, thresholds are very similar for the detection and the 3AFC-ID tasks, with means of 60.3 and 60.6 dB SPL, respectively. Threshold for the open-ID task is higher, with a mean of 71.6 dB SPL. Thresholds in the equal-rms condition are on average 5.1 dB lower than those in the associated equal-peak conditions. This difference is consistent with the 4.2-dB reduction in level. A set of three t-tests was performed, comparing the difference in steady masker thresholds within observer for each task; in no case was the mean difference significantly different from 4.2 dB ($p \geq 0.36$, two-tailed).

Masker AM tended to reduce thresholds relative to both the equal-peak and equal-rms conditions, but this effect was dependent on both AM rate and task. The effect of rate was broadly similar for the detection and 3AFC-ID tasks, with best thresholds for lower rates of masker AM and elevation with increasing modulation rate. There is some indication that the trend for poorer performance with increasing AM rates begins at a lower rate for the detection than the 3AFC-ID task, with knee points at 5 and 10 Hz, respectively. The effect of masker AM rate on these two conditions was large, with thresholds for the 3AFC-ID task spanning a range of 7.5 dB, and those for the detection task varying by 10.2 dB. In contrast, best performance for the open-ID task was obtained at the highest AM rates of 10–40 Hz, with poorer thresholds for lower AM rates. The effect of AM rate on open-ID thresholds was more modest than in the other conditions, with mean thresholds spanning a range of just 2.2 dB.

To assess this pattern of results thresholds were submitted to a repeated-measures analysis of variance (ANOVA), with five levels of AM RATE (2.5, 5, 10, 20, and 40 Hz) and three levels of TASK (detection, 3AFC-ID, open-ID). There were significant main effects of RATE ($F_{4,16}=14.0$, $p<0.0001$) and TASK ($F_{2,8}=218.5$, $p<0.0001$), and there was a significant interaction ($F_{8,32}=13.5$, $p<0.0001$). A preplanned linear contrast indicated a significant interaction with AM rate for the open-ID and 3AFC-ID tasks ($F_{1,4}=69.3$, $p<0.001$), consistent with the task-by-rate interaction described above. The interaction with AM rate for the 3AFC-ID and detection tasks just failed to reach significance ($F_{1,6}=7.1$, $p=0.056$).

Because the change in open-ID thresholds with AM rate was modest, it was of interest to determine whether the significant interaction between 3AFC-ID and open-ID tasks was due solely to variability in the degree to which increasing rates of AM increased 3AFC-ID thresholds, or whether the open-ID thresholds actually improved as a function of AM rate. The significance of the decrease in open-ID thresholds as a function of AM rate was assessed in two stages. First, thresholds were fitted with a linear regression, with four dummy variables coding for observer. This analysis resulted in a significant effect of observer ($F_{4,20}=3.82$, $p<0.05$). A correlation between the residuals from this analysis and the logarithm of modulation rate was computed. This second analysis resulted in a significant linear association between masker AM rate and open-ID thresholds ($r=-0.36$, $p<0.05$, one-tailed). While this fit accounted for only 13% of the variance in the data, it is consistent with an improvement in thresholds with increasing AM rate for this task.

The masking release associated with masker AM is shown in Fig. 2. The reduction in threshold relative to the equal-peak masker is plotted as a function of rate, with symbols following the same convention as in Fig. 1. The dashed

horizontal line indicates the reduction in masker level associated with AM (4.2 dB), and hence the improvement in threshold that would be expected if reductions in overall masker level were responsible for masking release. Values above this line reflect masking release associated with the transient improvements in SNR as a result of masker AM. Consistent with observations of absolute thresholds, above, these values of masking release show an interaction between masker AM rate and task; there is more masking release for low AM rates under conditions of high cue redundancy (detection and 3AFC-ID) and better performance at high rates under conditions of reduced redundancy (open-ID). This depiction of the data also highlights the fact that whereas there was a masking release in the mean data in all conditions, there was a nearly 15-dB difference in the masking release obtained across tasks with 2.5 Hz AM, and a relatively consistent masking release at 40-Hz. This result is consistent with the interpretation that performances in all three conditions benefit approximately equally for temporally dispersed brief modulation minima, but that the temporally sparse high-quality cues associated with low rates of AM are much more beneficial in high-redundancy (detection and 3AFC-ID) as compared to low-redundancy (open-ID) tasks.

C. Discussion

The results of experiment 1 support the hypothesis that the peak masking release associated with masker AM occurs at different rates of AM in tasks requiring differing levels of signal detail. Thresholds in tasks that can be performed based on sparse or coarse cues, in this case word detection or 3AFC-ID, are lowest for relatively low rates of AM, whereas thresholds in the open-set task, which requires more detailed information, are lowest at higher rates of AM.

1. Effects of task

Three speech tasks (detection, 3AFC-ID, and open-ID) were employed in order to manipulate the speech cues necessary to perform the task; it was argued that detection requires the sparsest cues, closed-set identification requires relatively minimal encoding of the speech signal, and open-ID requires relatively detailed encoding of multiple phonemes. Performance in these three tasks provides support for this ordering of task difficulty. Averaging across all masker conditions, performance is rank ordered following this presumed hierarchy of difficulty, with mean thresholds of 65.0 dB (open-ID), 50.9 dB (3AFC-ID), and 48.0 dB (detection). One caveat to this ranking of task difficulty is that the results of the detection and 3AFC-ID conditions were more similar to each other than to the open-ID condition, as reflected in the mean across all masker conditions and in the nearly identical steady masker thresholds in the detection and 3AFC-ID conditions.

Performance in the open-ID task in steady noise is consistent with published results. For example, Studebaker *et al.* (1994) measured percent correct for CID W-22 words as a function of SNR and found that a SNR of −5.5 dB was associated with 50% correct in speech-shaped noise, comparable to the average threshold of −3.9 dB SNR found for the open-ID steady masker conditions of the present experiment. It is commonly observed that speech recognition requires a higher SNR than detection of speech. In one demonstration of this effect Hawkins and Stevens (1950) measured thresholds for running discourse and found a 10-dB difference in detection as compared to recognition. This effect size is comparable to the approximately 10-dB threshold difference observed in the present open-ID and detection tasks. While these results are consistent, interpretation of this parallel is complicated by ambiguity in quantifying the relative context effects in the two paradigms, and, in particular, whether the open-ID of the present experiment is comparable the running discourse used in the Hawkins and Stevens (1950) study.

Differences between the open-ID and 3AFC-ID identification conditions are consistent with previous observations that words are easier to understand in the context of a sentence and that high-predictability sentences are recognized more accurately than low-predictability sentences (Miller *et al.*, 1951; Kalikow *et al.*, 1977). For example, Miller *et al.* (1951) reported percent correct for monosyllabic word recognition at a range of SNRs for open set and closed set, with a wide range of response options; these results indicate an ~18 dB difference in the 50% correct point between open-set and four-alternative forced-choice conditions (with SNRs of +4 and −14 dB, respectively), very close to the equal-peak masker results in the present data set for the open-ID and 3AFC-ID tasks (with SNRs of −3.4 and −14.4 dB, respectively).

The effect of context on recognition is sometimes described in terms of the "linguistic entropy" of the speech sample (van Rooij and Plomp, 1991; Bronkhorst *et al.*, 1993). When entropy is high, the listener has very little information with which to narrow the range of possible targets, but when entropy is low the listener can use the information present in the surrounding segment of speech to help interpret the sensory signal; entropy is the inverse of redundancy. The effects of linguistic entropy have been modeled in terms of variable levels of cognitive noise (Müsch and Buus, 2001). One way to think about the performance advantage associated with low linguistic entropy is in terms of a template-matching algorithm. When the pool of templates is small the odds of identifying the correct template are relatively good, whereas a very large pool of templates introduces more opportunities for error.

2. Effects related to modulation rate

The most striking finding related to masker modulation rate was that whereas there was a trend for better performance at lower modulation rates in the 3AFC-ID and detection tasks, the trend in the open-ID task was for better performance at the higher rates of modulation. This is also broadly consistent with the findings of Dirks and his colleagues. Dirks *et al.* (1969) reported that masking release for spondee words and sentences was greater for a masker AM rate of 1 Hz than 10 Hz. The opposite was observed for monosyllabic words, where a masker AM rate of 10 Hz was associated with greater masking release than a rate of 1 Hz. This rate effect was interpreted in terms of the minimal cues necessary to correctly identify a speech token given semantic

constraints inherent to the speech materials. In the present paradigm, the level of sensory detail necessary to correctly identify the speech token was manipulated by task, with 3AFC-ID requiring minimal cues (similar to high-redundancy/restrictive context materials) and open-ID requiring more complex cues (similar to low-redundancy/minimal context materials). In the present data set, the pattern of masking release as a function of AM rate is very similar for the 3AFC-ID and detection tasks. There is a nonsignificant trend for better performance at lower rates for the detection task, an effect that would be consistent with the interpretation of the effect of AM rate across the two identification tasks.

3. Effects of forward masking

Whereas the results are consistent with an interpretation of the AM rate effect in terms of the temporal distribution of glimpses of the speech, this account may be complicated by the fact that the SNR at threshold differs substantially across tasks. For example, at a 20-Hz rate of masker AM, threshold in the open-ID condition is 62.5 dB, whereas those in the 3AFC-ID and detection conditions are 49.6 and 47.8 dB, respectively. In poor SNR conditions, the speech signal is likely to be audible only during the temporal center of a masker modulation minimum, whereas in the higher SNR conditions, the speech signal may be audible for a larger proportion of the modulation period. As a consequence, the glimpses of speech associated with high rate AM could be effectively briefer at low as compared to high SNRs. If that is the case, then forward masking could play a larger role in performance of detection and 3AFC-ID tasks as compared to open-ID at high masker AM rates.

Supplemental data were collected in order to test the possible role of forward masking in the effect of AM rate. Stimuli were identical to those in the main experiment except that AM was applied to the speech signal instead of to the masker. In these conditions the masker was a 75-dB SPL speech-shaped noise, and the signal was a CNC word that was sinusoidally amplitude modulated at either 5 or 20 Hz. Signal level was adjusted adaptively to estimate threshold for 50% correct response, as described above, in both the 3AFC-ID or open-ID task. Thresholds were collected in random order, with a total of three to four estimates in each of four conditions (2 rates $\times$ 2 tasks). Seven normal-hearing observers were recruited to complete the supplemental conditions, all meeting the inclusion criteria noted above for the primary study. Observers ranged in age from 21.0 to 40.5 years (mean 29.5 years), and all had previously participated in a speech study using CNC words, including experiments 1 and 2 described in the present report, as well as very similar pilot experiments.

Results of these supplemental conditions are shown in Fig. 3, with mean threshold plotted as a function of modulation rate and error bars indicating one standard deviation. Symbol shape reflects task, consistent with Fig. 1. As in the AM-masker conditions, thresholds are higher in the open-ID than in the 3AFC-ID conditions, with a mean difference of 15.9 dB. Thresholds also differ for the two signal modulation rates, with better performance for 5-Hz AM in the 3AFC-ID condition and 20-Hz AM in the open-ID condition. A repeated-measures ANOVA was performed to assess the significance of this interaction. There were two levels of TASK (3AFC-ID and open-ID) and two levels of RATE (5 and 20 Hz). There was a main effect of TASK ($F_{1,6}=292.49$, $p<0.0001$), a main effect of RATE ($F_{1,6}=8.19$, $p<0.05$), and a significant interaction ($F_{1,6}=40.07$, $p<0.001$). Paired t-tests were performed to assess whether this interaction is due to rate effects on one or both of the tasks. These analyses confirmed that thresholds at the two signal AM rates were significantly different (one-tailed) for both the open-ID ($t_6=8.00$, $p<0.001$) and 3AFC-ID ($t_6=-2.68$, $p<0.05$) tasks.

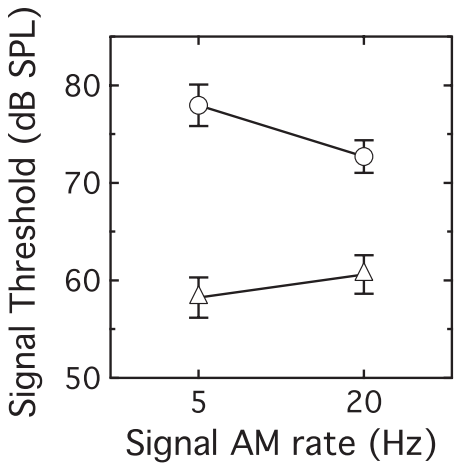


FIG. 3. Word identification thresholds are plotted as a function of signal AM rate. Symbols reflect the response condition, following the conventions of Fig. 1: 3AFC-ID (△) or open-ID (○). Error bars indicate ±1 standard deviation.

Because the speech rather than the masker was modulated in the supplemental conditions, forward masking would not be expected to play a large role in the results. The finding of an interaction between modulation rate and task lends support to the idea that the distribution of information required to perform each of the speech tasks is responsible for the AM-masker rate effects obtained in the main experiment. Previous work on amplitude-modulated speech, sometimes described as interrupted speech, has uncovered a nonmonotonic relationship between modulation rate and performance that is dependent on both modulation duty cycle and speaking rate (Huggins, 1964; Powers and Wilcox, 1977). These results have largely been explained in terms of the temporal distribution of cues required to support recognition. For example, if the "off" portion of the modulation period is long relative to the duration of a word, then some words will be missed in an open-set task. A similar explanation appears to be valid for the present data. A slow modulation period is associated with infrequent but relatively long-duration glimpses, sufficient to perform the 3AFC-ID task. Those cues are too temporally sparse to perform the open-ID task, however, where better performance is obtained when glimpses are more widely distributed over time.

The finding of better performance in the 3AFC-ID task at 5 than 20 Hz implies that the short glimpses associated with 20-Hz AM are less effective than those at 5 Hz even in the absence of forward masking. One factor that could un-

derlie better performance at the lower AM rates in the 3AFC-ID condition is the detrimental effect of sidebands associated with signal modulation. This suggestion is analogous to the switching artifact that has been proposed to limit performance with interrupted speech at high rates of interruption (Huggins, 1964).

4. The role of frequency region

It has been argued that whereas vowels are more perceptually salient, with higher mean energy content, consonants are more important in word identification (Bonatti *et al.*, 2005; Toro *et al.*, 2008). If that is the case, then it is also possible that detection and 3AFC-ID could be performed based on acoustically salient vowel information, whereas open-ID relies more on consonants. This possibility gets some support from the finding that across tokens, identification of vowels in speech-shaped noise is better than that for consonants, though there is wide variability across the different types of consonants (Phatak and Allen, 2007). There is also evidence that low- and high-frequency regions contribute differently to the perception of vowels and consonants. For example, recognition of vowels in speech-shaped noise is dominated by low-frequency cues related to F1 and to a lesser extent F2 (Parikh and Loizou, 2005), whereas the spectral cues, which differentiate consonants, are more robust at high than low frequencies (Phatak and Allen, 2007). These observations support the possibility that the available frequency region of speech could play an important role in the pattern of masker AM effects observed in the present experiment.

In normal-hearing listeners it has been suggested that temporal resolution may limit access to low-frequency cues more than those at high frequencies, a hypothesis motivated by several considerations: the possibility of inherently better temporal resolution in high- than low-frequency channels (Festen, 1987; Stuart and Phillips, 1996), better representation of the masker envelope at the output of wider high-frequency channels (Haggard *et al.*, 1990; Bacon *et al.*, 1997), or increased suppression at high frequencies in normal-hearing listeners (Bacon *et al.*, 1997; Lee and Bacon, 1998). In the context of the present paradigm, this line of reasoning would suggest that relatively high-frequency speech information could play a dominant role in the masking release observed with a full-spectrum speech target.

It is well known that masker AM provides greatly reduced benefits for listeners with moderate sensorineural hearing impairment as compared to normal-hearing listeners (e.g., Festen and Plomp, 1990; Gustafsson and Arlinger, 1994). Whereas some previous studies have implicated temporal resolution as a limitation in AM-related masking release, other work suggests that this factor is insufficient to explain the performance of hearing impaired listeners (Jin and Nelson, 2006), implicating reduced spectral resolution or the interaction between reduced spectral resolution and reduced speech redundancy as the dominant factor (Baer and Moore, 1994; Hall *et al.*, 2008). Several studies have shown a diminished benefit of masker AM minima for speech that has been degraded using a vocoder simulation (Kwon and Turner, 2001; Nelson *et al.*, 2003; Qin and Oxenham, 2003; Nelson and Jin, 2004). In some conditions performance worsens with the introduction of masker AM (Kwon and Turner, 2001), an effect attributed to masker modulation interfering with the processing of speech envelope fluctuations. One explanation for the poor intelligibility of vocoded speech in AM noise is that the spectral coarseness of the target fails to provide sufficient cues to segregate it from the masker. Qin and Oxenham (2005) suggested that cues related to voice fundamental frequency are very important to this segregation process, and that poor performance in these conditions is due to the fact that vocoded speech discards fine-structure cues to F0; this hypothesis is bolstered by the finding that restoring those cues significantly improves performance (Chang *et al.*, 2006; Qin and Oxenham, 2006). Other data suggest that similar effects may limit masking release obtained with introduction of masker AM for unprocessed speech (Lorenzi *et al.*, 2006). These results are consistent with the hypothesis that low-frequency fine-structure cues are critical for stream segregation, without which envelope patterns associated with masker AM could exert modulation masking that interferes with the use of envelope-based speech cues. By this reasoning, it is possible that the masking release trends observed in experiment 1 might have arisen largely from processes related to relatively low-frequency speech information.

A third and final possibility regarding the importance of different frequency regions of speech to masking release is that the release observed in experiment 1 is inextricably related to relatively *wideband* processes, and therefore depends critically on the synthesis of speech information across low and high spectral regions. This might be the case if spectral redundancy of speech cues were a precondition for AM-related masking release. This possibility is in accord with previous reports of greater masking release under conditions of high speech cue redundancy (Dirks *et al.*, 1969; Kwon and Turner, 2001), as well as the finding that masking release is reduced to a comparable degree for LP and high-pass (HP) speech stimuli when baseline performance in steady noise is comparable across filter conditions (Oxenham and Simonson, 2008). Fullgrabe *et al.* (2006) recently showed that different speech cues are perceived best at different rates of AM. The availability of a wide range of cues could increase the chances of correct identification in a masker characterized by a single rate of AM. Conversely, a spectrally impoverished cue set could severely restrict the speech information available at any single masker AM rate.

In summary, the discussion above highlights the importance of considering how the masking release effects such as those found in the first experiment may be related to absolute frequency region. The relative contribution of low- and high-frequency speech information to masking release was explored further in the second experiment.

III. EXPERIMENT 2

In experiment 1 it was shown that the masking release associated with masker AM varied as a function of AM rate, with different patterns of masking release for different speech tasks. These results were discussed in terms of audi-

bility, by virtue of unmasking of the cues required to perform each task as a function of time via masker AM. One potentially important factor in that paradigm was the relative contribution of low- versus high-frequency cues. Experiment 2 further assessed masking release for a LP and a HP filtered stimulus. If the effects observed in experiment 1 were driven solely by effects related to instantaneous signal-to-masker ratio, then the masking release due to masker AM should be very similar for LP and HP filtered speech in a speech-shaped noise masker under matched conditions, to the extent that the temporal distribution of cues is similar in these two frequency regions. If, on the other hand, the encoding of low- and high-frequency speech information is qualitatively different, then masking release could differ substantially across spectral regions.

One caveat that should be considered with respect to the present experiment is that the interaction of task and masker AM rate may rely on spectral as well as temporal redundancies. Up to this point the good performance at low rates of masker AM in the 3AFC-ID task has been discussed in terms of the temporal redundancy of cues, such that the cues available during a single, relatively long-duration glimpse could support correct identification, regardless of when that glimpse occurs in the word. Such good performance could also rely on the spectral redundancy of speech, however, as suggested by Oxenham and Simonson (2008). Reducing spectral redundancy by LP or HP filtering the stimulus, as in the present paradigm, could reduce the quality of each glimpse, such that low rates of masker AM no longer support good performance. Based on this reasoning, changes in the pattern of task-by-AM rate effects might be predicted for the present experiment.

A. Methods

1. Observers

A total of 17 observers participated in this experiment (15 females) ranging in age from 18.5 to 47.8 years (mean 27.1 years). All observers had pure tone thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz in the test ear (ANSI, 1996), and none reported a history of ear disease. Non-native English speakers were excluded from participation, and all listeners spoke with an American accent. Some observers had previously participated in psychoacoustic studies, but none using CNC speech materials.

2. Stimuli

As in experiment 1, testing involved either detection, 3AFC-ID, or open-ID for CNC words. All testing was performed in the presence of a continuous speech-shaped noise masker. Masker AM, when present, was achieved via multiplication with a raised sinusoid at a modulation rate of 2.5, 5, 10, 20, or 40 Hz. Steady masker conditions included equal-peak and equal-rms comparison conditions, associated with the same peak or the same rms level as the comparable AM-masker conditions. In contrast to experiment 1, both the signal and masker were passed through a LP or a HP filter. Filtering was achieved by passing the stimuli through a fourth order Butterworth filter twice, once forward and once backward, with a 1700-Hz cut-off frequency. This cut-off frequency was selected based on pilot listening in which this cutoff resulted in comparable open-set thresholds for both LP and HP filtered speech; a steady masker was used for these pilot conditions. This value is also consistent with the observation that the centroid of speech information is typically cited as falling between 1000 and 2000 Hz (Studebaker *et al.*, 1987; DePaolis *et al.*, 1996; Henry *et al.*, 1998). In the LP and HP conditions the masker was 75 or 70.8-dB SPL prior to filtering; HP filtering in the HP condition reduced the overall masker level by 16 dB. In order to assess the significance of this level reduction in the pattern of results, an additional HP filter condition was included, wherein the masker level was increased by 16 dB, for a level of 75 or 70.8-dB SPL after filtering. This condition is referred to here as HP+16.

3. Procedures

Observers were randomly assigned to the LP, HP, or HP+16 conditions. Following the procedures of experiment 1, each observer began with one of the two identification conditions (either open-ID or 3AFC-ID), completing thresholds in random order. The second session was spent on the alternate identification condition, and all observers completed detection conditions in the third and final listening session. In all cases signal level associated with 50% correct was estimating using a one-down, one-up track. Initial level adjustments were made in steps of 4 dB; this stepsize was reduced to 2 dB after the second track reversal. Tracks continued for 12 reversals, and the average signal level at the last 10 reversals was used as an estimate of threshold. Between two and three estimates were obtained in each condition, and the means of all thresholds obtained are reported below.

B. Results

The mean SNRs at threshold for the equal-peak steady masker conditions are reported in Table I; the standard error of the mean appears to the right of each value, and comparable results from the full-spectrum conditions of experiment 1 appear in the top row of the table for comparison. Thresholds were relatively consistent across the three filter conditions (LP, HP, and HP+16). As in experiment 1, thresholds tended to fall in rank order, from open-ID, to 3AFC-ID, to detection, and thresholds in the final two tasks were relatively similar. The SNR at threshold for the open-ID task differed by about 5 dB between the results of experiments 1 and 2, indicating a detrimental effect of LP and HP filtering on masked open-set recognition. The significance of this difference in the open-ID performance between full-spectrum and filter conditions was evaluated using a set of three t-tests, one for each of the filter conditions; all three were significant (α=0.05, two-tailed, and Bonferroni correction). In contrast, filtering had little or no effect on performance in the 3AFC-ID and detection conditions, for which SNR at threshold was relatively consistent across the two experiments and across filter conditions within experiment 2. Thresholds in the equal-rms condition (not shown) were on average 3.8 dB

TABLE I. Mean thresholds (SNR, in dB) for the equal-peak steady masker condition as a function of target filter conditions, masker level, and task. The standard error of the mean across thresholds ($n=5$ or $n=6$) appears in parentheses to the right of each estimate.

	Task		
Filter condition	Open-ID	3AFC-ID	Detection
Experiment 1, full-spectrum	−3.37 (0.70)	−14.44 (0.98)	−14.69 (1.13)
Experiment 2, low-pass (LP)	2.13 (1.03)	−14.56 (0.80)	−14.80 (0.57)
Experiment 2, high-pass (HP)	2.95 (1.07)	−13.73 (0.80)	−15.63 (0.93)
Experiment 2, high-pass (HP+16)	3.44 (1.01)	−14.54 (0.98)	−16.83 (0.63)

lower than those in the associated equal-peak conditions, with a standard error of 0.43 dB. This is consistent with the 4.2-dB reduced masker level.

The consistency of performance in the LP and HP filter conditions for the steady masker of the present experiment was assessed using a mixed model ANOVA. The mean threshold in the two steady masker conditions (equal-peak and equal-rms) was used in this analysis since the difference in masker level across these steady masker conditions did not significantly affect the SNR at threshold. There was one across-subject factor of COND (LP, HP, and HP+16) and one within-subject factor of TASK (open-ID, 3AFC-ID, and detection). The results of this analysis indicated a significant main effect of TASK ($F_{2,18}=518.88$, $p<0.0001$), but no main effect of COND ($F_{1,9}=1.14$, $p=0.31$) and no interaction ($F_{1,18}=1.28$, $p=0.30$). These results confirm that baseline performance is comparable across the three filter conditions, with comparable performance above and below the 1700-Hz filter cutoff.

Masking release was computed as the difference between thresholds measured with an AM masker and those from the associated equal-peak steady masker conditions. Results appear in Fig. 4, with masking release plotted as a function of modulation rate and error bars indicating one standard error of the mean. Masking release in all three filter conditions exhibited some common features: masking release associated with open-ID conditions was modest and that for detection was largest, with 3AFC-ID usually falling intermediate between these two. In all cases masking release varied across task more for low rates than for high rates of masker AM, where values tend to converge. These shared features were also noted in the results obtained with full-spectrum speech in experiment 1 (Fig. 2).

Despite these general commonalities, some aspects of the LP and HP filter data differed from the full-spectrum data of experiment 1, most notably a trend for smaller magnitudes of masking release in some conditions in the LP and HP conditions: averaging across all conditions, the mean masking release in experiment 1 was 12.3 dB, a value that can be compared to 11.2 dB in the LP and 10.6 dB in the HP conditions. In the LP condition the reduction in masking release is subtle and evident primarily in the open-ID conditions: whereas masking release in the open-ID condition of experiment 1 ranged from 6.9 to 9.1 dB, masking release in the LP conditions spanned from 4.8 to 7.9 dB.

The significance of the LP and HP filter on the pattern of masking release was assessed with a pair of mixed model ANOVAs. In each case the masking release obtained with full-spectrum stimuli from experiment 1 was compared with that from either the LP or the HP filter condition of the present experiment. These analyses included an across-subjects factor of COND (full-spectrum and filtered), a within-subjects factor of RATE (2.5, 5, 10, 20, and 40 Hz), and a within-subjects factor of TASK (open-ID, 3AFC-ID, and detection). As in the previous analyses, both of these analyses resulted in a highly significant main effect of RATE and TASK, as well as an interaction ($p<0.0001$). The result of interest here was in terms of the effect of COND. For the LP analysis, there was no main effect of COND ($F_{1,8}=0.58$, $p=0.47$), and none of the interactions with COND approached significance ($p>=0.36$). For the HP analysis, there was a significant main effect of COND ($F_{1,9}=18.38$, $p<0.005$) and a significant interaction between RATE and COND ($F_{4,36}=4.74$, $p<0.005$). These results indicate that the reduction in mean masking release with stimulus filtering

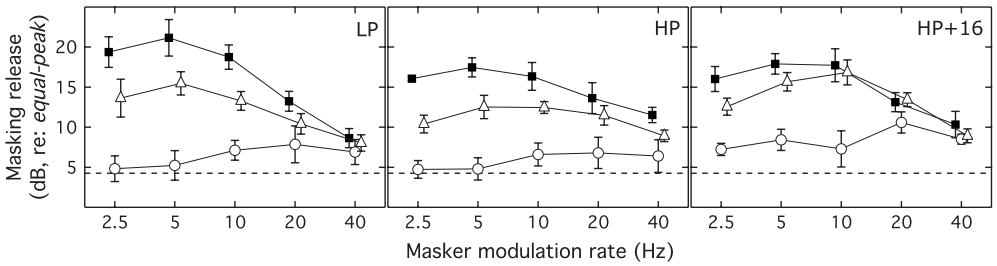


FIG. 4. Masking release is plotted as a function of masker AM rate relative to the threshold obtained in a steady masker with equal-peak level. Each panel shows results of a different filter condition, as indicated in the upper right corner. Symbols indicate the task: detection (■), 3AFC-ID (△), or open-ID (○). Error bars indicate ±1 standard error around the mean, and the horizontal dashed line shows the 4.2 dB improvement expected due to the reduction in masker level associated with AM. Data points are slightly offset on the abscissa to aid visual inspection.

was significant for the HP but not the LP conditions. The significant interaction between RATE and COND for the HP analysis reflects the fact that, compared to masking release in the full-spectrum conditions, masking release for the HP conditions did not vary as widely across tasks for the low rates of AM, and these functions did not converge as closely for the highest rate of masking release (compare Fig. 2 with middle panel of Fig. 4).

It is possible that the reduced masking release and reduced effect of masker AM rate on the HP condition are related to the 16-dB reduction in stimulus level associated with the 1700-Hz HP filter. A third ANOVA was therefore performed to compare the HP+16 and full-spectrum data. Again, there were highly significant main effects of RATE and TASK, as well as a significant interaction ($p<0.0001$). Of particular interest here, there was no main effect of COND ($F_{1,8}=0.05$, $p=0.82$), and none of the interactions with COND reached significance ($p>=0.12$). This result is consistent with the conclusion that the effect of the stimulus filter on the HP condition was mediated to some extent by a level effect rather than the elimination of speech cues below 1700 Hz.

The effect of masker AM rate on masking release in the open-ID condition was modest in the results of experiment 1, with AM rate accounting for only 13% of the variance in those data. Because of the small size of that result and its importance for interpretation of the task-by-AM rate interaction, an analysis was undertaken to assess the effect of masker AM rate on the open-ID task across all conditions of experiments 1 and 2. For this analysis, the differences between equal-peak and AM-masker conditions were computed for individual listeners. The resulting estimates of masking release were submitted to a mixed model ANOVA, with one across-subjects factor of COND (experiment 2: LP, HP, HP +16 and experiment 1) and one within-subjects factor of RATE (2.5, 5, 10, 20, and 40 Hz). This analysis resulted in a significant effect of RATE ($F_{4,68}=2.75$, $p<0.05$), but no significant effect of COND ($F_{3,17}=1.45$, $p=0.26$) and no interaction ($F_{12,68}=0.21$, $p=1.00$). A linear contrast performed on the effect of RATE was significant ($F_{1,17}=12.77$, $p<0.01$), consistent with the visual impression that masking release rose approximately linearly as a function of masker AM rate in the open-ID condition, with no reliable difference across filter conditions.

C. Discussion

Thresholds in the steady masker for both LP and HP filter conditions are statistically indistinguishable, consistent with the idea that the 1700-Hz cutoff corresponds to the centroid of speech cues for these stimuli. The SNR at threshold for the steady masker is nearly constant across filter conditions for a given task. Averaging across filter conditions, these thresholds are −15.3 dB for detection, −14.0 dB for 3AFC-ID, and 2.7 dB for the open-ID task. This pattern of results closely resembles that obtained for the full-spectrum conditions of experiment 1, with the exception of the open-ID task, for which thresholds were reliably about 5 dB higher in the filtered than full-spectrum conditions.

1. The role of frequency region in masking release

Masking release in the HP conditions of experiment 2 tended to be smaller than full-spectrum results from experiment 1, whereas those in the LP conditions were not reduced. In principle, this result could be due to the disproportionate contribution of low-frequency cues to masking release, such as cues based on the use of temporal fine structure. If this were the case, then increasing stimulus level should have little if any effect on this pattern of results. Contrary to that interpretation, masking release in the HP+16 condition was comparable to that obtained in the full-spectrum (experiment 1) and LP (experiment 2) conditions. This result supports the conclusion that masking release with the HP filtered target is smaller than that in comparable LP conditions not because of a failure of high-frequency information to convey the necessary speech cues, but rather due to the lower overall level of the stimulus in that frequency region. This result is consistent with published data indicating significant level effect on the intelligibility speech in AM noise (e.g., Dirks *et al.*, 1969; de Laat and Plomp, 1983; Festen, 1993; Stuart and Phillips, 1997).

The present paradigm bears some resemblance to the paradigm of Scott *et al.* (2001) and Elangovan and Stuart (2005). In those studies, masking release associated with an aperiodic masker was measured for NU6 words. LP filtering was shown to reduce the benefit of masker AM (Scott *et al.*, 2001), whereas HP filtering had a more modest effect (Elangovan and Stuart, 2005). These findings were interpreted as showing the importance of high-frequency channels for masking release, a consequence of better temporal resolution at high than low frequencies. Whereas the conclusions of these two studies by Stuart and his colleagues (Scott *et al.*, 2001; Elangovan and Stuart, 2005) are contradictory to the present results, the finding of comparable masking release above and below 1700 Hz in the present study is consistent with the results of Oxenham and Simonson (2009). In one set of conditions in that study, sentence recognition was assessed in speech-shaped noise and noise that was modulated by the envelope of a one-talker masker. Masking release was comparable for LP and HP filter conditions, a result supported and extended in the present experiments.

2. Effect of modulation rate on filtered stimuli

The effect of modulation rate was very similar across filter conditions, with largest effects at low rates for detection and 3AFC-ID. Masking release in the open-ID conditions was smaller, but increased as a function of masker AM rate. This general trend in masking release was similar across filter conditions and was also seen in the results of experiment 1 with full-spectrum stimuli. There was some indication of a reduced effect of masker AM rate on the HP as compared to previous full-spectrum data, but this difference was not apparent in the HP+16 condition. These results are consistent with the conclusion that the task-by-rate effect was consistent across low- and high-frequency regions of the speech signal, aside from a modest level effect.

One expectation touched on above, and proposed by Oxenham and Simonson (2009), is that masking release as-

sociated with masker AM requires some degree of cue redundancy relative to the cues necessary to perform the speech task; severely filtering the speech signal could reduce that redundancy and therefore reduce masking release. To the extent that LP and HP filtering the speech eliminates some of that redundancy, periodically unmasking "glimpses" of the signal could be less beneficial to performance. There was little evidence of a substantially reduced masking released for filtered stimuli in the current data set, however. Whereas the mean masking release associated with the open-ID, LP condition was approximately 2 dB less than that in the comparable full-spectrum conditions of experiment 1, this difference was not significant. It is possible that increasing the number of subjects might have revealed a small but significant effect. Whereas a reduction in redundancy might affect the optimal masker AM rate, it is possible that LP or HP filtering at 1700 Hz did not sufficiently reduce redundancy to allow such an effect to be observed. Future work will pursue more severe filtering conditions to assess this possibility.

IV. GENERAL DISCUSSION

In the present studies using CNC words, the pattern of masking release observed with the introduction of sinusoidal masker AM depends on the observer's task. Consistent with previous results obtained using low- and high-redundancy speech materials, lower rates of AM support better performance when the degree of detail required to correctly perform the task is relatively coarse, and high rates support better performance when fine detail is required. A similar task-by-AM rate interaction is seen when the signal (instead of the masker) is amplitude modulated. Therefore, the interaction between masker AM rate and task is unlikely to reflect the differential effects of forward masking at different SNRs. The pattern of masking release as a function of AM rate is relatively unaffected by LP or HP filtering the speech at 1700 Hz, provided that the data are compared for similar masker levels. This result was interpreted as indicating that absolute frequency effects, such as ability to encode temporal fine structure at low frequencies or hypothetically superior temporal resolution at high frequencies, are not required to account for the pattern of masking release as a function of AM rate.

There is continuing interest in the finding that masker AM provides greatly reduced benefits for listeners with moderate sensorineural hearing impairment as compared to normal-hearing listeners (Lorenzi *et al.*, 2006; Hopkins and Moore, 2009). While audibility plays some role in this result, deficits in temporal or spectral resolution might also limit performance in this task (Eisenberg *et al.*, 1995; Peters *et al.*, 1998), including effects related to the coding of temporal fine-structure cues (e.g., Hopkins and Moore, 2009). One interpretation of these findings is that reduced spectral and temporal resolution could reduce cue redundancy, such that sparse glimpses at the speech signal are not sufficient to support recognition, an idea recently proposed by Oxenham and Simonson (2009). The paradigm of the experiments described here could provide a tool for exploring the role of redundancy in the hearing impaired population.

The present results demonstrating an interaction between masker AM rate and task are difficult to reconcile with efforts to predict speech performance using the speech intelligibility index (SII) (ANSI, 1997). This model uses estimates of audibility of various spectral regions of the speech signal, in combination with speech importance functions, to compute a SII; the relationship between SII and percent correct is then obtained empirically. This basic model has been adapted for use with non-stationary maskers, by either averaging the SII associated with masker modulation maxima and minima (e.g., Horwitz *et al.*, 2007) or by computing the instantaneous SII as a function of time (Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2006). While not inherent to the SII model, context effects can be incorporated by taking into account the fact that the function relating percent correct to SII is steeper in high than low-predictability speech materials (Hargus and Gordon-Salant, 1995).

These adaptations of the SII model for use with a non-stationary signal are not consistent with the interaction between task and masker modulation rate observed here. Because the SII is based on audibility, it is not sensitive to the distribution of cues over time. An effect of modulation rate would be predicted by a SII model incorporating the limitations to audibility associated with forward masking and temporal resolution (such as Rhebergen *et al.*, 2006), but the results of the supplemental conditions of experiment 1 indicate that forward masking is not likely to be responsible for the task-by-rate interaction. Findings of the present study suggest that models of speech perception in modulated noise might benefit from inclusion of information regarding the temporal distribution of speech cues, including the degree of cue redundancy relative to the other sources of information available to the observer.

ACKNOWLEDGMENTS

This work was supported by a grant from NIH NIDCD (Grant No. R01 DC000418). We thank associate editor Ken Grant and two anonymous reviewers for their helpful comments and suggestions.

[1]Pilot testing of the self-scoring method indicated that verbalizing the response helped the subject commit to an unambiguous response prior to scoring. This method was judged to be superior to the more standard procedure, where an experimenter outside the booth monitors subject responses and performs scoring, because of difficulties differentiating odd pronunciation from errors in speech recognition and the possibility of experimenter bias. Whereas self-scoring could conceivably introduce errors, there is no reason to believe that such errors would vary systematically across stimulus conditions of the present study, as all observers were blind to the predictions of the study. Because the primary data of interest involved the pattern of results across masker conditions, any inaccuracy introduced by self-scoring is very unlikely to affect the results presented here.

ANSI (**1996**). *ANSI S3-1996, American National Standards Specification for Audiometers* (American National Standards Institute, New York).

ANSI (**1997**). *ANSI S3.5-1997, American National Standards Methods for Calculation of Speech Intelligibility Index* (American National Standards Institute, New York).

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (**1997**). "Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level," J. Acoust. Soc. Am. **101**, 1600–1610.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of

hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.
Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.
Bonatti, L. L., Pena, M., Nespor, M., and Mehler, J. (**2005**). "Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing," Psychol. Sci. **16**, 451–459.
Bronkhorst, A. W., Bosman, A. J., and Smoorenburg, G. F. (**1993**). "A model for context effects in speech recognition," J. Acoust. Soc. Am. **93**, 499–509.
Chang, J. E., Bai, J. Y., and Zeng, F. G. (**2006**). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," IEEE Trans. Biomed. Eng. **53**, 2598–2601.
de Laat, J. A. P. M., and Plomp, R. (**1983**). "The reception threshold of interrupted speech for hearing-impaired listeners," *Hearing, Physiological Bases and Psychophysics*, *Proceedings of the Sixth International Symposium on Hearing* (Springer-Verlag, Berlin), pp. 357–363.
DePaolis, R. A., Janota, C. P., and Frank, T. (**1996**). "Frequency importance functions for words, sentences, and continuous discourse," J. Speech Hear. Res. **39**, 714–723.
Dirks, D. D., and Bower, D. (**1970**). "Effect of forward and backward masking on speech intelligibility," J. Acoust. Soc. Am. **47**, 1003–1008.
Dirks, D. D., and Bower, D. R. (**1971**). "Influence of pulsed masking on spondee words," J. Acoust. Soc. Am. **50**, 1204–1207.
Dirks, D. D., Wilson, R. H., and Bower, D. R. (**1969**). "Effect of pulsed masking on selected speech materials," J. Acoust. Soc. Am. **46**, 898–906.
Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.
Elangovan, S., and Stuart, A. (**2005**). "Interactive effects of high-pass filtering and masking noise on word recognition," Ann. Otol. Rhinol. Laryngol. **114**, 867–878.
Festen, J. M. (**1987**). "Speech-reception threshold in a fluctuating background sound and its possible relation to temporal auditory resolution," *The Psychophysics of Speech Perception* (Nijhoff, Dordrecht, The Netherlands), pp. 461–466.
Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.
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.
Fullgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.
Gustafsson, H. A., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.
Haggard, M. P., Hall, J. W., and Grose, J. H. (**1990**). "Comodulation masking release as a function of bandwidth and test frequency," J. Acoust. Soc. Am. **88**, 113–118.
Hall, J. W., III, 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.
Hargus, S. E., and Gordon-Salant, S. (**1995**). "Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment," J. Speech Hear. Res. **38**, 234–243.
Hawkins, J. E., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.
Henry, B. A., McDermott, H. J., McKay, C., James, C. J., and Clark, G. M. (**1998**). "A frequency importance function for a new monosyllabic word test," Aust. J. Audiol. **20**, 79–86.
Hopkins, K., and Moore, B. C. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," J. Acoust. Soc. Am. **125**, 442–446.
Horwitz, A. R., Ahlstrom, J. B., and Dubno, J. R. (**2007**). "Speech recognition in noise: Estimating effects of compressive nonlinearities in the basilar-membrane response," Ear Hear. **28**, 682–693.
Howard-Jones, P. A., and Rosen, S. (**1993**). "The perception of speech in fluctuating noise," Acustica **78**, 258–272.
Huggins, A. W. (**1964**). "Distortion of the temporal pattern of speech: Interruption and alternation," J. Acoust. Soc. Am. **36**, 1055–1064.
Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: The effects of temporal resolution," J. Acoust. Soc. Am. **119**, 3097–3108.
Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.
Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.
Lee, J., and Bacon, S. P. (**1998**). "Psychophysical suppression as a function of signal frequency: Noise and tonal maskers," J. Acoust. Soc. Am. **104**, 1013–1022.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.
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.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Müsch, H., and Buus, S. (**2001**). "Using statistical decision theory to predict speech intelligibility. I. Model structure," J. Acoust. Soc. Am. **109**, 2896–2909.
Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.
Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.
Parikh, G., and Loizou, P. C. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear. Disord. **27**, 62–70.
Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.
Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**, 277–282.
Powers, G. L., and Wilcox, J. C. (**1977**). "Intelligibility of temporally interrupted speech with and without intervening noise," J. Acoust. Soc. Am. **61**, 195–199.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.
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.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.
Scott, T., Green, W. B., and Stuart, A. (**2001**). "Interactive effects of low-pass filtering and masking noise on word recognition," J. Am. Acad. Audiol. **12**, 437–444.
Stuart, A., and Phillips, D. P. (**1996**). "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-

hearing, and presbyacusic listeners," Ear Hear. **17**, 478–489.
Stuart, A., and Phillips, D. P. (**1997**). "Word recognition in continuous noise, interrupted noise, and in quiet by normal-hearing listeners at two sensation levels," Scand. Audiol. **26**, 112–116.
Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130–1138.
Studebaker, G. A., 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.
Toro, J. M., Shukla, M., Nespor, M., and Endress, A. D. (**2008**). "The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences," Percept. Psychophys. **70**, 1515–1525.
van Rooij, J. C., and Plomp, R. (**1991**). "The effect of linguistic entropy on speech perception in noise in young and elderly listeners," J. Acoust. Soc. Am. **90**, 2985–2991.

Pitch discrimination interference between binaural and monaural or diotic pitches[a)]

Hedwig E. Gockel[b)] and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Christopher J. Plack
Division of Human Communication and Deafness, University of Manchester, Manchester M13 9PL, United Kingdom

(Received 20 November 2008; revised 26 March 2009; accepted 20 April 2009)

Fundamental frequency (F0) discrimination between two sequentially presented complex (target) tones can be impaired in the presence of an additional complex tone (the interferer) even when filtered into a remote spectral region [Gockel, H., *et al.* (2004). J. Acoust. Soc. Am. **116**, 1092–1104]. This "pitch discrimination interference" (PDI) is greatest when the interferer and target have similar F0s. The present study measured PDI using monaural or diotic complex-tone interferers and "Huggins pitch" or diotic complex-tone targets. The first experiment showed that listeners hear a "complex Huggins pitch" (CHP), approximately corresponding to F0, when multiple phase transitions at harmonics of (but not at) F0 are present. The accuracy of pitch matches to the CHP was similar to that for an equally loud diotic tone complex presented in noise. The second experiment showed that PDI can occur when the target is a CHP while the interferer is a diotic or monaural complex tone. In a third experiment, similar amounts of PDI were observed for CHP targets and for loudness-matched diotic complex-tone targets. Thus, a conventional complex tone and CHP appear to be processed in common at the stage where PDI occurs.
 [DOI: 10.1121/1.3132527]

PACS number(s): 43.66.Hg, 43.66.Rq, 43.66.Ba [RLF] Pages: 281–290

I. INTRODUCTION

Despite many decades of research, the mechanism underlying the pitch perception of complex tones remains a matter of some dispute. One recent finding is that even when stimuli are represented quite differently in the auditory periphery, their pitches interact more centrally in an obligatory fashion that can impair performance in a forced-choice task. Specifically, when listeners are required to compare the fundamental frequencies (F0s) of two sequentially presented harmonic complexes, each filtered so that all of their components are unresolved by the peripheral auditory system, then performance can be disrupted by the addition of a group of resolved frequency components that are filtered into a lower frequency region (Gockel *et al.*, 2004, 2005, 2009b). The degree of this "pitch discrimination interference" (PDI) depends on the similarity between the F0s of the interfering and target complexes. As Gockel *et al.* (2004) pointed out, this suggests either that the pitches of resolved and unresolved complexes are initially processed by a common mechanism or that, if separate pitch mechanisms do exist, they are converted at some obligatory stage of processing into a common code.

In the experiments described here, we used PDI to study the commonality of processing between two stimuli that are initially processed very differently by the auditory system: resolved harmonic complexes and broadband noises that give rise to a binaurally generated pitch, termed Huggins pitch (HP). When samples of white noise are presented to each ear, which are identical in all frequency regions except for an interaural phase transition in a narrow frequency band, listeners perceive a faint pitch corresponding to the center frequency of that band (Cramer and Huggins, 1958). Each noise, when presented separately to one ear only, sounds just like white noise. When both noises are presented together, the perception is that of a noise, coming from the center of the head, and an additional tone, which is lateralized to one ear or the other (Raatgever and Bilsen, 1986). The perception of the HP crucially depends on the input from both ears being combined in some way. Thus, its percept must be derived from auditory processing at or higher than the level of the brainstem. Most current theories on the processing leading to the perception of a dichotic pitch assume the existence of an internal central spectrum that has a peak at the center frequency of the narrow band containing the interaural phase transition. How exactly this central spectrum is generated is still a matter of dispute (Raatgever and Bilsen, 1986; Culling *et al.*, 1998b; Hartmann and Zhang, 2003). Two well-known binaural models, the equalization-cancellation (EC) model (Durlach, 1960, 1972) and the central-activity pattern (CAP) model (Raatgever and Bilsen, 1986), both depend on interaural delay times. In the EC model, the left and right channel signals are subtracted after a preceding equalization stage, while in the CAP model, the left and right channel signals

[a)] Parts of this work were presented at the 153rd Meeting of the Acoustical Society of America, Salt Lake City, Utah, 4 June–8 June 2007 [J. Acoust. Soc. Am. **121**, 3068 (2007)].

[b)] Author to whom correspondence should be addressed. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

which are tuned in frequency and interaural time delay are added. The details of these models are beyond the scope of this paper. The binaural pitch is determined by the central spectrum. In the case of a complex pitch (containing multiple phase transition regions), the pitch of the binaural input has been assumed to be determined via a central pattern recognition process similar to those that can be applied to monaural or diotic pitch stimuli (Terhardt, 1974; Goldstein, 1973), which do not require binaural interaction (Raatgever and Bilsen, 1986). Note, however, that Akeroyd and Summerfield (1999) suggested a fully-temporal account of the perception of dichotic pitches, in which the output of an analysis of interaural timing based on the modified-EC model of Culling *et al.* (1998b, 1998a) feeds into a temporal-pitch model based on autocorrelation.

Our first experiment replicated and extended an earlier finding showing that listeners can perceive the "missing fundamental" of complex HP (CHP) when multiple phase transitions occur at frequencies that are integer multiples of a common F0, but when there is no transition at F0 (Bilsen, 1977); our listeners could match the binaural residue pitch of a CHP as accurately as they could match that of an equally loud diotic complex tone presented in noise. We then went on to show not only that PDI occurred between a group of monaurally or diotically presented resolved harmonics and CHP, but also that the amount of this interference was similar to that when the CHP was replaced by an equally loud resolved harmonic complex presented in a noise background. Our results thus show that CHP is processed in common with monaural pitches at an obligatory stage of processing, and that PDI can occur between stimuli for which the initial stages of processing are likely to be very different.

II. GENERAL METHOD AND PROCEDURE

In all experiments, a two-interval procedure was used. The details differed across experiments and will be described below for each experiment in turn. The mean F0 of the target complex tones was 200 Hz. The CHP stimulus had interaural phase transitions in frequency bands centered on the second to the fifth harmonics, i.e., centered on 400, 600, 800, and 1000 Hz. A stimulus which is generated by introducing a narrowband phase transition in a noise that is otherwise identical at the two ears leads to the perception of a pitch (termed HP^-), lateralized either to the left or the right, depending on the subject, with the diotic noise being perceived at the center of the head (Raatgever and Bilsen, 1986; Hartmann and Zhang, 2003; Zhang and Hartmann, 2008). The same stimulus configuration except for a global phase shift of 180° (termed HP^+) leads to the perception of a pitch that is perceived at the center of the head and a wideband noise with diffuse lateralization, i.e., lateralized to both left and right sides of the head with a tendency to spread toward the center. The HP^- configuration rather than a HP^+ configuration was used in the present study, as (i) the HP^- configuration results in a stronger pitch than the HP^+ configuration (Hartmann and Zhang, 2003) and (ii) it allowed us to also investigate the effect of relative lateralization of target sound and interfering sound for the maximum achievable difference between their perceived locations, i.e., to opposite sides (for more details, see Experiment II).

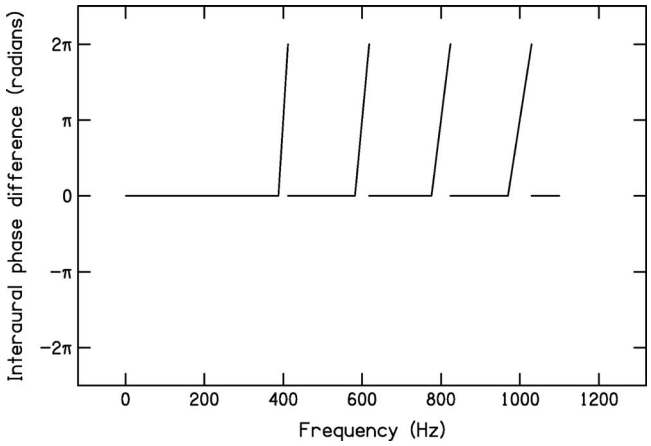


FIG. 1. Schematic of the CHP stimulus.

All stimuli were generated in MATLAB. The CHP was generated from a 1000-ms band-limited Gaussian noise sampled at 40 kHz. It was generated in the spectral domain by first applying a fast Fourier transform (FFT) to the noise and then modifying the phases of one of two matched buffers representing the left and right channels. A linear shift with frequency from 0π to 2π rad was added to the phases for frequency components from 3% below to 3% above the center frequency of the chosen harmonics. Applying an inverse FFT to the two spectral buffers gave the signal waveforms for the left and right channels. One out of 21 pre-generated realizations of the CHP, based on 21 realizations of Gaussian noise, was selected at random for each presentation of a CHP. The Gaussian noise extended up to 1.1 kHz in Experiment 1, where the F0 of the CHP was fixed at 200 Hz and up to 1.2 kHz in all other experiments, where the mean F0 of the CHP stimuli was roved between trials. Its spectrum level was 37.8 dB (re 20 μPa). The overall root-mean-square (rms) level of the CHP stimulus was 68.2 dB sound pressure level (SPL) in Experiment 1 and 68.6 dB SPL in the other experiments.

The duration of all stimuli was 1000 ms, including 40-ms raised-cosine onset and offset ramps. The silent interval between the two intervals within a trial was 500 ms. All tones were generated digitally. They were played out using a 16-bit digital-to-analog converter (CED 1401 plus), with a sampling rate of 40 kHz. Stimuli were passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 17.2 kHz (slope of 96 dB/oct) and presented using Sennheiser HD250 headphones.

III. EXPERIMENT 1: PITCH MATCHING TO THE MISSING FUNDAMENTAL OF A HP COMPLEX

A. Rationale

The objective of the first experiment was to determine whether a CHP tone, without a phase transition at the F0, is perceived by human listeners as having a residue pitch and to determine how salient that pitch is. We are aware of only one

other study investigating the perception of CHP in the absence of the fundamental. In that study (Bilsen, 1977), there were two phase transitions (at 600 and 800 Hz) and only two subjects, one of whom was Bilsen himself. Many listeners do not hear a residue pitch when a complex tone contains only two harmonics, even with monaural or diotic pitches (Smoorenburg, 1970). In Bilsen's study, his own matches of a pure tone frequency to the pitch of a CHP showed an average value of 205 Hz with a relatively narrow distribution, while the matches of the second subject showed a much wider distribution around 200 Hz. Here, the aim was to establish the pitch values and distribution of pitch matches for a greater number of listeners, for CHP with four phase transitions rather than two, which would be expected to lead to a clearer residue type pitch.

Listeners adjusted the F0 of a complex tone with harmonics 7–14 to match the residue pitch of a CHP with four phase transitions, centered on harmonics 2–5 of 200 Hz. Pitch matches to the CHP were compared with pitch matches to a diotic complex also containing harmonics 2–5 of 200 Hz, presented either in quiet or simultaneously with a diotic noise. When the diotic tone complex was presented with the diotic noise, the noise itself was identical to the noise used to generate the CHP stimuli (before phase shifts were applied), while the level of the tone complex was adjusted such that it had the same loudness as the tonal component in the CHP stimulus. The level of the diotic tone complex necessary to achieve this was determined for each subject individually in a loudness-matching experiment. Comparison of the distribution of the pitch matches to the CHP with that for the diotic complex tone, presented either in silence or at equal loudness in noise, provided information about the relative salience of the residue pitches derived from the three stimuli.

B. Methods

1. Loudness matching between diotic complex in noise and tonal component of CHP

The CHP and the diotic complex tone in noise, both "containing" harmonics 2–5 of a fundamental of 200 Hz, were presented in alternation. One of two virtual boxes lit up on a computer monitor in synchrony with each presentation of the CHP and the diotic complex. The CHP stimulus and the noise were fixed in level (68.2 dB SPL rms at each ear) and subjects had to adjust the level of the diotic complex so that its loudness was equal to that of the tonal component in the CHP. Subjects adjusted the level by moving a virtual slider on the monitor. The slider scale was marked from "−5" at the left hand side to "+5" on the right hand side, in steps of 1. Moving the slider from "0" to −5 attenuated the diotic complex by 5 dB on the next trial. Moving the slider from 0 to +5 increased the level of the diotic complex by 5 dB on the next trial. The slider could be moved to any position, thus allowing fine adjustments in level. After each presentation of the two stimuli, the slider's position was automatically moved to 0, before subjects could indicate the next desired adjustment of the level of the diotic tone complex. If subjects did not move the slider, the same stimulus pair was presented again. Subjects were encouraged to "bracket" the matching level several times, by making the diotic complex clearly softer than the tonal component of the CHP and then clearly louder (or vice versa), before making the fine adjustments. They were also encouraged to listen a few times to the same stimulus pair before indicating the next adjustment with the slider. Subjects pressed a virtual button on the monitor to indicate when they were satisfied with the loudness match. The matching level was defined as the level of the diotic tone complex presented immediately before the subject indicated a loudness match.

The starting level of the diotic complex was varied quasi-randomly in the range from 50 to 60 dB SPL per component, i.e., 56–66 dB SPL overall rms level. This range was chosen after some informal listening so that it covered levels at which the diotic tone was clearly louder or clearly softer than the tonal component in the CHP. For each presentation of the stimulus pair, 1 out of the 21 pre-generated CHP realizations was chosen randomly and the diotic noise that was presented simultaneously with the diotic tone in the other interval was identical to the noise used to generate the CHP stimulus (before phase shifts were applied).

The interval that contained the CHP was varied across blocks. A message on the monitor indicated to the subjects whether the level of the tone in the first or the second interval had to be adjusted to be equally loud to the tone heard in the other interval. At least 20, but more typically 40, loudness matches were collected for each subject (20 matches for each order of the two stimuli).

2. Pitch matching

Subjects adjusted the F0 of a complex tone with harmonics 7–14 to match the following: (1) the residue pitch of a CHP with phase transitions at 400, 600, 800, and 1000 Hz presented at a rms level of 68.2 dB SPL; (2) the residue pitch of a diotic complex with harmonics 400, 600, 800, and 1000 Hz presented in quiet at a rms level of 38 dB SPL; and (3) the residue pitch of the same diotic complex as in (2) but presented simultaneously with a diotic noise which was the same as in the CHP (1 out of 21 realizations randomly drawn for each trial), at a level that was determined for each subject so that the tonal component was equally loud as for the CHP (around 62 dB SPL rms level, see results below). The matching complex was presented diotically in quiet with a rms level of 35 dB SPL (26 dB SPL per component).

The pitch matching procedure was essentially the same as the loudness-matching procedure described above. Moving the slider from 0 to +5 or −5 increased/decreased the F0 of the matching sound to be presented on the next trial by 5 Hz. The slider could be moved to any position and returned to its central value at the start of each trial. The starting value of the F0 of the adjustable sound varied randomly in the range 200 ± 15 Hz. The interval that contained the matching sound was varied across blocks. A message on the monitor indicated to the subjects whether the F0 of the tone in the first or the second interval had to be adjusted to match the pitch of the tone heard in the other interval. For each subject and condition, 50 pitch matches were collected for each order of the two stimuli. The results from the two orders of the stimuli were averaged.

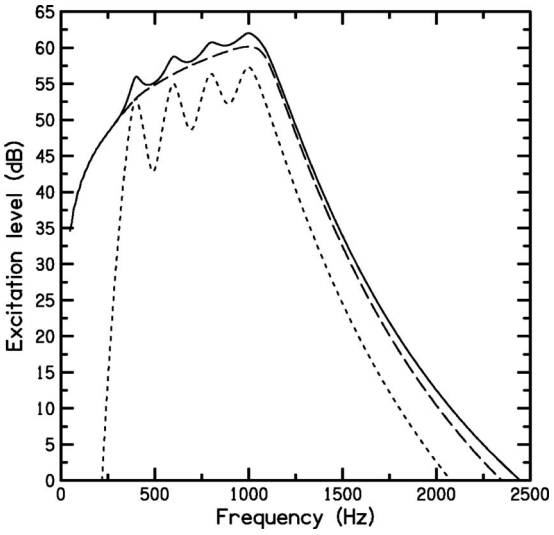


FIG. 2. Excitation patterns (following Moore *et al.*, 1997) for the white noise (dashed line), the diotic complex tone (dotted line) presented simultaneously with the white noise at the point of equal loudness with the tonal part of the CHP, and for the combination of the two (solid line).

3. Subjects

Four subjects, with various degrees of musical training, participated in all conditions of Experiment 1. They ranged in age from 19 to 28 years, and their quiet thresholds at octave frequencies between 250 and 4000 Hz were within 15 dB of the ISO 398-8 (2004) standard. To familiarize subjects with the procedure and equipment, they were given between 2 and 4 h of practice.

C. Results and discussion

1. Loudness matching between the diotic complex in noise and the tonal component of CHP

There were no significant differences between the loudness matches obtained for the two orders of stimulus presentation, and so the results were averaged across orders. The mean rms level of the diotic complex presented simultaneously with noise at the point of equal loudness with the tonal part of the CHP was [standard error (SE), in parentheses] 62.1 dB SPL (0.2) for subject 1, 61.9 dB SPL (0.1) for subject 2, 61.3 dB SPL (0.27) for subject 3, and 63.0 dB SPL (0.14) for subject 04. The mean (and SE) across subjects was 62.1 dB SPL (0.3).

The noise had a spectrum level in its passband of 37.8 dB (re 20 μPa), and a rms level of 65.6 dB in the frequency band from 400 to 1000 Hz, which contained the harmonics of the diotic complex. Therefore, equal loudness of the tonal component of the CHP and the diotic tone complex in noise occurred when the signal-to-noise ratio of the latter was only −3.5 dB. Figure 2 shows the excitation patterns (following Moore *et al.*, 1997) calculated separately for the noise background (dashed line) and for the diotic tone complex (dotted line) at the level of equal loudness with the tonal part of the CHP. The excitation pattern for the combined stimulus is shown by the solid line. At the level of equal loudness, the partial loudness of the diotic tone complex in the white noise was calculated as 22.5 phons or 0.192 sones (following Moore *et al.*, 1997). This level is 6.6 dB above the masked threshold for the diotic tone in the noise predicted by the loudness model of Moore *et al.* (1997). The latter result indicates that the loudness of the CHP corresponds approximately to that of the diotic tone complex presented at a sensation level of 6.6 dB SL. Overall, these results show that a discernible but relatively faint pitch was heard in the CHP, with good agreement between the four subjects.

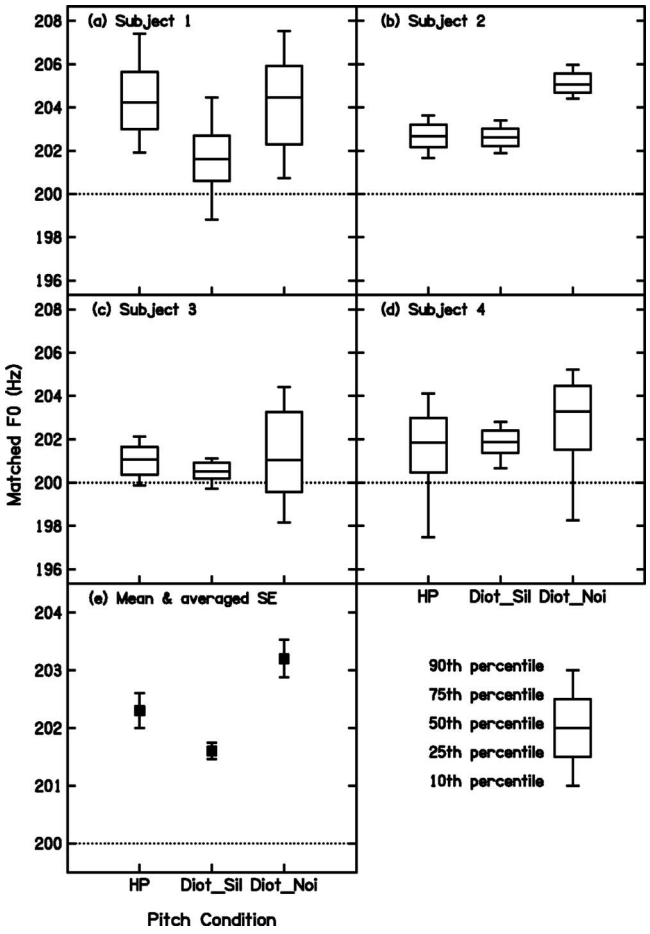


FIG. 3. Results of the pitch matching experiment (Experiment 1). The matched F0 of a diotically presented harmonic tone complex containing harmonics 7–14 when its pitch was perceived equal to that of a target complex tone, "containing harmonics 2-5" of an F0 of 200 Hz. The target complex could be either a CHP (condition "HP") containing phase transitions centered on the frequencies corresponding to the harmonics, or a harmonic tone complex presented diotically in quiet (condition "Diot_Sil"), or a harmonic tone complex presented diotically in the same noise as used for generating the HP, at a level so that the tone complex is perceived equally loud as the tonal part of the CHP stimulus (condition "Diot_Noi"). Figures 1(a)–1(d) show, for each of the four subjects, the 90th, 75th, 50th, 25th, and 10th percentiles of the distribution of the pitch matches in each of the three conditions. Figure 1(e) shows the mean of the pitch matches and the size of the "typical SE," i.e., the mean of the SEs calculated initially separately for each subject and condition.

2. Pitch matching

Figure 3 shows the means and distributions of the F0 of the diotic complex containing harmonics 7–14 when its pitch was matched to that of the target sound. There were no significant differences between the pitch matches obtained for the two orders of stimulus presentation, and results were averaged across the two orders. Figures 3(a)–3(d) show the

90th, 75th, 50th, 25th, and 10th percentiles of the distributions of the pitch matches for each of the four subjects. Figure 3(e) shows the mean matched F0 across subjects and the size of the typical SE, i.e., the average across the four individual SEs that were first calculated separately for each subject.

The mean F0 matched to the CHP was 202.3 Hz.[1] In spite of some differences between subjects with regard to their pitch matching reliability, the agreement between subjects was good. The matched F0 was somewhat above the "true F0" of the CHP of 200 Hz, but this was also true for the mean matched F0s to the other target sounds (201.6 Hz for the diotic tone in silence and 203.2 Hz for the diotic tone in noise). A repeated-measures one-way analysis of variance (ANOVA), using the mean of the matched F0s from each subject and condition as input, showed no significant difference between the mean pitch matches across the three different target sounds. There was a tendency for the SE of the matches to be smallest for the diotic tone in silence. However, a repeated-measures one-way ANOVA, with the SE of the matched F0s from each subject and condition as input, showed no significant difference between the SE of the pitch matches across the three different target sounds.

Overall these results show that the CHP, in the absence of a phase transition at the F0, evoked a residue pitch which corresponded well to that of a diotic complex tone with harmonics at frequencies corresponding to the center frequencies of the phase transitions in the CHP. The salience of the residue pitch of the CHP, as assessed by the SE of the pitch matches, corresponded well with that of the loudness-matched diotic complex in noise.

IV. EXPERIMENT 2: PDI BETWEEN A HP COMPLEX AND A CONVENTIONAL HARMONIC TONE

A. Rationale

Having established that a residue pitch can be perceived for a CHP with missing fundamental, the main objective of the second experiment was to investigate whether PDI would occur between the CHP as the target and a monaural or diotic complex tone as the added sound (interferer). The perception of the residue pitch of a CHP and a monaural or diotic residue pitch involve, at least initially, different processes: perception of the former requires binaural interaction while perception of the latter does not (see Sec. I). If PDI were observed between these two stimuli, it would indicate that PDI can occur between stimuli that are initially processed in a different way. This in turn would be consistent with the idea mentioned in Sec. I that the pitches of resolved and unresolved complexes could initially be processed by different mechanisms rather than the same mechanism, and that the reported PDI might have occurred at a later stage of processing where the pitch information was converted into a common code. Gockel *et al.* (2009b) recently reported significant PDI between complex tones presented to opposite ears, showing that PDI can occur at or after the stage where pitch information from the two ears has been combined. This strengthens the possibility that PDI might be observed between a CHP and a monaural or diotic pitch.

The second objective was to investigate whether the previously reported dependence of PDI on the similarity between the F0s of the target and the interferer would also be observed between a CHP target and a monaural or diotic interferer. To assess this, the F0 of the interferer either corresponded to the nominal F0 of the target or it was increased by 40%. If a similar dependence on F0 similarity was observed, it would support the interpretation that the impairment with a CHP target and PDI between monaural pitches are caused by similar processes.

The third objective was to investigate the role of perceived location of the target and interferer. Gockel *et al.* (2009b) observed significantly less PDI when the interferer was presented contralaterally to the target than when it was presented ipsilaterally. Thus, relative ear of entry of target and interferer played an important role in PDI. In the present study, the perceived location of the tonal part of the CHP varied across subjects, but was very reliable within a subject (Zhang and Hartmann, 2008). By presenting the interferer either to the left ear, or to the right ear, or diotically, its perceived location was varied relative to that of the tonal part of the CHP, the latter varying across subjects.

In the non-shifted F0 conditions, the interferer was a complex tone containing harmonics 7–14 with an F0 corresponding to the nominal F0 of the target. Thus the interferer contained higher harmonics than the target but still had a salient pitch (Moore and Glasberg, 1988; Houtsma and Smurzynski, 1990; Moore and Peters, 1992; Moore *et al.*, 2006). The level of the interferer was chosen such that it was perceived as equal in loudness to the tonal component in the CHP target stimulus. The level of the tone complex necessary to achieve this was determined individually for each subject in a loudness-matching experiment (see Appendix).

B. Method

Subjects had to discriminate between the F0s of two sequentially presented CHP, i.e., they had to indicate which of the two HP complexes, both with phase transitions around harmonics 2–5, had the higher F0. The mean F0 was varied across trials from 181.8 Hz (200/1.1) to 220 Hz (200×1.1), to encourage subjects to compare the pitch of the two targets presented in each interval. In each trial, in one, randomly chosen, interval the target complex had an F0 equal to $F0 - \Delta F0/2$, while in the other interval its F0 was $F0 + \Delta F0/2$. The difference in F0 between the two target tones in a trial, $\Delta F0$, was fixed, and percent-correct performance was measured. The size of $\Delta F0$ was chosen for each subject individually, in a preliminary experiment, so that performance in terms of d' was between about 1.6 and 1.8 in condition None, which was the easiest condition. The values of $\Delta F0$ were 1.4%, 0.9%, 1%, 1.8%, and 1.4% for subjects 1, 2, 3, 4, and 5, respectively. Correct-answer feedback was provided after every trial.

The target sounds were either presented alone (condition "None") or with a synchronously gated interferer. The interferer was either a harmonic complex containing harmonics 7–14 with an F0 that was equal to the mean of the F0s of the two target sounds, and thus its F0 was never identical to that

of a target sound, or a harmonic complex containing harmonics 5–10 with an F0 that was 40% higher than the mean target-F0. The F0 of the interferer was always identical in the two intervals of a trial and thus was non-informative for the task. In the non-shifted case, the interferer was presented either diotically or monaurally, either to the side where the tonal percept of the CHP was lateralized (condition "Ipsi") or to the opposite side (condition "Contra"). In these conditions, the level of the interferer corresponded to the individually determined levels of equal loudness between the interferer and the tonal percept of the CHP; across subjects, this corresponded to average rms levels of the interferer of about 40.6 and 36.6 dB SPL in the monaural and the diotic conditions, respectively (for details, see Appendix). In the pitch-shifted condition (condition "Dio_PS"), the interferer covered the same frequency range as the non-shifted interferers, and was presented diotically at the same rms level as the diotic non-shifted interferer.

The four interferer conditions were tested in blocks of 105 trials each. Each block with an interferer present was preceded by a block of 55 trials with the target alone. This was done so subjects knew the characteristics of the target sound. The first five trials within each block were considered as "warm-up" trials and results from those were discarded. One block was run for each interferer condition in turn, before additional blocks were run in any other condition. At least 400, but usually 500 trials were collected for each subject in each interferer condition. As each block with an interferer was preceded by a 55-trial block without an interferer, this means that performance in the target alone condition is based on at least 800, but usually 1000 trials. Subjects received from 3 to 8 h of practice until performance seemed to be stable within conditions, before data collection proper was started.

Five subjects participated in all conditions of Experiment 2. Four of them had also taken part in Experiment 1. They ranged in age from 19 to 45 years, and their quiet thresholds at octave frequencies between 250 and 4000 Hz were within 15 dB of the ISO 389-8 (2004) standard. All subjects had previously participated in other experiments on PDI.

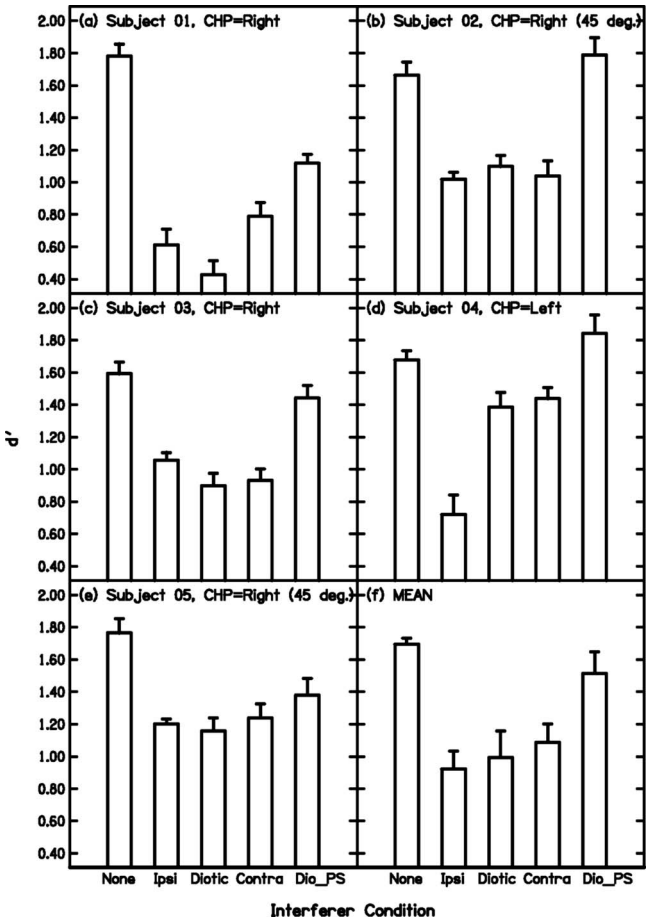


FIG. 4. Results of the PDI experiment (Experiment 2). Performance, d', for discrimination of the residue pitch of a CHP which was either presented alone (condition "None") or simultaneously with another complex tone (interferer) whose F0 was either centered between the F0 of the target sounds in the two intervals, or increased by 40% (condition "Dio_PS"). When the interferer was shifted in pitch, it was presented diotically. When the interferer was unshifted in pitch, it was presented either monaurally to the same side to which the CHP was lateralized (condition "Ipsi") or monaurally and to the opposite side of the CHP (condition "Contra") or diotically (condition "Diotic"). Figures 1(a)–1(e) show, for each subject separately, the mean d', averaged across blocks of 100 trials for the conditions with interferer and averaged across blocks of 50 trials in the absence of an interferer, together with the corresponding SEs, calculated across blocks. Also indicated in the top left corner is the lateralization of the tonal percept in the CHP complex for each subject; 45° indicates that, for this subject, the lateralization was half way to the side. Figure 1(f) shows the mean d' across all five subjects and the SEs across subjects.

C. Results and discussion

Figure 4 shows performance (d') for F0 discrimination for the CHP tone presented either alone (condition "None") or simultaneously with the interferer. The interferer was presented at a level leading to the same loudness as the tonal part of the CHP, as determined individually for each subject (see Appendix for details). In the absence of the interferer, d' values were between 1.6 and 1.8 for all of the subjects [Figs. 4(a)–4(e)]. This was as intended and shows that F0 discrimination for CHP is good for relatively small values of ΔF0 (0.9%–1.8%). It was not quite as good as for a complex tone containing resolved harmonics presented at moderate but higher sensation level, for the same values of ΔF0 (Gockel *et al.*, 2009a), but it was clearly better than for complex tones containing only unresolved harmonics (see, e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Gockel *et al.*, 2004, 2006). In the presence of an interferer with F0 centered between those of the target complexes in the two intervals (3 bars in the middle), performance deteriorated for all subjects. However, the degree of impairment varied across subjects and conditions. Subject 04 [Fig. 4(d), the first author] showed markedly higher performance when the perceived location of the interferer was different from that of the target than when it was similar, i.e., diotic and contralateral presentation both led to higher performance levels than ipsilateral presentation. The other four subjects did not seem to be able to take advantage of the difference between the lateralization of the target and the interferer. When the F0 of the interferer was increased by 40% above the nominal target F0 (condition Dio_PS), performance improved relative to that for the non-shifted conditions, for all

subjects. Figure 4(f) shows the mean data and SEs across subjects. On average, PDI, defined as the difference between the d' value in condition None and that observed in the presence of an interferer, was about 0.7 when the interferer's F0 was centered at the nominal target-F0 and about 0.2 when the interferer's F0 was shifted.

A repeated-measures one-way ANOVA, calculated with the mean d' values from all subjects and conditions as input, showed that there was a highly significant difference between conditions [$F(4,16)=12.92, p<0.001$]. *Post hoc* contrasts based on Fisher's least significant difference procedure showed that performance in condition None was significantly higher than in conditions Ipsi and Contra ($p<0.01$) and Diotic ($p<0.05$). In contrast, in the presence of the F0-shifted interferer performance did not differ significantly from that observed in condition None and was significantly higher than that observed in condition Diotic ($p<0.01$).

The results show that significant PDI does occur between a CHP target and a monaural or diotic interferer. The size of the PDI depended on the similarity of the F0s of the CHP target and the interferer. This indicates that the PDI for a binaural-pitch target and a conventional-pitch interferer, two stimuli which initially are processed in a different way, is likely to be caused by the same process as the PDI observed between conventional-pitch stimuli. The present results further support the idea that PDI occurs at least partly at or after the stage at which pitch-relevant information is combined across the two ears. In the present experiment, a difference in perceived lateralization of the target and interferer was an ineffective cue, except for one subject. In contrast, Gockel *et al.* (2009b) found that presentation of the target to one ear and the interferer to the other ear significantly reduced (but did not abolish) PDI in comparison to the case when both tones were presented to the same ear. Thus, relative ear of entry seems to have a more powerful influence on PDI than perceived lateralization. This could be another example of the small effect of perceived location in contrast to the significant effects of relative ear of entry reported for *concurrent* sound segregation (Culling and Summerfield, 1995; Hukin and Darwin, 1995; Gockel and Carlyon, 1998; Gockel, 2000; Darwin and Hukin, 2004).

V. EXPERIMENT 3: COMPARISON OF PDI WITH EITHER HP OR CONVENTIONAL PITCH COMPLEX AS TARGET SOUND

A. Rationale

After establishing that PDI can occur between a CHP target and a monaural or diotic conventional-pitch interferer, we assessed whether, for the same diotic pitch interferer, PDI would be larger if the target was a conventional-pitch complex than when it was a CHP. In other words, is there any benefit if the target and the interferer are initially processed in a different way compared to when they are processed in the same way?

B. Methods

Two different stimuli were used as targets. The first was the same CHP as used in Experiment 2. The second was a diotic tone complex, also containing harmonics 2–5 of 200 Hz, presented simultaneously with a diotic noise. This second target is the same as was used for pitch matching, as described for Experiment 1. Briefly, the diotic noise was identical to the noise used to generate the CHP stimuli (before phase shifts were applied) and the diotic tone complex was presented at a level leading to the same loudness as for the tonal component in the CHP stimulus, as determined for each subject in Experiment 1. Only the diotic (not the monaural) interferers from Experiment 2 were used, i.e., a tone complex containing harmonics 7–14 with an F0 equal to the mean of the F0s of the two target sounds (nominally 200 Hz) or containing harmonics 5–10 with an F0 that was nominally 280 Hz.

The procedure was the same as that used in the PDI experiment described above. The four conditions with an interferer were tested in blocks of 105 trials each. Each block with an interferer present was preceded by a block of 55 trials with the target alone. The first five trials within each block were considered as warm-up trials and results from those were discarded. Conditions with the two different targets were blocked within each session; in one half of the session the CHP was the target and in the other half, the diotic tone in noise was the target. The order of the two was counterbalanced across sessions. One block was run for each interferer condition in turn, before additional blocks were run in any other condition. At least 500 trials were collected for each subject in each condition.

Four subjects, aged between 19 and 45 years, participated in all conditions of Experiment 3. All of them had also taken part in Experiment 2. The values of ΔF0 were 1.4%, 0.9%, 1%, and 1.8%, for subjects 1, 2, 3, and 4, respectively.

C. Results and discussion

When the targets were presented alone, the mean d' values (with SEs across the four subjects given in brackets) were 1.70 (0.05) for the CHP and 1.86 (0.09) for the diotic tone. Figure 5 shows the PDI found using the two interferers. When the interferer's F0 was equal to the mean of the targets' F0s, the PDI was between 0.65 and 0.8. Both of these values were quite close to those observed in the previous experiment using the CHP as target. When the interferer's F0 was increased, the PDI values were between 0.24 and 0.14. Again, both of these values were quite similar to those obtained with the CHP as target in the previous experiment.

A repeated-measures two-way ANOVA, with factors type of target and F0 of the interferer, was calculated, using the mean PDI values from all subjects and conditions as input. The results showed a significant main effect of interferer F0 [$F(1,3)=52.36, p<0.01$]. The main effect of target type and the interaction were both not significant. Therefore, PDI did not differ significantly when the target was a conventional complex pitch and when it was a CHP. In other words, PDI was not significantly smaller when the target and the interferer were initially processed in a different way than when they were initially processed in the same way. To assess whether the presence of the pitch-shifted interferer produced significant impairment, one-sample t-tests were calcu-

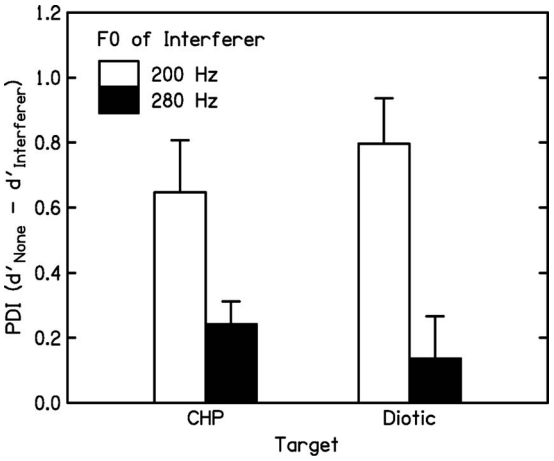


FIG. 5. Mean PDI, defined as d' value for the target alone minus d' value in the presence of the interferer, and corresponding SEs across four subjects observed in Experiment 3. The left-hand group of 2 bars shows PDI obtained for the CHP target. The right-hand group of 2 bars shows PDI obtained for the diotic tone complex presented in diotic noise as target. The white and the black bars show results when the interferer's F0 was centered at the nominal target F0 and when it was increased by 40% above the nominal target F0, respectively.

lated on the mean PDI values observed with the 280-Hz interferer from all subjects. The results showed that PDI for the pitch-shifted interferer was significantly larger than zero when the CHP was the target ($p<0.05$; two-tailed) but was not significantly above zero when the diotic tone was the target.

In summary, the results indicate that the size of the impairment in F0 discrimination of two sequentially presented target tones caused by a simultaneous interferer is not (or only slightly) affected by whether the target and the interferer are both diotic tones or whether the target is a CHP and the interferer is a diotic tone. In the latter case, the target and interferer are likely to be initially processed in a different way. The absence of a significant benefit from the difference in initial processing indicates that the processes underlying PDI most likely do not differentiate between the "origin" of their input.

VI. SUMMARY AND CONCLUSIONS

The F0 of a diotic complex tone containing harmonics 7–14 was adjusted to match the residue pitch of (i) a CHP stimulus and (ii) a loudness-matched diotic complex in noise, both containing harmonics 2–5 of a 200 Hz F0. Matches did not differ in value or accuracy between the binaural and the conventional-pitch targets (Experiment 1), showing that the residue pitch evoked by a CHP with a missing fundamental is comparable in salience to that of a loudness-matched diotic complex in noise.

F0 discrimination performance for a CHP target in the presence of a loudness-matched diotic or monaural tone complex (occupying a different spectral region from that of the target) was clearly impaired when the F0s of target and interferer were similar, but was not significantly (Experiment 2) or only slightly (Experiment 3) impaired when the interferer's F0 was 40% above that of the target. The tuning effect indicates that the impairment in F0 discrimination was likely to be based on similar processes to those underlying PDI effects between monaural or conventional pitches. The relative lateralization of the interferer and the CHP (the target) affected PDI only for one subject.

PDI with a diotic interferer was not significantly larger when the target was a diotic complex than when it was a CHP (Experiment 3). Thus, initial processing of the target and interferer in a different way did not lead to a significant reduction in PDI.

Overall, the results for a CHP tone were similar to those for a loudness-matched diotic complex tone with respect to the pitch matches and also with respect to the PDI that was observed in the presence of a diotic interferer. Thus, PDI can occur between stimuli which initially are likely to be processed in a different way. This further supports the notion that PDI is unlikely to occur at a very peripheral stage of auditory processing. It indicates that a conventional complex tone and a CHP are likely to be processed in common at the stage where PDI occurs. If PDI occurred *at* the stage of pitch extraction, the current results would support the idea of the existence of a central pitch processor common to binaural and monaural signals as expressed by Bilsen (1977). Alternatively, PDI may occur after the pitch extraction processes, which might not be in common for binaural and monaural pitches. If this is so, then the pitches must either be transformed into a common code at a later stage or not be independently accessible.

ACKNOWLEDGMENTS

This work was supported by the EPSRC Grant No. EP/D501571/1. We thank Brian Moore and Brian Glasberg for providing us with their program to calculate binaural loudness as described in Moore and Glasberg, 2007. We also thank Brian Moore, Richard Freyman, and three anonymous reviewers for helpful comments on an earlier version of this paper.

APPENDIX: LOUDNESS MATCHING BETWEEN THE TONAL COMPONENT OF THE CHP TARGET AND THE INTERFERER

This appendix describes the loudness-matching experiment between the tonal component of the CHP target and the interferer in Experiment 2.

1. Method

The procedure used was the same as that used for the loudness matches in Experiment 1. Here, subjects adjusted the level of a complex tone containing harmonics 7–14 with an F0 of 200 Hz which was presented in silence, so that it had the same loudness as the tonal part of the CHP stimulus. The latter was the same CHP stimulus with phase transitions around harmonics 2–5 of a nominal F0 of 200 Hz that was investigated in Experiment 1, except that the noise band extended up to 1.2 kHz rather than 1.1 kHz. The noise band was extended up to 1.2 kHz, to allow for the F0-randomization across trials applied in the following experiments on PDI (see Secs. IV and V). There were three conditions: The conventional-pitch complex was presented to the

TABLE I. Matched levels of a complex tone containing harmonics 7–14 of a 200 Hz F0 at the point of equal loudness with the tonal percept in a CHP containing phase transitions centered on frequencies corresponding to harmonics 2–5 of a 200 Hz F0 as a function of mode of presentation. Mode of presentation was (i) ipsilateral: monaural on the same side as lateralized tonal percept of the CHP; (ii) contralateral: monaural on the opposite side as lateralized tonal percept of the CHP; and (iii) diotic. Values are mean rms levels in dB SPL and the corresponding SEs in parentheses.

	Mode of presentation		
	Ipsilateral	Contralateral	Diotic
Subject 01	38.7 (0.45)	40.5 (0.41)	37.7 (0.27)
Subject 02	46.0 (0.18)	38.6 (0.14)	37.5 (0.13)
Subject 03	40.8 (0.21)	41.7 (0.21)	36.6 (0.21)
Subject 04	39.7 (0.49)	38.9 (0.51)	34.0 (0.20)
Subject 05	41.6 (0.22)	40.0 (0.26)	36.8 (0.24)
Mean	41.4 (1.13)	39.9 (0.50)	36.6 (0.59)

left ear, to the right ear, or diotically. These three conditions and the order of target and adjustable sounds were varied across blocks of usually ten matches and counterbalanced across subjects. The starting level of the conventional-pitch complex was varied randomly in the range from 21 to 31 dB SPL per component, i.e., 30–40 dB SPL rms level. This range was chosen after some informal listening so that it covered levels at which the diotic tone complex was clearly louder or clearly softer than the tonal component in the CHP stimulus. At least 40 loudness matches were collected for each condition and subject (20 matches for each order of the two stimuli), and results were averaged across the two orders of stimuli.

2. Results and discussion

Table I shows the mean rms level of the complex tone containing harmonics 7–14 of an F0 of 200 Hz at which the complex was judged to be equally loud to the tonal part of the CHP stimulus, for each of the three conditions. For four of the five subjects the lateralization of the tonal part of the CHP stimulus was toward the right, while for subject 04 the tone was clearly perceived at the left. Thus, ipsilateral presentation of the tone complex containing harmonics 7–14 meant presentation to the right ear for four subjects and presentation to the left ear for subject 04. A repeated-measures one-way ANOVA showed that there was a significant difference between the matched loudness levels in the three conditions $[F(2,8)=8.1, p<0.05]$.[2] *Post hoc* contrasts based on Fisher's least significant difference procedure showed that the level of the diotic tone required to match the loudness of the CHP was significantly lower than that in the monaural conditions ($p<0.05$) but, as expected, there was no significant difference between ipsilateral and contralateral presentations. In the diotic condition, the matched level was 4.05 dB lower than the average of the matched levels in the two monaural conditions, i.e., diotic presentation of the complex tone increased its loudness by an amount corresponding to a 4.05 dB increase in level.

Loudness calculations (following Moore and Glasberg, 2007) showed that, for monaural presentation, the matched level corresponds to a loudness value of 1.766 sones (47.6 phons) while for diotic presentation, the matched level corresponds to a loudness value of 1.977 sones (49.2 phons). Thus, the 4.05 dB difference between the matched levels in the monaural and the diotic conditions is close to the 5.6 dB predicted by the recent modification of the loudness model of Moore *et al.* (1997). While some empirical data suggest doubling of loudness with diotic presentation, more recent studies suggest less than that (for an overview see Moore and Glasberg, 2007), and the recent modification of the model was developed specifically to account for these more recent findings. Thus, the present data are reasonably in line with the recent findings.

The alert reader will have noticed the discrepancy between the loudness values calculated for the partial loudness of the diotic tone in noise containing harmonics 2–5 at the matched level in Experiment 1 (22.5 phons) and for the diotic interferer containing harmonics 7–14 presented in silence calculated here (49.2 phons). Both sounds were adjusted in level to be equally loud to the tonal component in the CHP stimulus, and thus one might expect their loudness values to be equal to each other. A possible explanation for why the diotic interferer was matched at a higher level is that subjects included the energy of the noise to a certain degree when they assessed the loudness of the tonal component of the CHP stimulus. The interferer was presented in quiet and thus would need to be adjusted to a higher level to be perceived as equally loud. Also, because the interferer had a very different timbre from the tonal part of the CHP stimulus, loudness comparison would have been difficult. In contrast, the diotic tone in noise containing harmonics 2–5 (Experiment 1) was perceptually very similar to the CHP. Therefore, loudness comparison would have been easier, and, importantly, it would be unlikely that the noise would contribute differentially to the loudness of the tonal percept in the two stimuli.

[1]It is unlikely that listeners did *not* perceive the residue pitch of the CHP stimulus and that this match was only obtained because the pitch chroma of the adjusted sound (about 200 Hz) matched the pitch chroma of the first phase transition (about 400 Hz) in the CHP stimulus. First, the pitch of the CHP stimulus and the pitch of the diotic tone in noise, which were presented sequentially in the loudness-matching part of Experiment 1, did not appear to have an octave relationship but rather they sounded the same. Second, Gockel *et al.* (2009b) showed that PDI was minimal when the F0s of target and interferer differed by as much as 1 octave. Thus, if the pitch of the CHP stimulus were mainly determined by the individual phase transition around 400 Hz, then one would not expect to observe the amount of PDI that was found in Experiments 2 and 3.

[2]Throughout the paper, if appropriate, the Huynh–Feldt correction was applied to the degrees of freedom (Howell, 1997). In such cases, the corrected significance value is reported.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A fully-temporal account of the perception of dichotic pitches," Br. J. Audiol. **33**, 106–107.

Bilsen, F. A. (**1977**). "Pitch of noise signals: Evidence for a "central spectrum"," J. Acoust. Soc. Am. **61**, 150–161.

Cramer, E. M., and Huggins, W. H. (**1958**). "Creation of pitch through binaural interaction," J. Acoust. Soc. Am. **30**, 413–417.

Culling, J. F., Marshall, D. H., and Summerfield, A. Q. (**1998a**). "Dichotic pitches as illusions of binaural unmasking. II. The Fourcin pitch and the dichotic repetition pitch," J. Acoust. Soc. Am. **103**, 3527–3539.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998b**). "Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the "binaural edge pitch"," J. Acoust. Soc. Am. **103**, 3509–3526.

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. (**2004**). "Limits to the role of a common fundamental frequency in the fusion of two sounds with different spatial cues," J. Acoust. Soc. Am. **116**, 502–506.
Durlach, N. I. (**1960**). "Note on the equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **32**, 1075–1076.
Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **2**.
Gockel, H. (**2000**). "Perceptual grouping and pitch perception," in *Results of the Eighth Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, M. Meis, and C. Reckhardt (BIS, Oldenburg, Germany).
Gockel, H., and Carlyon, R. P. (**1998**). "Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection," J. Acoust. Soc. Am. **104**, 3534–3545.
Gockel, H., Carlyon, R. P., and Moore, B. C. J. (**2005**). "Pitch discrimination interference: The role of pitch pulse asynchrony," J. Acoust. Soc. Am. **117**, 3860–3866.
Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.
Gockel, H., Moore, B. C. J., Plack, C. J., and Carlyon, R. P. (**2006**). "Effect of noise on the detectability and fundamental frequency discrimination of complex tones," J. Acoust. Soc. Am. **120**, 957–965.
Gockel, H. E., Carlyon, R. P., and Plack, C. J. (**2009a**). "Further examination of pitch discrimination interference between complex tones containing resolved harmonics," J. Acoust. Soc. Am. **125**, 1059–1066.
Gockel, H. E., Hafter, E. R., and Moore, B. C. J. (**2009b**). "Pitch discrimination interference: The role of ear of entry and of octave similarity," J. Acoust. Soc. Am. **125**, 324–327.
Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.
Hartmann, W. M., and Zhang, P. X. (**2003**). "Binaural models and the strength of dichotic pitches," J. Acoust. Soc. Am. **114**, 3317–3326.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.
Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, Belmont, CA).
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.
ISO 389-8 (**2004**). Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (International Organization for Standardization, Geneva).
Moore, B. C. J., and Glasberg, B. R. (**1988**). "Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments," in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).
Moore, B. C. J., and Glasberg, B. R. (**2007**). "Modeling binaural loudness," J. Acoust. Soc. Am. **121**, 1604–1612.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240.
Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.
Moore, B. C. J., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.
Raatgever, J., and Bilsen, F. A. (**1986**). "A central spectrum theory of binaural processing: Evidence from dichotic pitch," J. Acoust. Soc. Am. **80**, 429–441.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Smoorenburg, G. F. (**1970**). "Pitch perception of two-frequency stimuli," J. Acoust. Soc. Am. **48**, 924–942.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Zhang, P. X., and Hartmann, W. M. (**2008**). "Lateralization of Huggins pitch," J. Acoust. Soc. Am. **124**, 3873–3887.

Sequential stream segregation using temporal periodicity cues in cochlear implant recipients[a)]

Robert S. Hong[b)] and Christopher W. Turner
Department of Otolaryngology–Head and Neck Surgery and Department of Communication Sciences and Disorders, 121B Wendell Johnson Speech and Hearing Center, University of Iowa, Iowa City, Iowa 52242-1012

(Received 19 June 2008; revised 1 May 2009; accepted 1 May 2009)

Sequential stream segregation involves the ability of a listener to perceptually segregate two rapidly alternating sounds into different perceptual streams. By studying auditory streaming in cochlear implants (CIs), one can obtain a better understanding of the cues that CI recipients can use to segregate different sound sources, which may have relevance to such everyday activities as the understanding of speech in background noise. This study focuses on the ability of CI users to use temporal periodicity cues to perform auditory stream segregation. A rhythmic discrimination task involving sequences of alternating amplitude-modulated (AM) noises is used. The results suggest that most CI users can stream AM noise bursts at relatively low modulation frequencies (near 80 Hz AM), but that this ability diminishes at higher modulation frequencies. Additionally, the ability of CI users to perform streaming using temporal periodicity cues appears to be comparable to that of normal-hearing listeners. These results imply that CI subjects may in certain contexts (i.e., when the talker has a low fundamental frequency voice) be able to use temporal periodicity cues to segregate and thus understand the voices of competing talkers.
 [DOI: 10.1121/1.3140592]

PACS number(s): 43.66.Mk, 43.66.Ts, 43.66.Lj, 43.66.Hg [RYL] Pages: 291–299

I. INTRODUCTION

Cochlear implants (CIs) have been remarkable in their ability to restore hearing to deaf individuals. Many CI users experience near-total to total understanding of everyday speech in quiet conditions. However, their ability to understand speech deteriorates significantly in conditions with background noise. The understanding of speech in background noise is one of the biggest challenges facing CI listeners today. One approach (the approach advocated by this study) to examining the factors that may be most relevant to improving the understanding of speech in noise by CI users is to examine different auditory cues individually in a laboratory setting—specifically, by using psychoacoustic tests that assess auditory stream segregation. Auditory stream segregation is the process used to separate a complex sound into different perceptual streams, corresponding to the different individual sources from which the sound is derived. The ability to stream (pure tones) has been shown to be correlated with the understanding of speech in noise in both hearing-impaired (Mackersie *et al.*, 2001) and CI listeners (Hong and Turner, 2006), consistent with the idea that a better ability to segregate sounds leads to a better understanding of speech in competing backgrounds. Auditory streaming has been studied extensively in normal-hearing (NH) listeners, with the suggestion that any salient difference between sounds can potentially be used as a cue for segregating sounds (Moore and Gockel, 2002). In contrast, few studies have been performed to date on auditory streaming in CI users (Hong and Turner, 2006; Chatterjee *et al.*, 2006; Cooper and Roberts, 2007), with these studies generally suggesting that streaming may occur with CIs but on a limited basis.

The study of sequential stream segregation is performed in the laboratory with sequences composed of two different sounds (sound A and sound B) that alternate rapidly in time. When these sounds are similar, they are often heard in a single perceptual stream fluctuating between sounds A and B. As the sounds become more different, this percept changes to that of two different streams, one composed of repeating sound A's and the other of repeating sound B's. The traditional method of assessing sequential stream segregation in NH subjects is to ask subjects to subjectively indicate whether they hear one or two streams when listening to a sequence of alternating sounds. While such a method may be practical for NH subjects, it is unclear if such a method can accurately assess streaming in CI subjects, since sound is perceived differently through a CI than through NH ears. Thus, CI subjects may not understand what is meant by hearing one stream versus two streams, since they may not be able to perceive or describe such a difference. As a result, they may respond inaccurately with respect to streaming, because they may use some other percept (unrelated to streaming) to guide their decisions.

The research in this study attempts to bypass the concerns associated with the previously described methodology by using rhythmic discrimination procedures to assess sequential stream segregation in CIs. Roberts *et al.* (2002) and

[a)]Portions of this work were presented in "Auditory stream segregation using temporal periodicity cues in cochlear implants," Midwinter Research Meeting of the Association for Research in Otolaryngology, Denver, CO, 2007.
[b)]Author to whom correspondence should be addressed. Electronic mail: robert-hong@uiowa.edu

Stainsby *et al.* (2004) compared results obtained using the traditional subjective method to assess streaming and one based on a rhythmic discrimination task, with similar results obtained from each methodology. With respect to CI subjects, there are two primary advantages to using rhythmic discrimination to measure streaming. First, the concept of rhythm (unlike that of streaming) is likely to be familiar to CI users, since the rhythmic perception ability of CI users is similar to that of NH subjects (Gfeller and Lansing, 1991). Thus, CI users will more likely use the appropriate perceptual criteria to perform the task. Second, a rhythmic discrimination streaming task has a clear right and wrong answer, unlike the earlier described tasks where listeners are asked to indicate if they think they hear one or two streams. This allows an adaptive procedure to be implemented that can converge on a particular streaming threshold, providing the increased efficiency and reliability enjoyed by adaptive procedures (Leek, 2001) to the proposed tasks.

The ability of CI recipients to use rate pitch cues (also referred to in this paper as temporal envelope or temporal periodicity cues) for sequential stream segregation is not known. Shannon (1992) suggested that CI users have a poor ability to detect temporal envelope cues greater than 300 Hz. Thus, one may expect that the ability of CI users to use rate pitch cues for streaming, if present, would be limited to cues less than 300 Hz. However, even a potentially limited range of temporal envelope cues to low frequencies (<300 Hz) for streaming by CI users is important to study. Low-frequency temporal envelope cues correspond to the fundamental frequencies of many human voices, and fundamental frequency is one of the most important cues for the segregation of speech from background noise, at least in NH listeners (Brokx and Nooteboom, 1982). The ability to stream using low-frequency rate pitch cues is particularly important to examine in CI users because this may be the dominant cue used by CI listeners for streaming. Because place pitch cues are relatively unreliable compared to rate pitch cues for conveying pitch in current CI users (Geurts and Wouters, 2001; Laneau *et al.*, 2004), rate pitch cues may take on a relatively more important role for sound segregation. Our study uses sinusoidal amplitude-modulated (AM) noise bursts to determine the extent to which CI recipients can use rate pitch cues for sequential stream segregation. If low-frequency rate pitch cues are usable by CI listeners for streaming, then this would suggest that future speech processing strategies designed to enhance these cues may improve the perception of human vocal fundamental frequency and speech in background noise.

In our previous study of sequential stream segregation in CI recipients (Hong and Turner, 2006), we utilized the method requiring subjects to detect an irregular rhythm described by Roberts *et al.* (2002) to successfully assess auditory streaming of pure tones in CI users. However, attempts to adapt this method to measure streaming with temporal envelope cues using AM noise bursts were problematic. It appeared that subjects were able to perform the task while ignoring the envelope differences for different AM noise bursts; the common timbre of the noise, which can be a strong cue for grouping (Dannenbring and Bregman, 1976; Cusack and Roberts, 2000), allowed subjects to integrate all the sounds in the same stream and hear the irregular rhythm, irrespective of AM rate. This observation is also consistent with that of Vliegen *et al.* (1999), who, using a similar rhythmic discrimination task to that used in our previous CI streaming experiments, found little evidence of streaming for complex tones with unresolved harmonics, presumably for a similar reason. Thus, it was necessary to develop a different method to assess the streaming with temporal periodicity cues in CI users, one that required the successful segregation of sounds into two different streams as opposed to integration of sounds into a single stream to perform the task. This methodology is similar to that used by Hartmann and Johnson (1991) and Dowling (1973) to assess streaming, except it is based on interleaved rhythms as opposed to interleaved melodies. The interleaved rhythmic discrimination task that we developed to assess the sequential stream segregation of AM noise bursts is described in Sec. II.

Two types of streaming experiments involving interleaved rhythmic discrimination are used in this study to probe the AM streaming abilities of CI users. The goal of the first type of experiment is to assess the AM streaming ability of CI users under "optimal" conditions—that is, when the temporal modulation cue should be the strongest. Shannon (1992) demonstrated that CI users are most sensitive to temporal modulations at lower modulation frequencies (80–100 Hz AM), and so a modulation frequency of 80 Hz AM is chosen as a base frequency for this experiment. Furthermore, all the stimuli used for this experiment are 100% modulated, since the ability to detect modulations decreases as modulation depth decreases. This experiment thus measures the ability of listeners to stream a fully-modulated 80 Hz AM stimulus apart from a fully-modulated AM stimulus at a higher AM rate to determine if CI users have any ability to stream AM stimuli.

The second type of streaming experiment is designed to provide insight into the AM frequencies at which temporal envelope cues can be used for streaming. This second type of experiment measures the modulation depth at which listeners are no longer able to stream an 80, 200, or 300 Hz AM stimulus apart from an unmodulated noise. If CI users are able to perform AM streaming, this ability may decrease as the modulation AM frequency is increased from 80 to 300 Hz AM, given the increasingly poor ability of CI users to detect modulation at AM frequencies higher than 300 Hz (Shannon, 1992). Extending this hypothesis to the real world setting of speech perception in noise, such a trend would suggest that temporal envelope cues may be more useful to CI users for segregating a male speaker (Fo=80–120 Hz) than a child speaker (Fo=approximately 300 Hz) from a competing background.

II. METHODS

A. Participants

10 CI and 12 NH subjects participated in these experiments. All of the CI listeners used the Hi-Resolution™ strategy, a strategy that allows transmission of the temporal periodicity cues of interest in these experiments, with an

TABLE I. CI subject demographics. The age, duration of profound deafness, length of CI use at time of testing, device, and stimulation rate for all CI subjects is shown. All of the subjects used the Hi-Resolution™ strategy.

Subject	Age (years)	Duration of deafness (years)	Length of CI use (years)	Device	Stimulation rate (pulses/s channel)
CI1	30	11	4.6	Clarion CII HF	2520
CI2	79	0.6	3.4	Clarion CII HF	2320
CI3	71	12	1.8	Clarion 90K	5156
CI4	46	1	3.4	Clarion CII HF	2047
CI5	64	10	3.2	Clarion CII HF	2320
CI6	68	1	3	Clarion CII HF	5156
CI7	41	36	4.3	Clarion 90K	5156
CI8	52	46	6.4	Clarion CII HF	5156
CI9	47	23	4.1	Clarion CII HF	4679
CI10	56	0.1	3	Clarion 90K	2184

Advanced Bionics CI (Sylmar, CA). Additionally, the rate of stimulation on each channel was required to be greater than 1200 pulses/s, to allow for the accurate representation of temporal periodicity information up to 300 Hz (McKay *et al.*, 1994; Wilson, 1997), as shown in Table I.

To minimize effects of non-streaming related factors on performance, musicians with college level musical training were excluded from testing, since these subjects may be extraordinarily good at rhythm-related tasks. Age may also have an effect on the ability of subjects to perform temporally-related tasks, with a decline in performance observed by some studies for older subjects (Pichora-Fuller, 2003; Strouse *et al.*, 1998) but not others (Takahashi and Bacon, 1992). To prevent age from playing a potentially confounding role in the assessment of streaming, the NH subjects were roughly age-matched to the CI users; the average age of NH listeners is 47.9 years and CI listeners is 55.4 years.

Pure-tone audiometry was performed on all NH listeners to confirm that they had NH. All subjects had NH (<20 dB HL) across octave frequencies (from 250 to 8000 Hz), with the exception of one listener who had a 40 dB HL at 8000 Hz and a second listener who had a 25 dB HL at 4000 Hz.

B. Stimuli and procedures

1. Interleaved rhythmic discrimination streaming task

In this study, a three-interval two-alternative forced choice, interleaved rhythmic discrimination task was used to assess auditory streaming based on temporal envelope cues. In this task, subjects were presented a target rhythm and asked to identify which of two subsequent sequences of alternating sounds contained the target rhythm. If sound A and sound B were sufficiently different along any perceptual dimension, allowing listeners to stream sound A apart from sound B, then listeners were able to hear the target rhythm (composed of sound A's) in one of the two sequences (Fig. 1). Incidentally, the listener could either listen to the sound A's alone or the sound B's alone to hear the target rhythm, since the rhythm could be found among either sound, even though the target rhythm was played for the listener during the task only with sound A.

All of the stimuli used for sound A and sound B were AM, broadband noise bursts. Each noise burst was 125 ms in duration, corresponding approximately to the duration of syllables in speech, with 10 ms linear ramps at the onset and offset of each stimulus. The starting phase of sinusoidal amplitude modulation was random for each noise burst. The target rhythm was composed of four sound A's, separated by either a long silence (175 ms duration) or a short silence (25 ms duration), as shown in Fig. 1. Each of the two sequences was composed of a unique arrangement of eight sounds (four sound A's and four sound B's), with each sound separated by 25 ms of silence, as shown in the figure. The difference between the two sequences is denoted by the arrows, where sound A and sound B are reversed. The figure shows each sequence to be composed of a set of eight sounds played two times in a row, for a total duration of 2.4 s. In the actual task, each set of eight sounds was played four times in a row, such that each sequence actually lasted 4.8 s (twice as long as shown in Fig. 1). If the listener could perceptually stream sound A apart from sound B, the listener should be able to hear that sequence 2 contained the target rhythm (shown in bold and underlined in Fig. 1).

The stimuli were presented in a double-walled sound attenuated booth through a single loudspeaker situated directly in front of the listener at 72 dB SPL, with a 3 dB loudness rove from one sound to the next within each sequence. The root-mean-square power of each noise burst was equated before applying the loudness rove. The loudness rove was used to prevent listeners from using loudness cues, as opposed to rate pitch cues, to stream AM noise bursts, with a 3 dB range chosen to cover the range of different loudness that may be perceived by subjects for stimuli of different AM rates (Zhang and Zeng, 1997). All stimuli were stored at 44.1 kHz and passed through a low-pass 20 kHz filter before presentation to avoid aliasing. Feedback was provided on all tests.

To address the possibility that some subjects may not be able to perform the AM streaming task because of difficulties with discriminating rhythms in general, irrespective of streaming, all subjects were first screened with an initial task that required subjects to correctly identify the target rhythm in the absence of an interleaved rhythm. This test was iden-

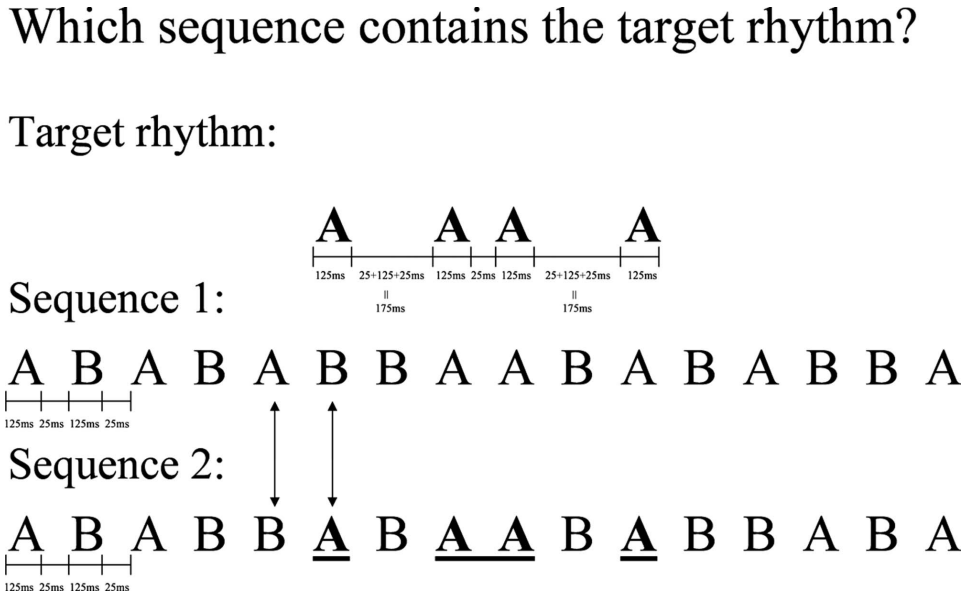


FIG. 1. Interleaved rhythmic discrimination task. The target rhythm is composed of four sound A's played in the rhythm shown. Each sequence is composed of a unique arrangement of eight sounds (four sound A's and four sound B's), which in the figure are repeated twice. The difference between the two sequences is denoted by the arrows, where sound A and sound B are reversed. If the listener can perceptually stream sound A apart from sound B, the listener will be able to hear that sequence 2 contains the target rhythm (shown in bold and underlined).

tical to the interleaved rhythm task, except all of the sound B's in the streaming task (as shown in Fig. 1) were replaced with silent gaps equal in duration to sound B. To pass the basic rhythm discrimination screening task, subjects needed to correctly identify the sequence that matched the target rhythm in five consecutive trials. Subjects that could not perform the simple rhythm discrimination task were excluded from further testing. In fact, one NH and one CI subject (not shown in Table I) did not pass the simple rhythm discrimination task and thus did not participate in any of the streaming experiments.

The streaming of AM noise bursts was assessed using four different conditions. In the first condition, the ability to stream was measured as a function of separation in amplitude-modulation rate between sound A and sound B, with both sounds fully modulated. Sound A was always an 80 Hz AM noise burst. Sound B was always a noise burst with a higher rate of AM compared to sound A, with sound B varied in logarithmic steps (sound B $=80.5, 81, 82, 84, \ldots, 1104$ Hz AM) to determine the AM frequency separation needed for auditory stream segregation. In the other three conditions, the ability to stream was measured as a function of modulation depth between a modulated sound A (fixed at 80, 200, or 300 Hz AM) and an unmodulated sound B. Sound A was varied in linear steps of modulation depth to determine the modulation depth needed to stream a modulated AM noise burst from an unmodulated one. In all four conditions, a two-down one-up adaptive staircase that converged on the 70.7% correct point of the psychometric function was used to determine the difference needed between sound A and sound B that resulted in a detectable target rhythm. In the first condition, where the task adapted based on AM frequency, there were eight total reversals in each run. The first two reversals were taken as practice and had a multiplicative step size of 4. The geometric mean of the final six reversals, with a multiplicative step size of 2, was taken as the threshold for a single run. In the other three conditions, where the task adapted based on modulation depth, there were ten total reversals in each run. The first four reversals were taken as practice, with a linear step size of 20% modulation depth for the first two steps and 10% modulation depth for the next two steps. The arithmetic mean of the final six reversals, with a linear step size of 4% modulation depth, was taken as the threshold for a single run. In each of the four conditions, a total of four consecutive runs were performed, with the mean of the final three runs taken as the threshold for each condition for each subject. Additionally, during the first practice run, it was sometimes necessary to introduce a loudness cue of up to 20 dB to sound A relative to sound B, to help the subjects understand the task. This cue was incrementally reduced until no longer present during the practice run, and it was not present in any of the subsequent three runs, where actual data were collected. Finally, subjects who exceeded the upper bound of the task after completing the required practice run for a particular condition were recorded as being unable to perform the task at that condition.

2. Amplitude-modulation discrimination tasks

Each of the four conditions tested in the AM interleaved rhythmic discrimination experiments had a corresponding AM discrimination test condition. The AM discrimination task, like the interleaved rhythm task, is a three-interval two-alternative forced choice task that adapts based on a two-down one-up rule. In this task, three consecutive AM noise bursts were played. Listeners were asked to decide if the second or third noise burst sounded different than the first, ignoring differences in loudness (since a 3 dB loudness rove was implemented among the noise bursts). The AM discrimi-

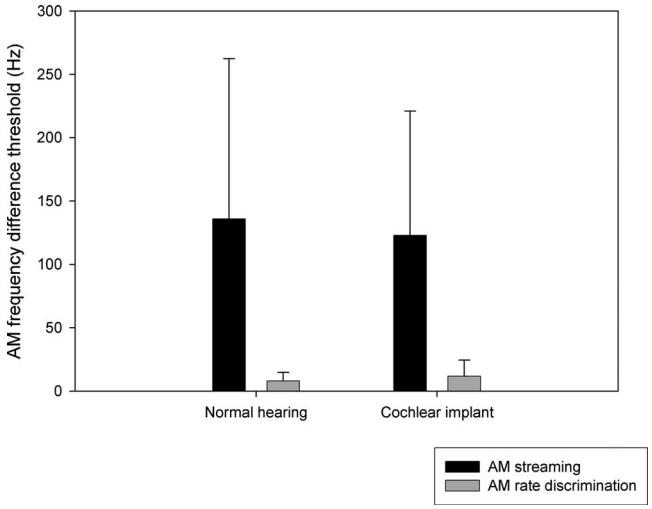


FIG. 2. Group performance on amplitude-modulation streaming and rate discrimination for tasks adapting based on AM frequency. The average results of the NH and CI groups are shown. All AM stimuli are fully modulated. The base frequency (sound A) is 80 Hz AM. The results demonstrate the threshold difference in AM frequency between sound A and sound B (with sound B always greater than 80 Hz AM) needed to stream or discriminate the two sounds. Error bars represent one standard deviation above the mean.

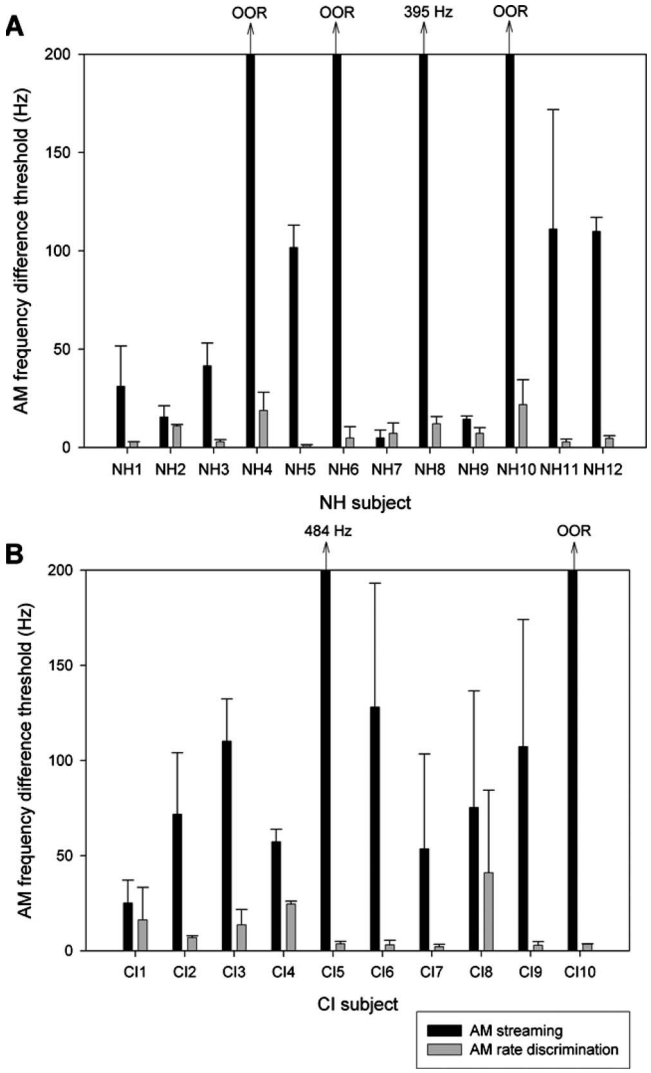


FIG. 3. Individual performance on amplitude-modulation streaming and rate discrimination for tasks adapting based on AM frequency. The results of each of the (a) 12 NH and (b) 10 CI listeners are shown. All AM stimuli are fully modulated. The base frequency (sound A) is 80 Hz AM. The results demonstrate the threshold difference in AM frequency between sound A and sound B (with sound B always greater than 80 Hz AM) needed to stream or discriminate the two sounds. Error bars represent one standard deviation above the mean. "OOR" indicates that the results are "out of range," indicating that the listener is unable to perform the task, even at the largest AM frequency differences. Error bars represent one standard deviation above the mean.

nation task that corresponded to the AM interleaved rhythm task adapting on AM frequency can be thought of as an AM rate discrimination task, where the base frequency was a fully-modulated 80 Hz AM noise burst. The three AM discrimination tasks that corresponded to the three AM interleaved rhythm tasks adapting on modulation depth can be thought of as three different AM detection experiments, with base frequencies of 80 Hz AM, 200 Hz AM, and 300 Hz AM, respectively. The composition of the different AM noise bursts, the step sizes, and the number of reversals used were identical to those found in the corresponding AM interleaved rhythmic discrimination task, as this allowed for a direct comparison of the ability to discriminate between AM noise bursts versus the ability to use such a perceptual difference between AM noise bursts for auditory stream segregation.

III. RESULTS

The average results for NH and CI subjects on the AM streaming task between fully-modulated AM noise bursts are shown in Fig. 2. In this task, sound A was always an 80 Hz AM noise burst. Sound B was a noise burst of a higher AM frequency, and thus the AM frequency difference threshold was the number of Hz above 80 Hz AM needed to be able to perform the task. As the figure demonstrates, the abilities of NH and CI subjects to stream AM stimuli on this task are comparable ($t_{df=20}=0.264$; $p>0.5$). However, there was a wide range of abilities among different individuals of both groups for AM streaming (Fig. 3). Some were excellent at streaming AM stimuli (e.g., NH 7 or CI1), while others could not perform the task at all (e.g., NH 4, NH 6, NH 10, and CI 10). For purposes of statistical analysis, all values for the AM frequency difference threshold above 300 Hz, including those where subjects were not able to perform the task (out of range=OOR), were taken as 300 Hz. This statistical cutoff was chosen since 300–400 Hz AM corresponds to the range reported by others where the ability to detect amplitude modulation, and thus presumably to use such cues for streaming, is severely limited for both NH and CI listeners (Shannon, 1992). (This poor ability to use a 300 Hz AM temporal cue for streaming was confirmed by other experiments in this study, as will be shown later in Fig. 4.)

Figure 2 also demonstrates that the ability of both NH and CI subjects to perform AM rate discrimination was significantly better than the ability to perform AM streaming (for NH, $t_{df=11}=3.604$, $p<0.005$; for CI, $t_{df=9}=3.378$, $p<0.01$). This suggests that it was not the ability to detect differences in AM rate that limited the ability of listeners to perform AM streaming. An analysis of the individual data shown in Fig. 3 further confirmed this point. All four sub-

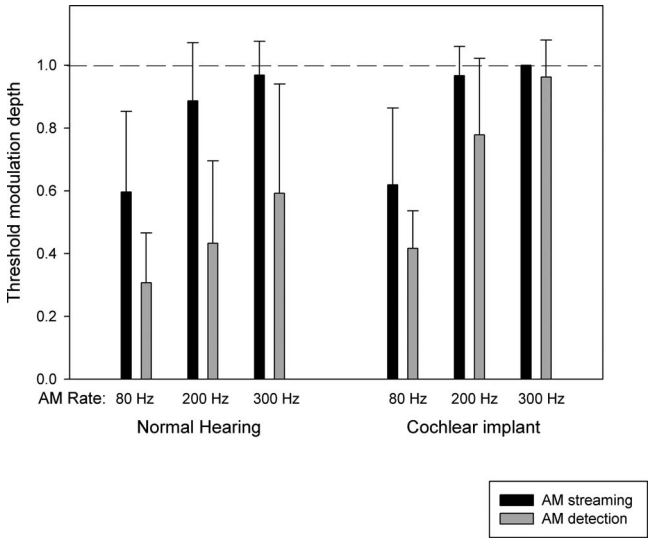


FIG. 4. Group performance on amplitude-modulation streaming and detection for tasks adapting based on modulation depth. The average results of the NH and CI groups are shown for stimuli at 80, 200, or 300 Hz amplitude modulation. The results demonstrate the threshold modulation depth for sound A that is needed to stream or discriminate it apart from an unmodulated sound B at each of these three AM rates. Error bars represent one standard deviation above the mean.

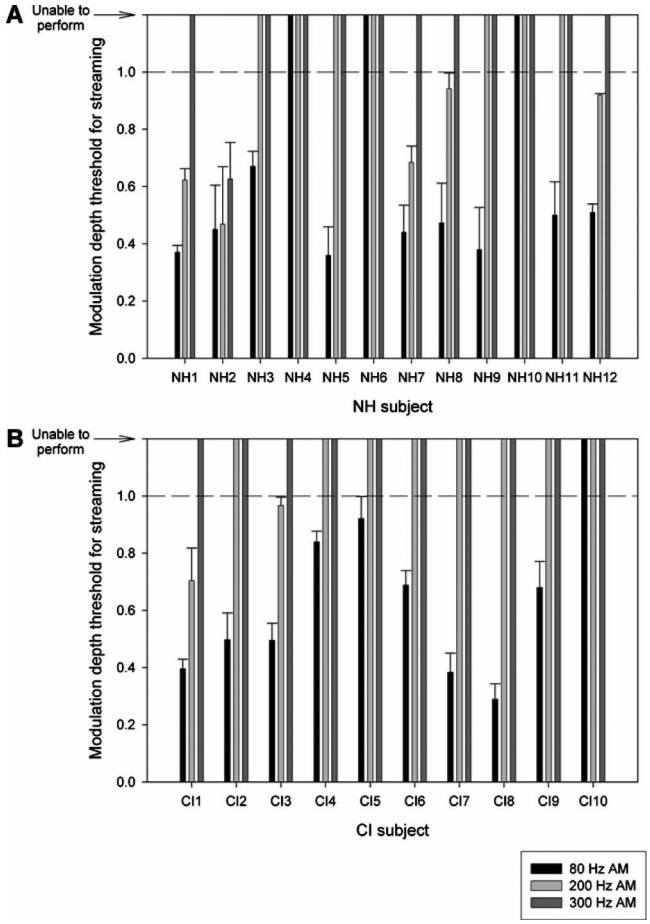


FIG. 5. Individual performance on amplitude-modulation streaming for tasks adapting based on modulation depth. The individual streaming results of each of the (a) 12 NH and (b) 10 CI listeners are shown for stimuli at 80, 200, or 300 Hz amplitude modulation. The results demonstrate the threshold modulation depth for sound A that is needed to stream it apart from an unmodulated sound B at each of these three AM rates. Values extending beyond the horizontal dotted line indicate that the subject was not able to perform the task for that particular condition.

jects who could not perform the AM streaming task (NH 4, NH 6, NH 10, and CI 10) were able to perform the AM rate discrimination task with relative ease.

The average results of both NH and CI subjects for AM streaming at different AM base frequencies as a function of modulation depth are shown in Fig. 4. There was no significant difference between NH and CI subjects on AM streaming at base frequencies of 80 Hz AM, 200 Hz AM, or 300 Hz AM ($p > 0.2$ for each of the three conditions). However, a repeated-measures analysis of variance (ANOVA) demonstrated that there was a significant difference within groups across base frequencies for both NH ($F_{2,22}=19.009$; $p < 0.001$) and CI ($F_{2,18}=23.46$; $p < 0.001$) listeners. Post-hoc pair-wise comparisons (Bonferroni) showed that for NH listeners, AM streaming at 80 Hz AM was significantly better than at 200 Hz AM ($p < 0.005$) and also than at 300 Hz AM ($p < 0.001$); however, AM streaming at 200 Hz AM was not significantly better than at 300 Hz AM ($p=0.165$). For CI subjects, AM streaming results were similar to those of NH listeners, with streaming at 80 Hz AM significantly better than at 200 Hz AM ($p < 0.005$) and also than at 300 Hz AM ($p < 0.005$), and no significant difference between streaming at 200 Hz AM and 300 Hz AM ($p > 0.5$).

The individual AM streaming results for NH and CI listeners at different base frequencies are shown in Fig. 5. Within both groups, there was a range of streaming abilities. At a base frequency of 80 Hz AM, the majority of both NH and CI listeners were able to perform streaming as a function of modulation depth (9 out of 12 NH and 9 out of 10 CI subjects). In contrast, at a base frequency of 300 Hz AM, only one listener (NH 2) out of all the subjects tested (either NH or CI) was able to perform AM streaming. The ability to perform streaming at a base frequency of 200 Hz AM varied among both NH and CI subjects.

Comparison of AM streaming versus AM detection abilities at different base frequencies suggests that both NH and CI subjects were able to detect smaller differences in modulation depth than they were able to use for streaming (Fig. 4). Paired t-tests demonstrated this to be true for NH subjects across each of the three base frequencies ($p < 0.005$), although for CI subjects this was the case only at 200 Hz AM ($p < 0.05$), though such a difference approached significance at 80 Hz AM ($p=0.064$). Analyses of the individual results for AM streaming (Fig. 5) and for AM detection (Fig. 6) further highlight the point that the ability to stream AM stimuli did not appear to be limited by the ability to detect AM stimuli. For example, 8 out of the 11 NH listeners who could not perform AM streaming at 300 Hz AM could perform AM detection at 300 Hz AM, while for CI listeners, the same held true for 3 out of the 8 CI listeners at 200 Hz AM. Finally, it is noted that the lack of significant difference between streaming and detection for CI listeners at 300 Hz AM was because most CI subjects could neither detect [Fig. 6(b)] nor stream at this base frequency [Fig. 5(b)]; for these subjects at this test condition, it is unclear if

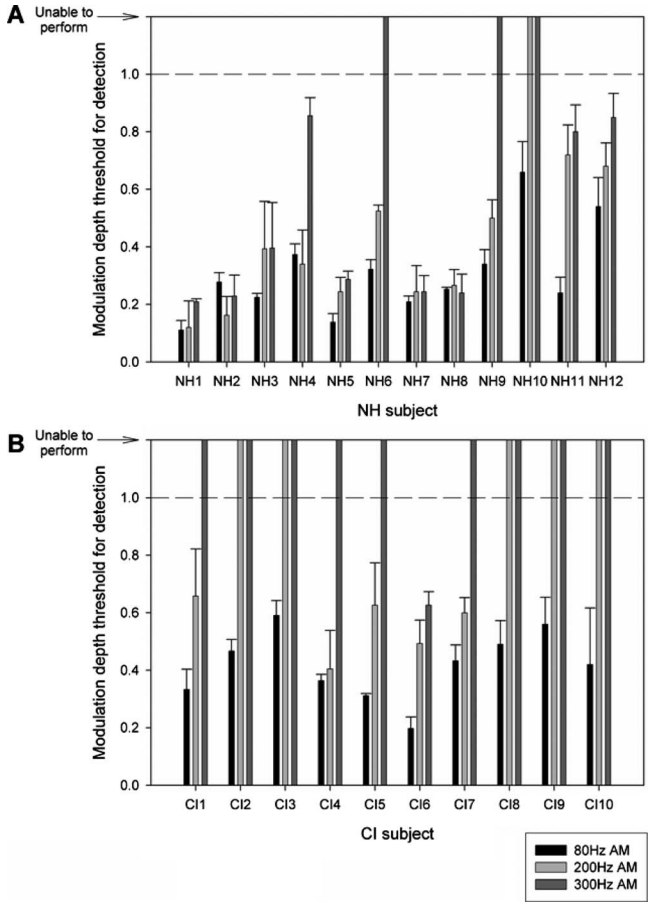


FIG. 6. Individual performance on amplitude-modulation detection for tasks adapting based on modulation depth. The individual results of each of the (a) 12 NH and (b) 10 CI listeners are shown for stimuli at 80, 200, or 300 Hz amplitude modulation. The results demonstrate the threshold modulation depth needed to detect the amplitude modulation at each of these three AM rates. Values extending beyond the horizontal dotted line indicate that the subject was not able to perform the task for that particular condition.

their ability to stream was limited by their ability to discriminate, since they could not perform either task.

Correlations were also performed for NH and CI subjects between age and each of the measures of AM streaming/discrimination/detection. It has been reported that age may play a role in performance on temporally-based tasks (e.g., Strouse *et al.*, 1998). Because the age range of the listeners from each of the two groups was wide (NH =22–68 years old; CI=30–79 years old), correlations could be performed to examine if age also affected the results of the tasks in this study, which involved temporal modulations of the envelope. The results demonstrated significant positive correlations for NH subjects between age and AM streaming thresholds at 200 Hz base frequency (r=0.809; p<0.001), AM detection thresholds at 200 Hz (r=0.723; p<0.01), and AM detection thresholds at 300 Hz (r=0.683; p<0.02). No correlation was expected or found for AM streaming at 300 Hz base frequency, since only 1 out of 12 NH subjects could perform the task, though interestingly, the one who could perform the task was the youngest subject tested. None of the other correlations were significant for NH subjects. Additionally, no significant correlations were found for the CI group.

IV. DISCUSSION

The first objective of this study is to determine whether or not CI recipients can use temporal envelope cues for auditory stream segregation. The results from these experiments (Figs. 2–5) demonstrate that the large majority of CI recipients (night out of ten subjects) can to some extent use such cues for streaming. This is the first study to the knowledge of the authors that has demonstrated that CI users can use temporal periodicity cues for auditory stream segregation. Additionally, these results suggest that at least some of the ability of CI recipients to perform pure-tone auditory stream segregation (as observed by Hong and Turner, 2006), specifically in the lower frequency regions (less than about 200 Hz), can be attributed to their ability to use temporal periodicity cues for streaming.

The next objective, given that most CI users could perform auditory streaming with temporal periodicity cues, is to determine how the AM streaming abilities of CI listeners compare to those of NH listeners. The results suggest that as a group, the AM streaming abilities of NH and CI subjects were similar (Figs. 2 and 4), though there was considerable variation among individuals of each group. Such a result is consistent with other findings in literature, which have suggested that, in general, the temporal perception abilities of CI recipients are similar to those of NH listeners (Moore and Glasberg, 1988; Shannon, 1989, 1992).

Another question of interest is to determine the extent to which AM cues may be useful for segregating sounds in situations where such cues are degraded (i.e., when the modulation depth is not 100%). Temporal envelope cues in the real world, such as those corresponding to the fundamental frequency of a talker, are often degraded by background noise and reverberation. In examining the abilities of CI users to perform streaming using AM cues, it is clear that listeners are better able to stream using degraded AM cues at lower AM frequencies (80 Hz AM) than at higher AM frequencies (200 Hz AM), and that this ability is essentially nonexistent at the highest AM frequencies tested (300 Hz AM). This finding suggests that temporal envelope cues may be more useful in the real world for segregating the voice of a male (lower Fo) than a female or child (higher Fo) from competing backgrounds. However, further research is needed to establish the relevance of our findings to the real world setting, because the perceived strength of temporal periodicity cues represented via broadband AM noise bursts (as used in this study) may be stronger than that found in speech signals with a more limited bandwidth. Thus, the extent to which our results can be extrapolated to the streaming of limited bandwidth speech remains to be determined.

The results of this study suggest that the age of listeners may influence performance on AM detection and AM streaming tasks; significant correlations among NH listeners between age and AM detection/streaming at higher AM rates (200 and 300 Hz AM) were observed, with older subjects performing more poorly on these tasks. With respect to AM detection, our findings differ from those of Takahashi and Bacon (1992), which suggested that age was largely unrelated to AM detection; we do not have a good explanation for

these differences. With respect to AM streaming, we suspect that the observed effects of age may be due to the temporal nature of the rhythmic discrimination task used in this study, since performance on temporally-based tasks decreases with age (Trainor and Trehub, 1989; Strouse *et al.*, 1998). An alternative interpretation is that auditory streaming ability with temporal periodicity cues diminishes with age; however, this is less likely, since a number of studies have suggested that sequential stream segregation is preserved in older adults (Trainor and Trehub, 1989; Snyder and Alain, 2007). It is important to note, however, that age does not factor into the primary comparison made in this study (i.e., examining the ability of NH versus CI listeners to perform AM streaming), since the average age of each group was controlled to be similar (47.9 versus 55.4 years old).

Grimault *et al.* (2002) measured the ability of NH listeners to use AM cues for auditory streaming. The test paradigm and stimuli used by Grimault *et al.* (2002) were different than those of this study, as they used a traditional test paradigm asking listeners to judge whether they heard one or two streams. Nevertheless, the results of the two studies appear to be relatively consistent with respect to the ability of NH listeners to perform AM streaming, providing evidence that the interleaved rhythmic discrimination task provides a valid assessment of streaming ability. Grimault *et al.* (2002) found that for fully-modulated stimuli, the streaming threshold occurred when sound A was 100 Hz AM and sound B was roughly 200 Hz AM. Similarly, in this study, the streaming threshold for fully-modulated stimuli occurred when sound A was 80 Hz AM and sound B was roughly 215 Hz AM (or 163 Hz AM, if you exclude the results of the three NH subjects who could not perform the task). Additionally, both Grimault's study and this study demonstrate that performance on streaming worsened with shallower modulation depths, when amplitude-modulation cues are weaker.

Comparisons between AM rate discrimination (and AM detection) thresholds with AM streaming thresholds obtained for both NH and CI listeners suggest that for both groups of listeners, it is easier to discriminate (or detect) AM stimuli than it is to stream them. In fact, in a number of cases, listeners were able to perform the discrimination/detection task but not the streaming task with the same stimuli (Figs. 3, 5, and 6). The finding that discrimination/detection thresholds are smaller than streaming thresholds is consistent with those of other authors. For example, Grimault *et al.* (2002) observed that NH listeners had smaller thresholds on AM rate discrimination and AM detection tasks than AM streaming tasks. Similarly, Rose and Moore (2005) observed that both NH and hearing-impaired listeners had smaller thresholds for pure-tone frequency discrimination than pure-tone streaming. The observation that streaming thresholds are larger than discrimination/detection thresholds suggests that the ability to stream is not necessarily limited by the ability to discriminate stimuli, implying that auditory stream segregation involves more complex neural processing than that needed for auditory discrimination.

A recent study by Gaudrain *et al.* (2008) examined the ability of listeners to perform obligatory stream segregation of vowel sequences using noise-excited envelope vocoders simulating CIs. Obligatory stream segregation is measured with test paradigms that require integration of rapidly alternating sounds into a single stream to successfully perform the task. They did not find any convincing evidence of obligatory streaming using temporal periodicity cues in these CI simulations. Our results differ from those of Gaudrain *et al.* (2008) in that our results suggest that CI listeners can use temporal periodicity cues for streaming, at least up to ~200 Hz. There are a few possible reasons for this discrepancy between our results and those of Gaudrain *et al.* (2008). First, our study measures schema-based stream segregation as opposed to obligatory stream segregation—that is, the test paradigm used in our study requires segregation of sounds into two different streams as opposed to integration of sounds into a single stream to complete the task. Thresholds for schema-based versus obligatory stream segregation are known to differ, as they represent two distinct boundaries of sequential streaming (van Noorden, 1975), and thus it is not theoretically inconsistent that temporal periodicity cues may be useful for CI recipients for schema-based but not obligatory stream segregation. Second, our study uses sinusoidal AM noise bursts while Gaudrain *et al.* (2008) utilized noise-vocoder vowels. It is possible that temporal periodicity cues are stronger in AM noise bursts than in noise-vocoder vowels, leading to the different conclusions of the two studies. Finally, our study examines CI recipients directly while Gaudrain *et al.* (2008) used CI simulations in NH individuals. While CI simulations are useful for obtaining a rough estimate of CI performance, they generally do not reflect the wide range of listening abilities found in actual CI users, leading to a third possible reason for the differing results of the two studies.

The results of this study suggest that temporal periodicity cues may be helpful for CI recipients in segregating sounds. It is important to note, however, that the CI users that participated in these tasks were carefully selected to all use the Hi-Res strategy with stimulation rates of at least 1200 pulses/s per channel—conditions that ensured that temporal periodicity cues were available and accurately presented to CI listeners. It is unlikely that all current CI recipients would be able to take advantage of AM cues for auditory stream segregation, because of the following two possible limitations to their implant. First, these cues are not even transmitted by some CIs; for example, in some Nucleus signal-processing strategies, the low-pass cutoff for the temporal envelope is effectively 125 Hz, preventing rate pitch information above 125 Hz from being transmitted by the implant (Bom Jun Kwon, Cochlear Corporation, personal communication). Second, even if temporal envelope cues are transmitted, they are often distorted as many common signal-processing strategies do not provide electrical stimulation at fast enough rates to accurately convey the full range of perceivable rate pitch cues [i.e., up to ~300 Hz rate pitch (Eddington *et al.*, 1978; Shannon, 1983; Zeng, 2002)]. It has been demonstrated that the electrical carrier pulse rate must be at least four to five times the frequency of the desired rate pitch for the temporal envelope cue to be accurately represented by electrical stimulus (McKay *et al.*, 1994; Wilson, 1997), suggesting that a carrier rate of at least 1200 pulses/

s channel (=4 × 300 Hz) is necessary. The results of these experiments encourage signal-processing strategies to include more information in the temporal domain, as this may be helpful for segregating sounds in the real world, potentially leading to better speech perception in background noise by CI recipients.

V. CONCLUSIONS

A novel adaptive rhythmic discrimination task requiring successful segregation of sounds into two different streams to assess sequential stream segregation is described in this study.

This is the first demonstration to our knowledge that CI listeners can utilize purely temporal cues for sequential stream segregation.

CI and NH listeners demonstrate similar abilities to use temporal periodicity cues for streaming.

The ability of both CI and NH listeners to stream using temporal periodicity cues is poorer than their ability to detect/discriminate such cues. This suggests that CI signal-processing strategies designed to improve speech perception in noise need to consider not only improvements in signal detection but also signal streaming, since listeners in some instances can detect certain cues but not use such cues for sound segregation.

Improvements in the encoding by CIs of temporal periodicity cues up to ~200 Hz may be useful for CI listeners in segregating sounds and, by extension, for understanding speech in background noise.

ACKNOWLEDGMENTS

This research was supported by Grant Nos. R01 DC000377, P50 DC00242, and 5 T32 DC000040-13 from the National Institutes of Health. The authors thank Anna Hong for help in collecting the data, and Arik Wald for help with programming the experiments.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (**2006**). "Auditory stream segregation with cochlear implants: A preliminary report," Hear. Res. **222**, 100–107.

Cooper, H. R., and Roberts, B. (**2007**). "Auditory stream segregation of tone sequences in cochlear implant listeners," Hear. Res. **225**, 11–24.

Cusack, R., and Roberts, B. (**2000**). "Effects of differences in timbre on sequential grouping," Percept. Psychophys. **62**, 1112–1120.

Dannenbring, G. L., and Bregman, A. S. (**1976**). "Stream segregation and the illusion of overlap," J. Exp. Psychol. Hum. Percept. Perform. **2**, 544–555.

Dowling, W. L. (**1973**). "The perception of interleaved melodies," Cogn. Psychol. **5**, 322–337.

Eddington, D. K., Dobelle, W. H., Brackman, D. E., Mladejovsky, M. G., and Parkin, J. L. (**1978**). "Auditory prostheses research with multiple channel intracochlear stimulation in man," Ann. Otol. Rhinol. Laryngol. **87**, 1–59.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J.-C. (**2008**). "Streaming of vowel sequences based on fundamental frequency in cochlear-implant simulations," J. Acoust. Soc. Am. **124**, 3076–3087.

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.

Gfeller, K. E., and Lansing, C. (**1991**). "Melodic, rhythmic, and timbral perception of adult cochlear implant users," J. Speech Hear. Res. **34**, 916–920.

Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). "Auditory stream segregation on the basis of amplitude-modulation rate," J. Acoust. Soc. Am. **111**, 1340–1348.

Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**, 155–184.

Hong, R. S., and Turner, C. W. (**2006**). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," J. Acoust. Soc. Am. **120**, 360–374.

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.

Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.

Mackersie, C. L., Prida, T. L., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **44**, 19–28.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," J. Acoust. Soc. Am. **83**, 1093–1101.

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acust. Acta Acust. **88**, 320–333.

Pichora-Fuller, M. K. (**2003**). "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," Int. J. Audiol. **42**, S59–S67.

Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am. **112**, 2074–2085.

Rose, M. M., and Moore, B. C. J. (**2005**). "The relationship between stream segregation and frequency discrimination in normal hearing and hearing-impaired subjects," Hear. Res. **204**, 16–28.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man: I. Basic Psychophysics," Hear. Res. **11**, 157–189.

Shannon, R. V. (**1989**). "Detection of gaps in sinusoids and biphasic pulse trains by patients with cochlear implants," J. Acoust. Soc. Am. **85**, 2587–2592.

Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.

Snyder, J. S., and Alain, C. (**2006**). "Sequential auditory scene analysis is preserved in normal aging adults," Cereb. Cortex **17**, 501–512.

Stainsby, T. H., Moore, B. C. J., Medland, P. J., and Glasber, B. R. (**2004**). "Sequential streaming and effective level differences due to phase-spectrum manipulations," J. Acoust. Soc. Am. **115**(4), 1665–1673.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Wesley, D. G. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.

Trainor, L. J., and Trehub, S. E. (**1989**). "Aging and auditory temporal sequencing— Ordering the elements of repeating tone patterns," Percept. Psychophys. **45**, 417–426.

van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, Netherlands.

Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am. **106**, 938–945.

Wilson, B. S. (**1997**). "The future of cochlear implants," Br. J. Audiol. **31**, 205–225.

Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

Zhang, C., and Zeng, F.-G. (**1997**). "Loudness of dynamic stimuli in acoustic and electric hearing," J. Acoust. Soc. Am. **102**, 2925–2934.

Detection of the break in interaural correlation is affected by interaural delay, aging, and center frequency

Ying Huang, Xihong Wu, and Liang Li[a)]
Department of Psychology, Department of Machine Intelligence, Speech and Hearing Research Center, and Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871, People's Republic of China

(Received 4 May 2008; revised 4 May 2009; accepted 11 May 2009)

This study investigated whether interaural integration is affected by introducing an interaural delay. In Experiment 1, both younger adults with normal hearing and older adults in the early stages of presbycusis were able to detect a transient break in interaural correlation (BIC) in the temporal middle of interaurally correlated wideband noises. However, their duration thresholds for detecting the BIC became larger with increasing interaural time difference (ITD) from 0 to 6 ms, and the threshold increase for older participants was larger than that for younger participants. In Experiment 2, to investigate whether the effect of changing ITD on the BIC detection is frequency-dependent, 1/3-octave narrowband noises with various center frequencies were used as stimuli. Results show that the duration threshold for detecting the BIC was higher for high-frequency noises than for low-frequency noises. Also, with increasing ITD from 0 to 4 ms, the threshold increase was larger for high-frequency noises than for low-frequency noises. The results suggest that there are age- and frequency-related temporal declines in maintaining fine-structure signals for interaural integration. These declines may affect the recognition of sound sources in reverberant environments.
 [DOI: 10.1121/1.3147504]

PACS number(s): 43.66.Pn, 43.66.Lj, 43.66.Mk, 43.66.Qp [MW] Pages: 300–309

I. INTRODUCTION

In reverberant environments, listeners receive both the direct sound waves from sound sources and numerous reflections from various surfaces. To perceptually separate the target sound signal from irrelevant sound signals in such conditions, the auditory system needs to not only perceptually group direct sound waves emanating from the target source with reflections of the target source but also perceptually group direct sound waves emanating from an irrelevant source with reflections of the irrelevant source by both calculating correlations between sound waves and integrating highly correlated sound waves. In humans, the integration of the direct source wave with its reflections causes a perceptual consequence called fusion. That is, when the time interval between the sound and a delayed copy of the sound is sufficiently short, attributes of the lagging sound are perceptually captured by the leading sound (Li *et al.*, 2005), causing a single fused sound image that is perceived to be at or near the location of the leading sound. This phenomenon is generally named as the "precedence effect" or "the law of the first wavefront" (Freyman *et al.*, 1991; Litovsky and Shinn-Cunningham, 2001; Wallach *et al.*, 1949; Zurek, 1980; for a review see Litovsky *et al.*, 1999). The precedence effect plays a role in suppressing the perception of echoes and facilitating the recognition and localization of sources in reverberant environments. If the delay between the leading sound and the correlated lagging sound is sufficiently large, listeners will perceive a second sound image near the location of the lagging source. The minimum delay which allows a listener to distinctly perceive the lagging sound is called the echo threshold (e.g., Haas, 1951; Litovsky *et al.*, 1999; Rakerd *et al.*, 2000).

Processing the similarity and dissimilarity of sound waves arriving at the two ears is critical for the precedence effect and other binaural perceptual phenomena (Blauert and Divenyi, 1988; Litovsky *et al.*, 1999; Scharf, 1974; Shinn-Cunningham *et al.*, 1995; Trahiotis *et al.*, 2005; Yang and Grantham, 1997a, 1997b), and this interaural integration is important for auditory perception in noisy, reverberant environments (Bregman, 1990). Human listeners with normal hearing are very sensitive to small differences between a wideband noise delivered at one ear and its copy delivered at the other ear (Akeroyd and Summerfield, 1999; Boehnke *et al.*, 2002; Gabriel and Colburn, 1981; Goupell and Hartmann, 2006; Pollack and Trittipoe, 1959). Changing the interaural correlation[1] of wideband noises modifies the percept of the noises (Blauert and Lindemann, 1986; Hall *et al.*, 2005). For example, as reported by Blauert and Lindemann (1986), when the interaural correlation of wideband pink noises was 1, listeners perceived a single compact auditory event precisely localized in the middle of the head. When the interaural correlation was 0, listeners perceived two respective events, one at each ear. When the interaural correlation was 0.25, 0.50, or 0.75, listeners perceived one diffused event in the median plane, and two additional ones lateralized symmetrically with respect to the median plane. Thus, some perceptual dimensions of the sounds such as the compactness, number of images, and lateral position depend on the interaural correlation.

[a)]Author to whom correspondence should be addressed. Electronic mail: liangli@pku.edu.cn

The perceptual representation of interaurally correlated noises is also determined by the delay time between the two ears [interaural time difference (ITD)]. If identical steady-state wideband noises are presented at the two ears with the ITD of 0 ms, a single compact noise image is perceived at the middle point inside the head of a normal-hearing listener. When an ITD shorter than 1 ms, e.g., 0.5 ms, is introduced, the image is located between the middle of head and the leading ear. When the ITD is increased to 1 ms, the image is perceived at the leading ear. With a further increase in the ITD to a higher value within the range of the precedence effect, e.g., 4 ms, a single fused noise image is still located at the leading ear. Theoretically, when an ITD of several milliseconds is introduced, fine-structure information of the noise at the leading ear has to be maintained in the central auditory system for that period of time; otherwise, instead of one single fused image, multiple images would be perceived. Moreover, if there is a temporal decay of the central representation of fine-structure details of the noise, especially for that of high-frequency components, the binaural integration would decline with increasing ITD, even though the perceptual fusion is still maintained. Since processing fine-structure information is largely based on phase locking of neural firing and phase locking tends to breakdown as the frequency increases, it is of interest to know how the vulnerability of the interaural integration to ITD is frequency dependent.

Human listeners are also able to detect a transient break in correlation (BIC) between the two ears (i.e., a transient change of interaural correlation from 1 to 0 and to 1), and measuring the duration threshold for detecting this transient change in interaural correlation is a way of estimating the ability to temporally resolve fast changes in interaural configurations (Akeroyd and Summerfield, 1999; Boehnke *et al.*, 2002). Note that introducing a change in interaural correlation for wideband noises does not change the energy and spectrum in the signals, but it can change the loudness of the signals (Culling, 2007).

If there is a degeneration of the interaural integration of fine-structure details by introducing an ITD, a BIC should be less detectable (increase in the duration threshold for detecting the BIC). Thus, measuring the duration threshold for detecting the BIC at various ITDs provides a way of investigating whether the interaural integration of acoustic details is affected by an interaural delay. One of the purposes of this study was to investigate the effect of ITD on interaural integration of fine-structure acoustic details by examining whether the duration threshold for detecting the BIC embedded in either wideband or narrowband noises is affected by ITD.

The sensitivity to the interaural correlation appears to be frequency dependent (Akeroyd and Summerfield, 1999; Culling *et al.*, 2001; Mason *et al.*, 2005). For example, when the center frequency of the narrowband noise is 250 Hz and the bandwidth is 100 Hz, human listeners can detect the occurrence of the BIC with the mean duration threshold of 6.5 ms. When the center frequency becomes 1000 Hz and the bandwidth is still 100 Hz, the threshold increases to 35 ms (Akeroyd and Summerfield, 1999). However, it should be noted that because the bandwidth of the auditory filter varies roughly on a logarithmic scale with changing frequency (Glasberg and Moore, 1990), using a bandwidth that varies logarithmically is more appropriate for studying the center-frequency effect on interaural integration of fine-structure information of narrowband noises. Another purpose of this study was to examine whether the duration threshold for detecting the BIC embedded in narrowband noises is affected by the center frequency when the narrowband-noise bandwidth is fixed at 1/3 octaves across various center frequencies.

Recognizing acoustic signals (e.g., comprehending speech) is particularly difficult for older adults in noisy, reverberant environments (Nábělek and Robinson, 1982; Nábělek, 1988; Huang *et al.*, 2008b; Helfer, 1992; Helfer and Wilber, 1990). Since interaural integration is important for both the precedence effect (Blauert and Divenyi, 1988; Litovsky *et al.*, 1999; Scharf, 1974; Shinn-Cunningham *et al.*, 1995; Trahiotis *et al.*, 2005; Yang and Grantham, 1997a, 1997b) and auditory perception in noisy, reverberant environments (Bregman, 1990), it is possible that there is an age-related decline in interaural integration. However, previous studies have failed to find any age-related effects on a fusion-related localization task when the inter-click delay was in the range 0.7–8 ms for inducing the phenomenon of "the law of the first wavefront" (Cranford *et al.*, 1993), and on the echo threshold when stimuli was short-duration (4 ms) noises (Roberts and Lister, 2004), long-duration (from 250 to 350 ms) 1/4-octave-wide noise (with the center frequency of 1000, 2000, or 3000 Hz) (Lister and Roberts, 2005), or short-duration (about 2 ms) tone bursts (Schneider *et al.*, 1994). Nevertheless, two studies seem to have found some age-related effects. (1) Cranford *et al.* (1993) reported an age effect on the fusion-related localization task when the inter-click delay was no larger than 0.5 ms. However, it is still not clear whether this effect was caused by age-related hearing loss and/or binaural imbalance because the authors used the mean bilateral high-frequency pure tone averages across 1, 2, and 4 kHz as the criteria for assigning participants into the normal-hearing group or the hearing-loss group. It is still not clear whether group differences in the pattern of "error" occurred when the bilateral clicks started simultaneously or when the click on one side led the click on the other side. (2) Using 4 ms bursts of white noise as stimuli, Roberts *et al.* (2002) reported that listeners with hearing loss (mean age =68 years) had longer echo thresholds than listeners with normal hearing (mean age=29 years). However, their later studies (Roberts and Lister, 2004) have shown that there was no effect of aging or hearing loss on the echo threshold under dichotic or anechoic conditions. Specifically, under the reverberant condition, older adults with normal-hearing sensitivity (ONH) exhibited the highest thresholds, followed by those for younger adults with normal-hearing sensitivity (YNH) and older adults with impaired-hearing sensitivity (OIH). The mean echo thresholds for the ONH group were significantly higher than those of the OIH group, but the thresholds of the YNH group were not significantly different from those of the ONH group or the OIH group for the reverberant condition, showing no aging effects. Thus, there is a need to

further investigate whether changes in interaural integration occur in people with the early stages of presbycusis.

This study investigated whether the duration threshold for detecting the BIC embedded in the temporal middle of interaurally correlated noises was influenced by increasing the ITD in younger adults with normal hearing and older adults in the early stages of presbycusis (Experiment 1). Moreover, this study also investigated whether the ITD-related modulation of the detection of the BIC in narrowband noises depends on the center frequency (Experiment 2). Instead of a linearly constant bandwidth as used by Akeroyd and Summerfield (1999), this study used the logarithmically-constant bandwidth of 1/3 octaves for the comparison across narrowband noises with various center frequencies.

II. EXPERIMENT 1

A. Methods

1. Participants

Ten younger university students (19–28 years old, mean age=23.0, six females) and eight older adults (65–74 years old, mean age=68.4, five females) participated in this study. None of the participants had any history of hearing disorders, and none used hearing aids. The participants gave their written informed consent to participate in the experiment and were paid a modest stipend for their participation.

The younger participants all had normal and symmetrical (no more than 15-dB difference between the two ears) and no more than 25-dB pure-tone hearing thresholds between 125 and 8000 Hz. In total, 43 older adults who self-reported to have normal hearing were examined for their hearing sensitivity, but only 8 of them passed the hearing test. These eight older participants had symmetrical (no more than 15-dB difference between the two ears) and no more than 25-dB pure-tone hearing thresholds between 125 and 500 Hz, and symmetrical and no more than 40-dB pure-tone hearing thresholds between 1000 and 4000 Hz. Although hearing thresholds at 8000 Hz were measured, they were not used for screening participants. Figure 1 presents average hearing levels for the two age groups as a function of the testing-tone frequency.

As Fig. 1 shows, the thresholds of older participants were generally higher than those of younger participants, and threshold differences between younger and older participants continued to increase with frequency. Although these older adults had clinically normal hearing, they were best characterized as being in the early stages of presbycusis.

2. Apparatus and stimuli

During a testing session, the participant was seated in a sound-attenuating chamber (EMI Shielded Audiometric Examination Acoustic Suite). Gaussian wideband noise signals (0–22.05 kHz) with the duration of 1000 ms (including 30-ms rise-fall times) were synthesized using the "randn()" function in the MATLAB function library at the sampling rate of 44.1 kHz with 16-bit amplitude quantization. The stimuli were transferred using the Creative Sound Blaster PCI128, passed through an AURICAL system, and presented to listeners by two headphones (Model HDA 200). Calibration of

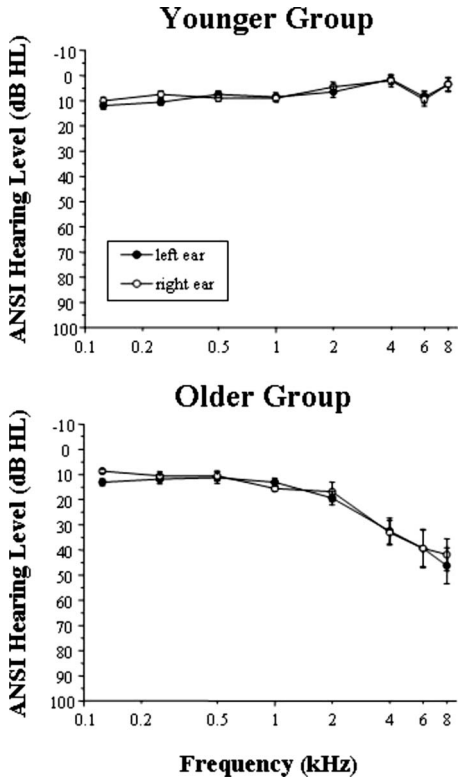


FIG. 1. Average hearing thresholds in the left ear (closed symbols) and the right ear (open symbols) for the younger-participant group (top panel) and those for the older-participant group (bottom panel) tested in Experiment 1. ANSI: American National Standards Institute (S3.6-1989). Error bars represent the standard errors of the mean.

sound level was carried out with the Larson Davis Audiometer Calibration & Electroacoustic Testing System (AUDit and System 824, Larson Davis) with "A" weighting. Although a new random noise was generated for each trial, the sound level was fixed at 58 dB sound pressure level.

3. Procedures

Two 1000-ms presentations of correlated noises were delivered over headphones. The right-headphone noise in one of the presentations was an exact copy of the left-headphone noise. The right-headphone noise in the other presentation was also identical to the left-headphone noise except for the substitution of a randomly selected independent noise fragment (i.e., BIC) introduced into the temporal middle of the correlated 1000-ms noises. The introduction of the BIC made the interaural correlation drop from 1 to a value near 0 (but not 0 because two randomly generated noises are not necessarily orthogonal) and then return to 1. For wideband noises used in this experiment, the interaural correlation for the BIC was not larger than 0.045. The duration of the BIC was systematically manipulated during the testing (see below).

In each trial, the inserted BIC had equal possibility to be randomly assigned to one of the two presentations. The delay between the two presentations (from the end of the first presentation to the onset of the second presentation) was 1000 ms. For each presentation, the noise presented on the left headphone started simultaneously with that presented on the

right headphone or led the right-ear noise by 2, 4, or 6 ms. For each participant, the order of presenting the ITDs was randomized. A new noise was used for each trial. The participant's task was to identify which of the two presentations contained the transient change.

The participant initiated a trial with a particular ITD by typing a letter on the computer keyboard. The two-interval forced-choice procedure was used for measuring the BIC detection threshold. A three-down-one-up paradigm was used for systematically manipulating the BIC duration (Levitt, 1971). The starting BIC duration was set at 250 ms, which was a sufficiently large duration based on results of our pilot experiments. The BIC duration was decreased following three consecutive correct identifications of the presentation containing the BIC, and increased following one incorrect identification. The initial step size of changing the fragment duration was 50 ms, and then the step size was altered by a factor of 0.5 with each reversal of direction until the minimum step size of 0.5 ms was reached. Feedback was given visually after each trial. A test session was terminated following ten reversals in direction and the threshold for that session was defined as the average BIC duration for the last six reversals. Test sessions were repeated four times for each participant, and the average over the two lowest session thresholds defined the participant's threshold. To ensure that each participant understood the experimenter's instructions and became familiar with the procedure, a brief training session was used before the experiment.

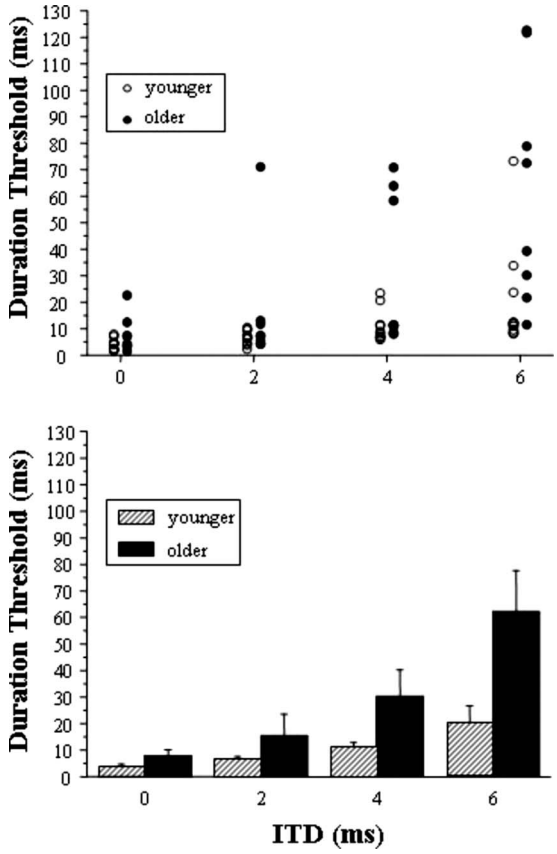


FIG. 2. Comparison of the duration threshold for detecting the BIC embedded in wideband noises between younger participants and older participants at each of the four ITDs (ITD=0, 2, 4, or 6 ms) in Experiment 1. The BIC was embedded in the 1000-ms interaurally correlated Gaussian noises. The top panel shows individuals' thresholds at each of the ITDs and the bottom panel shows group-mean thresholds at each of the ITDs. Error bars represent the standard errors of the mean.

B. Results

All the younger and older participants were able to detect the BIC when the BIC duration was sufficiently long at each of the ITDs (0, 2, 4, or 6 ms). Duration thresholds for younger and older individuals at the four ITDs are shown in the top panel of Fig. 2, and group-mean duration thresholds at the ITDs for younger and older participants are shown at the bottom panel of Fig. 2. Clearly, the duration threshold for detecting the BIC became larger with increasing ITD for each of the two age groups. Also, the thresholds were generally larger for older participants than for younger participants at all the ITDs.

A 4×2 two-way-mixed-subject analysis of variance (ANOVA) showed that the main effect of ITD was significant $[F(3,48)=17.55, p<0.001]$, the main effect of age group was significant $[F(1,16)=6.40, p=0.022]$, and the interaction between ITD and age group was significant $[F(3,48)=5.16, p=0.004]$. Further separate ANOVAs showed that when the ITD was 0 or 2 ms, the group effect was not significant ($p>0.05$ for both). However, when the ITD was increased to 4 or 6 ms, the group effect was significant [4 ms: $F(1,16)=4.49, p=0.050$; 6 ms: $F(1,16)=7.33, p=0.016$]. Separate ANOVAs also showed that the ITD effect was significant for both the younger group $[F(3,27)=5.93, p=0.003]$ and the older group $[F(3,21)=10.23, p<0.001]$. These analyses indicate that changing the ITD significantly affects the detection of the BIC and older participants needed a larger BIC duration than younger participants to correctly detect the occurrence of the change in interaural correlation particularly when a sufficiently large ITD was introduced.

C. Discussion

The stimulus design used in the present study was effectively the same as used in both the Akeroyd and Summerfield (1999) study and the study of Boehnke *et al.* (2002), that is, binaurally-presented noises underwent from correlated to uncorrelated and then to correlated, without introducing any substantial changes in energy and spectrum. In the present study, when no interaural delay was introduced (ITD=0 ms), the group-mean duration threshold for detecting the BIC was 4.0 ms for younger participants, which is larger than the mean threshold (2.3 ms) of the BIC measured in eight participants (20–35 years old) in the study of Boehnke *et al.* (2002) using a broadband noise (0–22 050 Hz) but smaller than the mean threshold of BIC (5.3 ms) measured in six participants (whose ages were not provided) in the Akeroyd and Summerfield (1999) study using bandpass noises (100–500 Hz). The present study also shows that although older participants needed larger durations (the

group-mean threshold=7.8 ms) to detect the BIC than younger participants, the group difference was not statistically significant.

The detection of BIC depends on the central computation of the dynamic interaural correlation of noise fine structure. Both younger participants and older participants tested in the present study were able to detect the BIC even though an ITD of 6 ms was introduced, indicating that they could integrate fine-structure information of steady-state wideband noises at the two ears across an interaural delay at least 6 ms. These results are in agreement with previous studies showing that listeners with normal hearing are able to lateralize much larger ITDs than those experienced in free-field listening (e.g., Blodgett *et al.*, 1956; Mossop and Culling, 1998). However, the duration threshold for detecting the BIC increased as the ITD became larger, suggesting that the contrast between the central representation of the BIC and that of the rest of the noise (the temporal flanks of the BIC) decreases with increasing ITD. Moreover, when the ITD was increased to 4 or 6 ms, the group-mean duration threshold for older participants was significantly larger than that for younger participants. Thus there is an age-related decline in the ability to integrate fine-structure information of wideband noises presented at the two ears particularly when an interaural delay is introduced.

Since the perceived size of the noise image depends on the interaural correlation (Blauert and Lindemann, 1986; Goupell and Hartmann, 2006), detection of changes in correlation may be essentially detection of the size change of the auditory event. It is well known that detection of a small decrease in interaural correlation from a reference correlation continues to become worse as the reference correlation decreases (Boehnke *et al.*, 2002; Culling *et al.*, 2001; Gabriel and Colburn, 1981; Koehnke *et al.*, 1986; Pollack and Trittipoe, 1959). Thus it is much easier to detect the change from the perfectly correlated value (+1) to a slightly decorrelated value than from a slightly decorrelated value to a more decorrelated value. With increasing ITD, the central representation of fine structure of noise may be progressively diminished, and a compact, correlated image (when the correlation is 1) may be broadened, leading to reduced sensitivity to changes in the interaural correlation. Moreover, the reduction in the sensitivity to the correlation change is greater with increasing ITD in older participants than in younger participants.

Why does the compact image of interaurally correlated noise become broadened when the ITD increases? Stern *et al.* (1988) described a cross-frequency weighted-image model to explain perceptual-image lateralization of complex binaural stimuli based on theoretic operations on trajectories of the maxima of the frequency-dependent interaural cross-correlation function of the binaural stimuli. According to this model, listeners' subjective laterality of stimuli is determined by a weighted combination of the ITD centroids of the trajectories of the maxima following peripheral bandpass filtering and rectification. The frequency-independent "straightness" of the trajectory of maxima and the centrality of the trajectory of maxima are the two bandwidth-dependent weighting functions and they compete with each other for determining the consistent ITD. The straightness weighting function is to weigh heavily the trajectories that are straighter, which very likely represent a true ITD of a stimulus. The centrality weighting function is to emphasize the contribution of a greater number of trajectories with various values of internal delays. Thus an interesting question can be raised: If a trajectory with any magnitude is able to make its (weighted) contribution to the image laterality and the relative weight of a trajectory depends on its both straightness and centrality, does ITD also affect the stimulus compactness in addition to the stimulus laterality? Yost (1981) estimated the lateral position within the head for binaurally-presented tones with various frequencies and showed that although cross correlation or a coincidence network can account for the results at any one frequency, it cannot account for the results across frequencies. In other words, the image is not at the same position for a particular value of ITD at all frequencies. For tones with frequencies less than 1000 Hz, the lateral position corresponds more closely to the interaural phase difference than it does to the ITD. The model responses for expressing the image position within the head appear to be located at positions closer to midline as the tone frequency increases, and the range of the image distributions (which can be used to describe the degree of diffuseness) is also affected by the tone frequency. Thus introducing an ITD for interaurally correlated wideband noises may not only lateralize the auditory event (making the event move away from the midline and more toward one ear) but also weaken the compactness of the noise image, thereby reducing the sensitivity to a drop in interaural correlation (possibly due to compressed changes in event size). This view is supported by the Mossop and Culling (1998) study showing that with the increase in reference ITD, the just-noticeable differences between the reference-noise image and the testing-noise image in intracranial position became larger (the discrimination of laterality cues became poorer). This would provide an explanation for the elevated threshold for detecting the BIC in both younger and older listeners when the ITD became larger and the temporal flanks of BIC became more diffuse (i.e., the perception of laterality is fading).

Goupell and Hartmann (2006, 2007a, 2007b) showed that in the task of distinguishing between the slightly incoherent noises (coherence=0.992) and diotic noises (coherence=1.0), for narrowband noises, the incoherence was much more readily detectable in noises with larger fluctuations in interaural phase difference or in interaural level difference than in noises with smaller fluctuations. However, as the bandwidth increased, the incoherence became equally detectable in all the different noises, consistent with a model in which detection is predictable from interaural coherence alone. In their studies, when the coherence values were the same for noises with interaural phase fluctuations and interaural level fluctuations of difference size, incoherence was more easily detected for noises with large interaural fluctuations than for noise with smaller fluctuations. However, as the noise bandwidth increased, the distributions of interaural phase fluctuations and interaural level fluctuations across noises became narrower, and the value of noise coherence was sufficient to predict listeners' detection performance.

Since Gaussian wideband noises were used in Experiment 1 of this study, any potential age-related reduction in central representation of fluctuations in interaural phase difference and/or in interaural level difference might not contribute to the increase in threshold for detecting the BIC in the older group. Instead, the age-related threshold increase might be caused by declines in either monaural fine-structure processing (e.g., increase in filter bandwidth and reduction in phase-locking or synchrony) and/or binaural fine-structure processing (e.g., increase in "binaural sluggishness").

Some previous studies have shown that older normal-hearing listeners have smaller masking level differences (MLDs) than younger-adult listeners (Grose *et al.*, 1994; Olsen *et al.*, 1976; Pichora-Fuller and Schneider, 1991, 1992, 1998; Zurek and Durlach 1987; Strouse *et al.*, 1998). Since detection of interaural-correlation difference is closely associated with the MLD (Durlach and Pang, 1986; Bernstein and Trahiotis, 1992), the assumed age-related decline in processing fine-structure information would be associated with the age-related deficits of interaural integration. In the Pichora-Fuller and Schneider (1992) study, for example, the threshold of detecting a 500-Hz pure tone against band-limited white noise (0.1–5 kHz) for older participants differed significantly from that for younger participants only when the interaural delay of the noise masker was increased to a value between 0.5 and 5 ms. Results of the present study show that when the ITD was 0 or 2 ms, the mean threshold for detecting the BIC in older participants was not significantly larger than that in younger participants. However, when the ITD was increased to 4 or 6 ms, the mean threshold for detecting the BIC in older participants was significantly larger than that in younger participants. Thus the results of this study are generally in agreement with the Pichora-Fuller and Schneider (1992) report. The age-related decline in the ability to interaurally integrate correlated sounds when an interaural delay is introduced may be related to the difficulties experienced by older listeners in noisy, reverberant environments (Huang *et al.*, 2008b). However, it is not clear whether the age-related decline in the ability to integrate correlated sounds between the two ears is due to the age-related reduction or loss in monaural neural firing synchronization (monaural jitter) or the assumed age-related increase in binaural jitters. Pichora-Fuller and Schneider (1992) proposed that the age-related differences in binaural unmasking could be explained by the difference of change in temporal jitter as a function of internal interaural delay if the following two premises are satisfied: (1) temporal jitter in the binaural system does not vary as a function of internal interaural delay in old subjects, and (2) temporal jitter increases with internal interaural delay in young subjects. However, further evidence, especially physiological evidence, is still needed to clarify the issues of age-related temporal jitter.

III. EXPERIMENT 2

The results of Experiment 1 show that detection of the BIC embedded in wideband noises is affected by the ITD. As mentioned in Sec. I, the sensitivity to the interaural correlation is frequency dependent. Thus it is also important to know whether for narrowband noises the ITD effect on the interaural fine-structure integration is center-frequency dependent. In Experiment 2, we examined whether the sensitivity to the BIC in narrowband noises depends on the center frequency and, particularly, whether the center-frequency effect interacts with the ITD effect. Since fine-structure processing decreases with increasing frequency (due to degradation of phase locking), it was predicted that for narrowband noises, BIC detection would decrease with increasing center frequency (indeed, according to the "duplex" theory, BIC detection would become strikingly poor when the center frequency reaches 3000 Hz). In Experiment 2, the bandwidth was kept logarithmically-constant at 1/3 octaves across noises with various center frequencies.

A. Method

1. Participants

Twelve young university students (20–25 years, mean age=22.7 years, eight females) with normal hearing participated in Experiment 2. They gave their written informed consent to participate in the experiment and were paid a modest stipend for their participation. The criteria for evaluating participants' hearing were the same as those used in Experiment 1.

2. Apparatus and materials

Similar to Experiment 1, the participant was tested in the EMI Shielded Audiometric Examination Acoustic Suite. To match the stimuli that were used in our recent neurophysiological studies of the effects of center frequency and ITD on humans' scalp event-related potentials to the BIC (Huang *et al.*, 2008a), the duration of Gaussian wideband noises was 2000 ms, including 30-ms rise-fall times, and the noises were synthesized using the randn() function in the MATLAB function library at the sampling rate of 48 kHz with 16-bit amplitude quantization.

In narrowband conditions, stimuli had a fixed logarithmically-constant bandwidth of 1/3 octaves and a center frequency of 200, 400, 800, 1600 or 3200 Hz. The stimuli were filtered with 512-point bandpass finite impulse response (FIR) filters. In wideband conditions, the stimuli were low-pass filtered at 10 kHz with a 512-order FIR filter. In order to avoid the spectral artifacts due to energy splatter outside the pass band, each stimulus was filtered after any uncorrelated fragment was introduced. The stimulus transduction, display, and calibration were the same as those used in Experiment 1.

Similar to Experiment 1, introducing a BIC in narrowband noises made the interaural correlation drop from 1 to a value close to 0 (but not 0 because the two randomly generated noises were not necessarily orthogonal) and then return to 1. However, the interaural correlation during the BIC for narrowband noises exhibited larger fluctuations than that for wideband noises. The maximum correlation for the BIC across narrowband noises used in this experiment varied with the center frequency (200 Hz: 0.255; 400 Hz: 0.200; 800 Hz: 0.150; 1600 Hz: 0.136; 3200 Hz: 0.116).

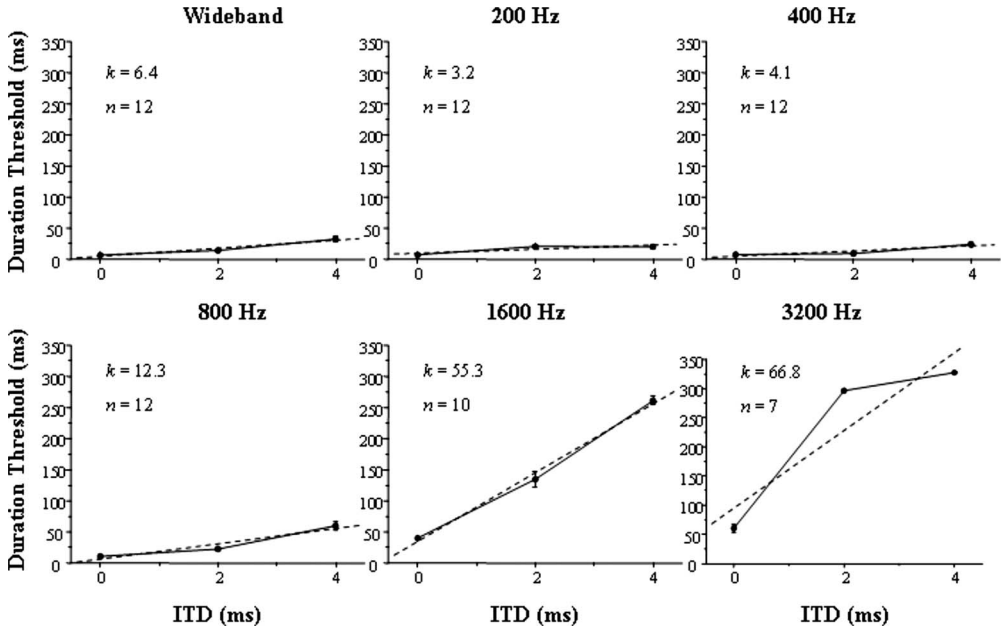


FIG. 3. The mean duration threshold across participants tested in Experiment 2 for detecting the BIC as a function of the ITD when the noise was wideband noise (top left panel) or narrowband noise with the center frequency of 200 Hz (top middle panel), 400 Hz (top right panel), 800 Hz (bottom left panel), 1600 Hz (bottom middle panel), or 3200 Hz (bottom right panel). The broken line in each panel is the (linear) best fitting of the duration threshold as a function of the ITD. *k* represents the slope of the best-fitting line. *n* represents the number of participants who could detect the BIC at all the ITDs under the particular noise-type condition. Error bars represent the standard errors of the mean

3. Procedures

The testing procedure and participants' task in Experiment 2 were the same as in Experiment 1. For each noise presentation, the noise at the left headphone either started simultaneously with that at the right headphone or led that at the right headphone by 2 or 4 ms. There were two within-subject factors: (i) noise type (wideband noise or narrowband noise with the center frequency of 200, 400, 800, 1600, or 3200 Hz) and (ii) ITD (0, 2, or 4 ms). The order of ITD was counterbalanced among 12 participants using the Latin-square design. Under each ITD condition, the order of noise type was in a random manner. A test session was terminated following ten reversals in direction, and the threshold for that session was defined as the average duration for the last six reversals. Test sessions were repeated three times for each participant, and the average over the three session thresholds defined the participant's threshold for the testing condition.

B. Results

When the BIC duration was sufficiently long, all the participants could detect the occurrence of the BIC around the temporal middle of the sustained noise under each of the noise-type conditions, except that two participants (Participant Nos. 1 and 8) could not detect the occurrence of the BIC embedded in the 3200-Hz narrowband noise even when the BIC duration was at the maximum value (330 ms) used in this experiment.

When the ITD was 2 ms, two participants (Participant Nos. 1 and 8) could not detect the BIC in the 1600-Hz narrowband noise and four participants (Participant Nos. 1, 7, 8, and 9) could not detect the BIC in the 3200-Hz narrowband noise. When the ITD was 4 ms, two participants (Participant Nos. 1 and 8) could not detect the BIC in the 1600-Hz narrowband noise and five participants (Participant Nos. 1, 2, 7, 8, and 9) could not detect the BIC in the 3200-Hz narrowband noise.

Figure 3 shows BIC duration thresholds for participants who could detect the BIC in a particular type of noise at all the ITDs. The broken line in each panel represents the linear best fitting of the duration threshold as a function of the ITD. The slope (k) of the best-fitting line is also indicated. Since some participants could not detect the BIC in high-frequency noises especially at long ITDs, the numbers of participants are not identical across panels.

For each noise type, the duration threshold was elevated with increasing ITD. For example, at the 400-Hz center frequency, as the ITD increased from 0 to 2 ms, the mean duration threshold increased from 7.6 to 9.2 ms. As the ITD increased to 4 ms, the threshold increased to 24.2 ms. At the 800-Hz center frequency, as the ITD increased from 0 to 2 ms, the mean duration threshold increased from 10.8 to 22.8 ms. As the ITD increased to 4 ms, the threshold increased to 60.0 ms. Obviously, the effect of the ITD on the duration threshold for narrowband noises was affected by the center frequency. Detection of the BIC in high-frequency narrowband noises was much more vulnerable to the change in the ITD than that in low-frequency noises. This important feature can be indicated by the slope of the best-fitting line in each panel. With the increase in the center frequency, the slope increased considerably. The largest effect occurred when the center frequency was 3200 Hz.

For the seven participants who could detect the BIC in each of the noise types at all the ITDs, a 3 (ITD) $\times$ 6 (noise type) within-subject ANOVA showed that the main effect of ITD, the main effect of noise type, and the interaction between the two factors were all significant ($p < 0.001$ for all).

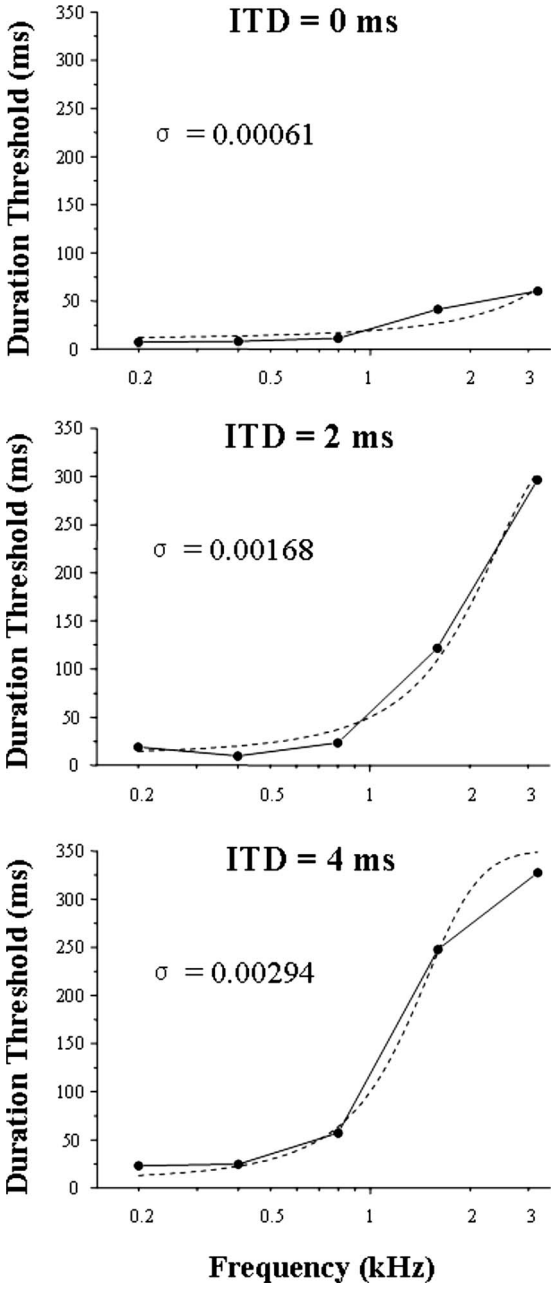


FIG. 4. The mean duration threshold across seven (young) participants tested in Experiment 2 for detecting the BIC in narrowband noises as a function of the center frequency when the ITD was 0 ms (top panel), 2 ms (middle panel), or 4 ms (bottom panel). The broken curve in each panel represents the logistic function for fitting the duration threshold as a function of the center frequency. σ is the slope parameter of the psychometric function.

Separate one-way within-subject ANOVAs showed that at each of the ITDs, the effect of noise type was significant ($p < 0.001$ for all), confirming that detection of the BIC was noise-type dependent. Moreover, separate one-way within-subject ANOVAs showed that for each of the six noise types the effect of ITD was significant ($p < 0.001$ for all).

To further estimate the interaction of ITD and center frequency on the BIC detection, Fig. 4 shows the BIC duration thresholds for narrowband noises as a function of the center frequency for the seven participants who could detect the BIC in each of the noise types at all the ITDs, when the ITD was 0 ms (top panel), 2 ms (middle panel), or 4 ms (bottom panel). As indicated in Fig. 4, with increasing ITD from 0 to 4 ms, the detection of BIC became more affected by the center frequency.

We used the following logistic function to fit the threshold data across center frequencies at each of the ITDs (the broken curve) using the Levenberg–Marquardt method (Wolfram, 1991):

$$y = \frac{350}{1 + e^{-\sigma(x-\mu)}},$$

where y is the threshold at the frequency x, μ is the frequency corresponding to 50% of the maximum y value on the psychometric function at an ITD condition, and σ determines the slope of the psychometric function. The σ value at each of the ITDs is indicated in Fig. 4. Clearly, with increasing ITD from 0 to 4 ms, the slope parameter σ became larger, indicating that the BIC detection threshold increased faster with the center frequency as the ITD increased. Note that the same function can also fit the data across ITD at each of the center frequencies.

C. Discussion

The results of Experiment 2 show that the duration threshold for detecting the BIC in narrowband noises with the fixed bandwidth (1/3 octaves) largely depended on the center frequency. With increasing center frequency from 200 to 3200 Hz, the duration threshold for detecting the BIC progressively increased. Thus the results support the view that detecting the dynamic changes in interaural correlation is easier for low-frequency narrowband noises than for high-frequency narrowband noises (Akeroyd and Summerfield, 1999). Although Bernstein and Trahiotis (1999) suggested that temporal processing for binaural detection must be accounted for differently for high-frequency and low-frequency stimuli, a gradual degradation of processing dynamic changes in interaural correlation may occur as the frequency changes from low to high. Moreover, since the duration threshold for detecting the BIC in the wideband noise was very close to that for low-frequency (200 or 400 Hz) narrowband noise, low-frequency components in wideband noises seem to make a majority of the contribution to the BIC detection. The results also support the view proposed by Akeroyd and Summerfield (1999) that high-frequency components in wideband noises would not interfere with the detection of the transient break in interaural correlation. The frequency dependency of the sensitivity to the dynamic change in interaural correlation may be associated with both the frequency dependency of the discrimination of interaural-correlation difference (Culling *et al.*, 2001) and the frequency dependency of the perceived auditory source width of interaurally correlated noise stimuli (Mason *et al.*, 2005). Since fine-structure ITDs of acoustic stimuli cannot be processed well when the frequency becomes higher than 1500 Hz, in Experiment 2 a sharp increase in the duration threshold occurred when the center frequency changed from 800 to 1600 Hz. Note that since the maximum value of the BIC used in Experiment 2 was 330 ms, the change in duration

threshold when the center frequency changed from 1600 to 3200 Hz (see Figs. 3 and 4) might be underestimated due to a potential ceiling effect particularly when the ITD was 2 or 4 ms.

The center-frequency effect also interacts with the ITD effect on the detection of BIC. For each of the noise types used in this experiment, with the change in the ITD from 0 to 4 ms, the perceptual fusion of the noises presented to the two ears was not broken but the threshold of detecting the BIC increased. Particularly, the rate of the threshold elevation became monotonically larger as the center frequency of narrowband noise was increased (Fig. 4). These results suggest that the interaural integration of high-frequency acoustic components degenerates faster with increasing ITD than that of low-frequency components, supporting the reports that a great range of ITDs can be lateralized when low frequencies are present than when only high frequencies are available (Blodgett *et al.*, 1956; Mossop and Culling, 1998). Thus, center frequency cooperates with ITD in determining the interaural integration of correlated noises.

IV. SUMMARY

This study provides evidence that the interaural integration of correlated noises is affected by ITD, aging, and center frequency (the last for narrowband noises). Even when listeners do not experience a breakdown of the perceptual fusion of two correlated noises presented at the two ears with the increase in the ITD, there is a temporal degeneration of the interaural integration of fine-structure acoustic information. In younger listeners with normal hearing, this temporal degeneration of the interaural integration for narrowband noise is center-frequency dependent: high-frequency noises degenerate faster with increasing ITD than low-frequency noises. Moreover, there is an age-related decline in the ability to integrate binaural noises particularly when an ITD is introduced.

Processing the similarity and dissimilarity of sound waves arriving at the two ears contributes to binaural perceptual phenomena including the precedence effect (Blauert and Divenyi, 1988; Litovsky *et al.*, 1999; Scharf, 1974; Shinn-Cunningham *et al.*, 1995; Trahiotis *et al.*, 2005; Yang and Grantham, 1997a, 1997b). Since interaural integration is important for auditory perception in noisy, reverberant environments (Bregman, 1990), it will be necessary in the future to investigate the relationship between the age-related changes in interaural integration and age-related difficulties in recognizing sound sources (e.g., speech) in such adverse situations.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 30711120563, 30670704, and 60535030), and the "973" National Basic Research Program of China (2009CB320901).

[1]In this manuscript, we use the term "interaural correlation" following the definition by Grantham (1995): "The interaural correlation of a sound is the correlation between the sound waveform presented to the left ear and the sound waveform presented to the right ear after one waveform is shifted in time to maximize the correlation." We also notice that a related term, "coherence," has been used by some authors. For example, as proposed by Blauert (1982), "the degree of coherence k is defined as the maximum absolute value of the normalized cross-correlation function of two signals." According to our present knowledge, the term interaural correlation, but not the term "interaural coherence," has been used in a large number of (over 100) papers published in primary journals in the field. Particularly, to keep the consistency with the two previous studies that are most related to the present study (e.g., Akeroyd and Summerfield, 1999; Boehnke *et al.*, 2002), interaural correlation, instead of interaural coherence, is used in this manuscript.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.
Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high-frequencies," J. Acoust. Soc. Am. **91**, 306–316.
Bernstein, L. R., and Trahiotis, C. (**1999**). "The effects of signal duration on NoSo and NoS pi thresholds at 500 Hz and 4 kHz," J. Acoust. Soc. Am. **105**, 1776–1783.
Blauert, J. (**1982**). *Spatial Hearing* (MIT, Cambridge, MA), p. 238.
Blauert, J., and Divenyi, P. L. (**1988**). "Spectral selectivity in binaural contralateral inhibition," Acustica **66**, 267–274.
Blauert, J., and Lindemann, W. (**1986**). "Spatial-mapping of intracranial auditory events for various degrees of interaural coherence," J. Acoust. Soc. Am. **79**, 806–813.
Blodgett, H. C., Wilbanks, W. A., and Jeffress, L. A. (**1956**). "Effect of large interaural time differences upon the judgment of sidedness," J. Acoust. Soc. Am. **28**, 639–643.
Boehnke, S. E., Hall, S. E., and Marquardt, T. (**2002**). "Detection of static and dynamic changes in interaural correlation," J. Acoust. Soc. Am. **112**, 1617–1626.
Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
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.
Culling, J. F. (**2007**). "Evidence specifically favoring the equalization-cancellation theory of binaural unmasking," J. Acoust. Soc. Am. **122**, 2803–2813.
Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**). "Interaural correlation sensitivity," J. Acoust. Soc. Am. **110**, 1020–1029.
Durlach, N. I., and Pang, X. D. (**1986**). "Interaural magnification," J. Acoust. Soc. Am. **80**, 1849–1850.
Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). "Dynamic processes in the precedence effect," J. Acoust. Soc. Am. **90**, 874–884.
Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapers from notched-noise data," Hear. Res. **47**, 103–138.
Goupell, M. J., and Hartmann, W. M. (**2006**). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," J. Acoust. Soc. Am. **119**, 3971–3986.
Goupell, M. J., and Hartmann, W. M. (**2007a**). "Interaural fluctuations and the detection of interaural incoherence. II. Brief duration noises," J. Acoust. Soc. Am. **121**, 2127–2136.
Goupell, M. J., and Hartmann, W. M. (**2007b**). "Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models," J. Acoust. Soc. Am. **122**, 1029–1045.
Grantham, D. W. (**1995**). "Spatial hearing and related phenomena," in *Hearing*, edited by B. C. J. Moore (Academic, London).
Grose, J. H., Poth, E. A., and Peters, R. W. (**1994**). "Masking level differences for tones and speech in elderly listeners with relatively normal audiograms," J. Speech Hear. Res. **37**, 422–428.
Haas, H. (**1951**). "On the influence of a single echo on the intelligibility of speech," Acustica **1**, 49–58.
Hall, D. A., Barrett, D. J. K., Akeroyd, M. A., and Summerfield, A. Q. (**2005**). "Cortical representations of temporal structure in sound," J. Neurophysiol. **94**, 3181–3191.
Helfer, K. S. (**1992**). "Aging and the binaural advantage in reverberation and noise," J. Speech Hear. Res. **35**, 1394–1401.
Helfer, K. S., and Wilber, L. A. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Hear. Res. **33**, 149–155.

Huang, Y., Huang, Q., Chen, X., Qu, T.-S., Wu, X.-H., and Li, L. (**2008b**). "Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults," Hear. Res. **244**, 51–65.

Huang, Y., Kong, L.-Z., Fan, S.-L., Wu, X.-H., and Li, L. (**2008a**). "Both frequency and interaural delay affect ERP responses to binaural gap," NeuroReport **19**, 1673–1678.

Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). "Performance in several binaural-interaction experiments," J. Acoust. Soc. Am. **79**, 1558–1562.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Li, L., Qi, J. G., He, Y., Alain, C., and Schneider, B. (**2005**). "Attribute capture in the precedence effect for long-duration noise sounds," Hear. Res. **202**, 235–247.

Lister, J. J., and Roberts, R. A. (**2005**). "Effects of age and hearing loss on gap detection and the precedence effect: Narrow-band stimuli," J. Speech Lang. Hear. Res. **48**, 482–493.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., and Shinn-Cunningham, B. G. (**2001**). "Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression," J. Acoust. Soc. Am. **109**, 346–358.

Mason, R., Brookes, T., and Rumsey, F. (**2005**). "Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli," J. Acoust. Soc. Am. **117**, 1337–1350.

Mossop, J. E., and Culling, J. F. (**1998**). "Lateralization of large interaural delays," J. Acoust. Soc. Am. **104**, 1574–1579.

Nábělek, A. K. (**1988**). "Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss," J. Acoust. Soc. Am. **84**, 476–484.

Nábělek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.

Olsen, W. O., Noffsinger, D., and Carhart, R. (**1976**). "Masking level differences encountered in clinical populations," Audiology **15**, 287–301.

Pichora-Fuller, M. K., and Schneider, B. A. (**1991**). "Masking-level differences in the elderly: A comparison of antiphasic and time-delay dichotic conditions," J. Speech Hear. Res. **34**, 1410–1422.

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.

Pollack, I., and Trittipoe, W. J. (**1959**). "Binaural listening and interaural noise cross correlation," J. Acoust. Soc. Am. **31**, 1250–1252.

Rakerd, B., Hartmann, W. M., and Hsu, J. (**2000**). "Echo suppression in the horizontal and median sagittal planes," J. Acoust. Soc. Am. **107**, 1061–1064.

Roberts, R. A., Besing, J., and Koehnke, J. (**2002**). "Effects of hearing loss on echo thresholds," Ear Hear. **23**, 349–357.

Roberts, R. A., and Lister, J. J. (**2004**). "Effects of age and hearing loss on gap detection and the precedence effect: Broadband stimuli," J. Speech Lang. Hear. Res. **47**, 965–978.

Scharf, B. (**1974**). "Localization of unlike tones from two loudspeakers, in sensation and measurement," in *Papers in Honor of S. S. Stevens*, edited by H. R. Moskowitz, B. Scharf, and J. C. Stevens (Reidel, Dordrecht), pp. 309–314.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980–991.

Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (**1995**). "Cross-frequency interactions in the precedence effect," J. Acoust. Soc. Am. **98**, 164–171.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Trahiotis, C., Bernstein, L. R., Stern, R. M., and Buell, T. N. (**2005**). "Interaural correlation as the basis of a working model of binaural processing: An introduction," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 238–271.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336.

Wolfram, S. (**1991**). *Mathematica: A System for Doing Mathematics by Computer* (Addison-Wesley, New York).

Yang, X., and Grantham, D. W. (**1997a**). "Echo suppression and discrimination suppression aspects of the precedence effect," Percept. Psychophys. **59**, 1108–1117.

Yang, X., and Grantham, D. W. (**1997b**). "Cross-spectral and temporal factors in the precedence effect: Discrimination suppression of the lag sound in free-field," J. Acoust. Soc. Am. **102**, 2973–2983.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.

Zurek, P. M. (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 952–964.

Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.

Testing the binaural equal-loudness-ratio hypothesis with hearing-impaired listeners[a)]

Jeremy Marozeau[b)]
Department of Speech-Language Pathology and Audiology (106A FR), Communication Research Laboratory, and Institute for Hearing, Speech and Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115 and The Bionic Ear Institute, 384-388 Albert Street, East Melbourne, Victoria 3002, Australia

Mary Florentine
Department of Speech-Language Pathology and Audiology (106A FR), Communication Research Laboratory, and Institute for Hearing, Speech and Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 2 July 2008; revised 4 March 2009; accepted 23 April 2009)

The primary purpose of the present experiment was to test whether the binaural equal-loudness-ratio hypothesis (i.e., the loudness ratio between monaural and binaural tones presented at the same Sound Pressure Level, SPL, is independent of SPL) holds for hearing-impaired listeners with bilaterally symmetrical hearing losses. The outcome of this experiment provided a theoretical construct for modeling loudness-growth functions. A cross-modality matching task between string length and tones was used to measure three loudness functions for eight listeners: two monaural (left and right) and one binaural. A multiple linear regression was performed to test the significance of presentation mode (monaural vs binaural and left vs right), level, and their interaction. Results indicate that monaural loudness functions differ between the ears of two listeners. The interaction between presentation mode (binaural/monaural) and level was significant for one listener. Although significant, these differences were quite small. Generally, the binaural equal-loudness-ratio hypothesis appears to hold for hearing-impaired listeners. These data also indicate that loudness-growth functions in two ears of an individual are more similar than loudness-growth functions in ears from different listeners. Finally, it is demonstrated that loudness-growth functions can be constructed for individual listeners from binaural level difference for equal-loudness data.
 [DOI: 10.1121/1.3133703]

PACS number(s): 43.66.Pn, 43.66.Sr, 43.66.Cb, 43.66.Ba [RL] Pages: 310–317

I. INTRODUCTION

A tone presented binaurally is louder than the same tone presented monaurally (Fletcher and Munson, 1933). Fletcher and Munson (1933) (see, also Fletcher, 1953) assumed that the binaural-to-monaural loudness ratio is independent of level. This assumption of a constant ratio between the binaural and monaural loudness-growth functions for the same sound is known as the binaural equal-loudness-ratio hypothesis (BELRH). The BELRH is supported by empirical data for normal listeners (Marozeau *et al.*, 2006), but remains an untested assumption for listeners with hearing losses of primary cochlear origin. The first of four aims of the present experiment is to obtain a set of data that is comprehensive enough to permit testing of the BELRH for individual hearing-impaired (HI) listeners.

Testing the BELRH for HI listeners is important for two reasons. First, the BELRH was assumed to be true by Whilby *et al.* (2006) to model loudness-growth functions for HI listeners. This assumption needs to be tested. Second, current hearing-aid fitting procedures often assume that the binaural-to-monaural loudness ratio is about the same for all people with the same amount of hearing loss. This assumption may not be valid and could be partly responsible for the uncomfortable loudness experienced by hearing-aid users. A survey (Kochkin, 2005) of 1500 hearing-aid users indicates that only 60% of them reported being satisfied when asked about comfort with loud sounds.

Testing the BELRH for HI listeners presents a potential problem that is not encountered when testing normal-hearing listeners. For normal-hearing listeners, loudness functions for right and left ears usually grow at the same rate (Marks, 1978, 1980). Therefore, either ear can be used to obtain the monaural loudness-growth function for comparison with their binaural loudness-growth function. For HI listeners, it is unclear whether the loudness-growth functions are similar in both ears of the same listener, even if they have bilaterally symmetrical losses. Although most fitting procedures for binaural hearing aids assume the same loudness-growth functions in both ears of individual HI listeners with symmetrical hearing losses, it is possible that hearing losses could affect the two ears of an individual in different ways. Data clearly indicate that two ears from different HI listeners with the same amount of hearing loss can have different loudness

[a)] A portion of this work was presented in a talk to the American Auditory Society. [Marozeau, J., and Florentine, M. (2008). AAS bulletin, 33, p. 29].
[b)] Author to whom correspondence should be addressed. Electronic mail: jmarozeau@bionicear.org

TABLE I. Summary of the individual data (gender, age); clinic measurements (thresholds at 0.25, 0.5, 1, 2, and 4 kHz in dB Hearing Level (HL) for both ears and presumed etiology); frequency in kHz selected for the experiment and threshold in dB SPL measured in the laboratory; individual results (exponent α and binaural-to-monaural ratio). Abbreviations for etiology are congenital or hereditary hearing loss (CHHL), presbycusis (Pres.), noise induced hearing loss (NIHL), and family history of hearing loss (FHHL).

Listeners			Clinic threshold (dB HL)								Laboratory measurement		Expt. result	
	Gender	Age	Ear	0.25*k*	0.5*k*	1*k*	2*k*	4*k*	8*k*	Et.	Test frequency (kHz)	Threshold (dB SPL)	*a*	Ratio
HI-1	M	65	L	55	60	65	65	75	70	CHHL	1.05*k*	70	0.314	1.297
			R	50	60	60	60	60	75			72		
HI-2	M	65	L	10	5	40	45	70	70	Pres.	1*k*	65	0.201	1.15
			R	10	15	40	45	75	85			64		
HI-3	M	76	L	20	30	40	45	75	85	Pres.	2*k*	65	0.279	1.342
			R	15	40	40	45	75	nr			67		
HI-4	M	60	L	55	60	65	65	70	90	NIHL	1*k*	65	0.236	1.163
			R	55	65	65	70	75	90			68		
HI-5	F	60	L	20	15	45	35	50	55	Pres.	0.95*k*	45	0.25	1.533
			R	15	15	40	60	45	60			47		
HI-6	M	61	L	5	10	30	55	95	95	NIHL	0.95*k*	33	0.158	1.337
			R	10	10	25	40	85	95			34		
HI-7	M	74	L	25	30	45	50	65	50	Pres. FHHL	1*k*	53	0.19	1.141
			R	25	30	40	45	50	50			55		
HI-8	F	73	L	45	40	45	45	75	75	Pres.	1*k*	55	0.22	1.355
			R	40	35	40	50	80	80					

functions (Dix *et al.*, 1948; Knight and Margolis, 1984; Hellman and Meiselman, 1990; Brand and Hohmann, 2001; for a review see Marozeau and Florentine, 2007). In fact, HI listeners with similar audiograms can have different loudness-growth functions (Hellman, 1994; Florentine *et al.*, 1997), as can normal-hearing listeners (Epstein and Florentine, 2005, 2006). Therefore, the second aim of the present experiment is to obtain a set of data from the right and left ears of individual listeners with symmetrical hearing losses to test the assumption that their monaural loudness-growth functions are similar. These data will be compared with binaural loudness-growth functions from the same listeners.

The third aim of the present experiment is to compare the binaural and monaural loudness-growth functions for individual HI listeners to those of normal-hearing listeners. There are two different measures used to study binaural loudness summation. One is the previously described binaural-to-monaural loudness ratio. The other is the binaural level difference for equal loudness (BLDEL), which was used by Whilby *et al.* (2006). When binaural and monaural loudness-growth functions are plotted together on a logarithmic scale, the binaural-to-monaural loudness ratio corresponds to the vertical distance and the BLDEL corresponds to the horizontal distance between the two functions. If the BELRH holds, the two functions are parallel and it would be possible to derive a loudness-growth function directly from the BLDEL data. This is possible because the slope of the loudness-growth function would be proportional to the inverse of the BLDEL.

The last of the four aims proposed for the present experiment is to compare the BLDEL of normal and HI listeners measured using an indirect method with the data of Whilby *et al.* (2006). This is an important cross-check of the method to derive loudness functions from BLDELs by Whilby *et al.* (2006) for HI listeners.

To accomplish all four aims of the present experiment as efficiently and effectively as possible, a reliable cross-modality matching procedure used by Epstein and Florentine (2005) and Marozeau *et al.* (2006) to measure loudness functions in normal-hearing listeners was used to test HI listeners in the present experiment.

II. METHOD

A. Stimuli

The stimuli were pure tones with equivalent rectangular durations of 200 ms, including 6.67-ms raised-cosine rises and falls. Frequencies were chosen individually for each listener around 1 or 2 kHz and can be found in Table I. For each listener, the frequency of the tone was selected to have a threshold difference of less than 3 dB between the two ears. Although problems with threshold microstructure that could influence loudness judgments near threshold have been observed (Horst *et al.*, 2003), none were observed in our listeners. Tones were presented monaurally (right and left ears) and binaurally at the same Sound Pressure Level (SPL) to both ears. Having the same threshold at the two ears prevented the potentially confounding issues of the differences in loudness at equal-SL (Sensation Level) vs equal-SPL. The level of the tone changed in 5-dB steps from the first multiple of 5-dB above threshold to 100 dB SPL, resulting in a total of approximately nine levels per listener. The upper limit of 100 dB SPL was chosen to be loud, but did not cause

any tolerance issues for any of the listeners. The Institutional Review Board of Northeastern University protocol was strictly followed.

B. Procedure

The experiment consisted of two parts: absolute threshold measurements and cross-modality matches. Except for the selection of the stimuli, the procedure was the same as that used by Marozeau *et al.* (2006) and is summarized below.

1. Absolute thresholds

Absolute thresholds were measured separately for each ear using a two-interval, two-alternative forced-choice paradigm with feedback. The listeners' task was to indicate which 250-ms visually marked interval contained the signal by pressing a key on a small computer terminal. Each threshold measurement consisted of three interleaved tracks, each of which ended after five reversals. For each track, the level of the signal was initially set approximately 10 dB above the expected threshold of the listeners. The step size was 5 dB until the second reversal, after which it decreased to 2 dB. The threshold for each track was calculated as the average signal level of the last two reversals. The average of the three interleaved tracks was considered the absolute threshold. Absolute thresholds for right and left ears were then compared. If the difference was greater than 3 dB a different frequency was selected.

2. Cross-modality matching

For each listener, three modes of presentation were used: monaural (right and left ears) and binaural. As noted in Sec. II A, approximately nine levels were tested for each presentation mode. This yielded about 27 stimuli for each listener.

Listeners judged the loudness of each stimulus using a string-length cross-modality matching procedure. This procedure was chosen because it has been shown to yield reliable individual data. (For further information, see Epstein and Florentine, 2005.) Listeners were asked to cut a piece of a string that was as long as the sound was loud from a virtually unbounded ball of very thin, but strong string (i.e., embroidery floss). After cutting each piece of string, the listener taped it into a notebook, turned the page, and pressed a button to indicate completion of the response. This response initiated presentation of the next stimulus after a 700-ms delay.

A total of six cross-modality matches were made for each monaural stimulus and a total of 12 matches were made for each binaural stimulus. The 216 matches per listener ($6+6+12$ matches $\times 9$ levels) were divided into four testing sessions administered in random order over two days, with two sessions per day separated by a 15-min break. Stimuli were presented binaurally (at the same SPL to each ear) and monaurally, always to the same ear within one session. The two test days were separated by less than two weeks, except for HI-7. The first part of HI-7's data was collected in a pilot experiment performed two months before with slight variation of protocol in which tones were presented at the same SL, which is within 2 dB of equal SPL at the two ears and within the variability of the measurement for the other listeners. Each level was presented three times within a session.

At the start of each trial, a new tone level and one of the two modes (binaural or monaural) were randomly selected from all other stimuli that had not yet been presented three times and had a level within 30 dB of the level of the previous trial. The 30-dB level restriction was included to avoid surprising the listener with a sudden large level increase or decrease, which may cause the listener to miss attending to a stimulus. If no stimuli fulfilled these criteria, but some other stimuli still had been presented fewer than three times, a dummy trial was inserted. The dummy trial had the same mode and a level 30-dB above or below the preceding level, depending on the levels of the stimuli that remained to be presented. The dummy trials were not included in the final analysis.

C. Apparatus

A PC-compatible computer with a 24-bit sound card (Lynx Two-b) played the tone that was generated with MATLAB. The sampling rate was 48 kHz. The computer also recorded the listeners' responses and executed the adaptive procedure. The output of the sound card was led to a headphone buffer (TDT HB6), which fed the earphones of the Sony MDR-V6 headset. For routine calibration performed before each session, the output of the headphone buffer was sent back to the sound card, such that the computer could sample the waveform and calculate its rms voltage.

D. Listeners

Eight naïve HI listeners with symmetrical sensorineural hearing losses of primarily cochlear origin participated in this experiment. Their hearing losses ranged from mild to severe. Table I shows gender, age, presumed etiology, and audiometric data. The audiometric data were obtained using a calibrated audiometer (ANSI, 2004) and a modified Hughson–Westlake procedure (Harrell, 2002, p. 73).

E. Data analysis

The geometric mean of string lengths for each stimulus was computed for each listener and level using all available data. The standard deviation was determined from the logarithms of the string lengths. The group mean and standard error were calculated across the individual listener's geometric means for each presentation mode (monaural right, monaural left, or binaural) and level. The resulting data were transformed back into the string-length domain to show the probable range of each individual listener's responses.

A correction factor was applied to each session in order to minimize any possible difference between the binaural data obtained across sessions caused by changes in the listeners' internal judgment standard. First, the binaural data from the first session in which the monaural data were presented to the left ear were selected as a reference. A factor was obtained with the aid of a mean-square fit to minimize the differences between the logarithm of the binaural data of each subsequent session and the reference data. Each factor

was then applied in the linear domain to the monaural data as well in order to not impact the binaural-to-monaural ratio. The factor applied was usually less than 0.2 (with a median of 0.1). To examine the effects of stimulus variability, a multiple linear regression was performed on the logarithms of the string lengths obtained for each level and mode using the statistics package R (www.r-project.org). Three models were tested:

$$\text{model 1: } \log(S) = aL + b + \text{err}, \quad (1)$$

$$\text{model 2: } \log(S) = aL + b + Nc + \text{err}, \quad (2)$$

$$\text{model 3: } \log(S) = aL + b + Nc + NLd + \text{err}, \quad (3)$$

where S is the string-length estimation, L is the level, and N is the Boolean factor condition. This factor will be set to zero for left monaural stimuli, and unity for right monaural stimuli, when analyzing the effect of ear (Sec III B). It will be set to zero for monaural stimuli, and unity for binaural stimuli, when testing the BELRH (Sec. III C). If the sum of squares of the error for model 2 is significantly lower than the sum of squares for model 1, then the effect of the presentation mode is significant. If the sum of squares of the error for model 3 is significantly lower than the sum of squares for model 2, then the effect of the interaction condition and level is significant. For all tests, $p \leq 0.01$ is considered significant.

III. RESULTS

A. Mean and variability

Figure 1 shows the two monaural loudness functions for each of the eight listeners. Figure 2 shows the binaural loudness functions and the average of the two monaural functions for each listener. String lengths ranged from 0.1 cm (the precision of the measurement) to 37.8 cm with an arithmetic mean of 4.2 cm [standard deviation (std) of the mean is 2.6 cm between subjects, and mean of the std is 1.33 within subjects] for the monaural left condition, 4.12 cm (std of 2.76 cm between subjects and 1.34 within subjects) for the monaural right condition, and 5.4 cm (std of 2.71 cm between subjects and 1.29 within subjects) for the binaural condition. The intra-subject variability differs among individual listeners. Whereas HI-4, HI-7, and HI-8 have relatively small variability, HI-3 and HI-5 have relatively large variability. However, the variability observed for the HI listeners is within the same range as the normal-hearing listeners measured by Marozeau *et al.* (2006).

It should be noted that a few listeners did not hear every presentation of the 5-dB-SL tone and did not cut any string. Because at least four estimations were available for each stimulus, statistical analyses were performed on the remaining data.

B. Monaural loudness functions

Figure 1 shows that monaural loudness functions for the two ears of an individual HI listener are more similar than monaural loudness functions for different HI listeners. A multiple linear regression was performed between the data

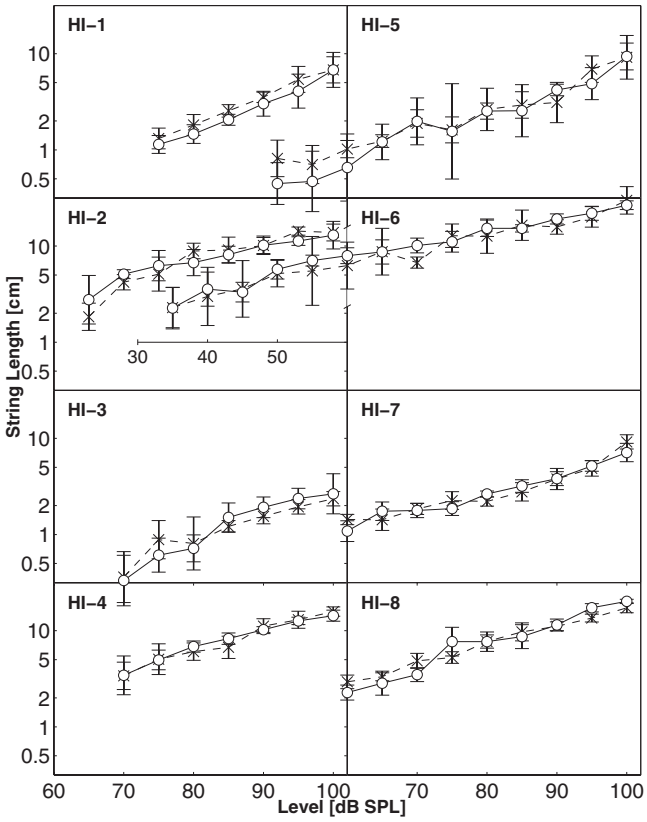


FIG. 1. Individual monaural loudness functions for tones from all eight HI listeners. The geometric means of string lengths are plotted on a log scale as a function of level. Data are shown separately for left (x) and right (o) ears. The vertical bars show ±1 standard deviation of the log of the string lengths. Note that data for listeners HI-5 and HI-6 are continued into the left panel while maintaining the same relative scale.

for the left and right ears of individual listeners in order to test if the two ears were significantly different. The analysis indicates that the effect of level was always significant ($p < 0.001$, as expected, for all eight listeners). The effect of ear presentation (left vs right) was significant ($p=0.004$, $R^2 = 2\%$) for only one listener out of eight, HI-1, indicating that the loudness function for his left ear was significantly higher than the one for his right ear. The interaction between ears and level was significant ($p=0.002$, $R^2=1\%$) only for listener HI-8, indicating that the slope of loudness function for his left ear was significantly shallower than the one for his right ear. For six of the eight listeners the ear and the interaction between ear and level were not significant ($p > 0.01$). This provides support for the contention that the two monaural loudness functions are not essentially different. It is worth noting that if the two loudness functions of HI-1 were plotted in dB SL, the gap between the functions will decrease accordingly. The two functions show a difference of 3 dB. This is consistent with a threshold elevation in his right ear of 2 dB (a previous measurement in the clinic reported a 5-dB difference in threshold between his ears). When the statistical analysis was performed in SL, no significant effects of ear or interaction were found for this listener $p > 0.01$.

C. Binaural vs monaural loudness function

Because the difference between the two monaural loudness functions for HI-1 and HI-8 was small enough, as

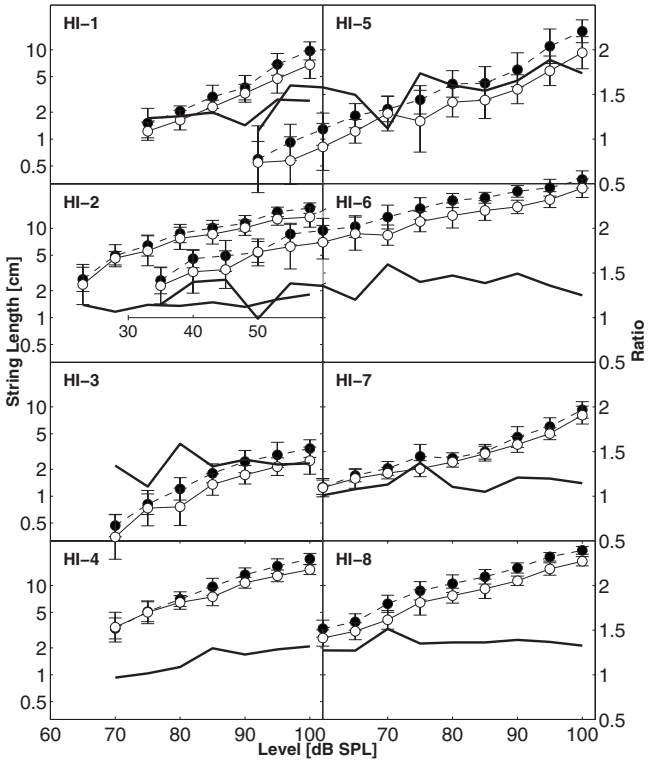


FIG. 2. Individual binaural (filled circles) and the averaged monaural (open circles) loudness functions for tones for eight HI listeners. The geometric means of string lengths (on the left ordinate) are plotted as in Fig. 1. The thick line shows the ratio of string lengths (on the right ordinate) for equal-SPL monaural and binaural tones.

shown by the R^2, not to impact the rest of the analysis, the two monaural loudness functions were averaged and compared to the binaural function for each listener in Fig. 2. As expected, the binaural tones were perceived louder than the monaural tones. Statistical analysis supports this observation; the factor presentation mode (binaural vs monaural) was significant for all eight listeners ($p<0.004$). Despite the variability, the ratios between the monaural and binaural data appear relatively constant across level for most listeners. The interaction was significant only for HI-4 ($p=0.002$, $R^2=14\%$). As the analysis and Fig. 1 show, the two monaural loudness functions were similar; therefore, it appears that for this listener the binaural ratio increased with level. For the seven other listeners no significant interactions were found.

The binaural and monaural data for each listener were fitted with two parallel straight lines using the method of least-squares. Although the shape of the loudness functions differs somewhat from a straight line, this method was used as a rough approximation to derive the overall slope. The exponents of the power functions and the binaural-to-monaural ratio for each listener are summarized in Table I. The exponent varies from 0.16 to 0.31, with an average of 0.23. The ratio varies from 1.14 to 1.53, with an average of 1.29. The average exponent is lower than the 0.3 of the power law (Stevens, 1955), and this ratio is lower than the doubling of loudness expected from perfect binaural loudness summation (Marks, 1978; Hellman, 1991). However, both are within the range of what has been found for normal-hearing listeners in literature (see Marozeau *et al.*, 2006).

IV. DISCUSSION

A. Testing the binaural equal-loudness-ratio hypothesis for hearing-impaired listeners

The first aim of the present experiment was to obtain a set of data to permit testing of the BELRH for HI listeners. Results of the present experiment support the BELRH, except for listener HI-4. The binaural loudness function for HI-4 is significantly steeper than his monaural loudness function. It is unclear why his binaural loudness summation is dependent on level. No apparent difference can be found in his audiological profile and the monaural loudness functions for both his ears were not significantly different.

Although the present experiment provides comprehensive data for eight individual HI listeners, the sample is not large enough to make precise inferences to all HI listeners. However, it seems reasonable to assume that the BELRH is supported for most HI listeners with bilaterally symmetrical-hearing losses, which is the most common type of hearing loss (http://www.nidcd.nih.gov/funding/programs/ot/inner_ear_summary.html).

B. Monaural loudness functions

The second aim of the present experiment was to obtain a set of data comprehensive enough to test the assumption that the right and left ears of an individual listener with symmetrical hearing losses have similar monaural loudness-growth functions. Although listeners HI-1 and HI-8 show significant differences between the ears, only listener HI-8 shows a loudness function with a shallower slope in the left ear than the right ear. (Recall that for listener HI-1 the loudness function for his left ear was significantly higher than the one for his right ear, but the slopes were not significantly different.) Although the difference in slope is small, it may be important because it illustrates the fact that there may be a difference in the loudness-growth functions between the two ears of the same listener even with the same threshold in both ears.

C. Loudness-growth functions in normal-hearing and impaired-hearing listeners

The third aim of the present experiment was to compare the binaural and monaural loudness-growth functions for individual HI listeners to those of normal-hearing listeners. Two types of loudness-growth functions have been described for listeners with sensorineural hearing losses of primary cochlear origin: the rapid growth type and softness imperception (for a review see Marozeau and Florentine, 2007). In the rapid growth type (same as the classical view of recruitment) the model is described as follows: (1) loudness at threshold is the same for normal-hearing and HI listeners, (2) loudness at and near threshold grows more rapidly than for normal-hearing,[1] and (3) loudness is the same or approaches that of normal listeners at high levels. The second type of loudness growth is called softness imperception, because it refers to the inability of the listener to hear soft sounds. This model is described as follows: (1) Loudness at threshold is higher for HI listeners than normal-hearing listeners, (2)

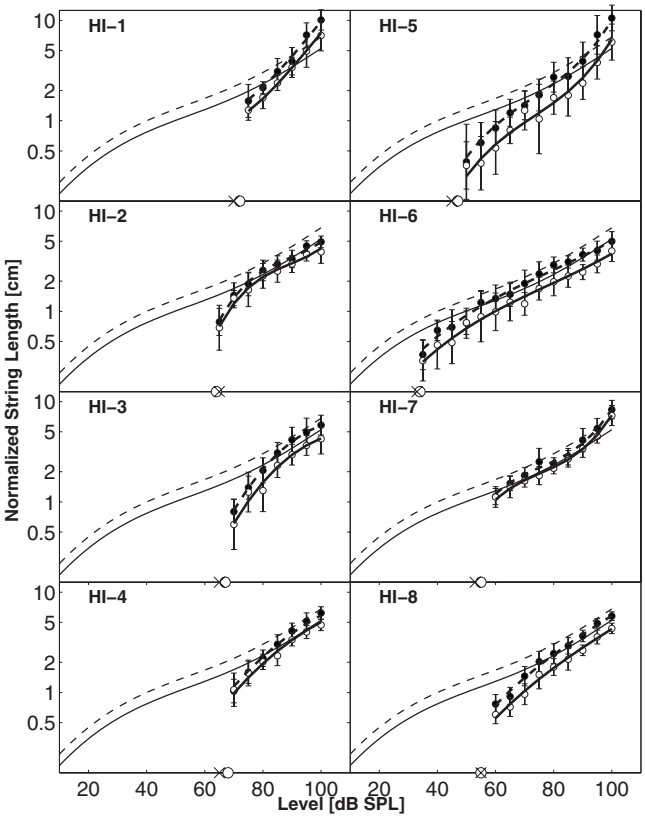


FIG. 3. Individual normalized data (see text) replotted from Fig. 2 together with polynomial fits for the binaural (thick dashed lines) and the monaural (thick continuous lines) loudness functions. The averaged normal data for eight listeners from Marozeau *et al.*, 2006 are also shown for the corresponding binaural (thin dashed line) and monaural (thin continuous line) loudness functions for 1-kHz tones. Absolute thresholds are shown for left (x) and right (o) ears on the abscissa.

loudness growth at and near threshold is similar for normal-hearing and HI listeners, (3) the loudness of some HI listeners exhibits a reduction in the amount of gain (also known as loss of compression) at moderate levels, and (4) the loudness-growth function approaches that of normal-hearing listeners at higher levels.

Data from individual HI listeners from the present experiment are compared in Fig. 3 to the average data from eight normal-hearing listeners obtained from Marozeau *et al.* (2006) using the same task as in the present study. In order to be able to compare the two sets of data across different individual scales, the present data were normalized using the same procedure as Whilby *et al.* (2006): The polynomial fit was set such that each individual average binaural loudness at 85 dB SPL matched the overall average binaural loudness of the average normal listeners at 80 dB SPL. Although this normalization was based on an assumption that is likely to be inaccurate in some cases, it cannot be too far off either because (1) listeners reported that the sounds were loud at these levels, and (2) loudness matching data in literature are more nearly equal at high levels for listeners with hearing losses of primarily cochlear origin.

Considerable individual differences can be observed in the rate of loudness growth with increasing level among the HI listeners. For example, HI-2 and HI-3 show behavior consistent with the rapid growth type, and HI-7 shows a behavior consistent with softness imperception. The loudness functions of HI-2 and HI-3 show rapid growth close to threshold and that quickly approaches normal. On the other hand, the loudness function of listener HI-7 is normal at all tested levels. The lowest level tested was 5 dB above threshold. This implies that for this individual listener either loudness increases at an improbable rate between threshold and 5 dB SL, or loudness at threshold is greater than normal at or very near threshold. The other HI listeners shown in Fig. 3 exhibit an intermediate behavior between these two extremes.

The present data are consistent with data in literature. Marozeau and Florentine (2007) reviewed five published experiments from different laboratories, obtained using different methods, to measure individual loudness functions for normal and HI listeners. They found that (1) individual differences are greater for HI listeners than for normal listeners, and (2) some HI listeners seem to show rapid growth, some softness imperception, and some a combination of both. Therefore, a sufficient number of individual functions for HI listeners exist in literature and they show clear individual differences. (*N.B.* loudness data should not be averaged across HI listeners because these important differences will be missed).

D. The binaural level difference for equal-loudness data

As explained in the Introduction, the fourth and final aim of the present experiment is to take the BLDEL data of normal and HI listeners (indirectly derived from the present loudness measurements) and compare them with the data of Whilby *et al.* (2006). The BLDELs were extracted by measuring the level difference of the fitted monaural and binaural functions for every listener. (For details of the fitting procedure and assumptions to derive loudness functions, see Whilby *et al.*, 2006.) Figure 4 shows these data compared to those of Whilby *et al.* (2006) obtained using a loudness-matching procedure. As shown by Whilby *et al.* (2006), the BLDEL data vary with level and are non-monotonic. Values range from 1 dB for HI-2 at 65 dB SPL to 10 dB for HI-5 at 70 dB SPL. Six out of eight HI listeners show BLDEL data that vary within the range of the HI data shown by Whilby *et al.* (2006) and two listeners (HI-5 and HI-6) show BLDEL data that follow the normal range. It is noteworthy that the latter two listeners have the mildest hearing losses of all the listeners. Therefore, there appears to be a direct relationship between the BLDEL, the slope of the loudness-growth functions, and the binaural-to-monaural ratio. This provides support for the method proposed by Whilby *et al.* (2006).

E. Validation of a method to derive loudness functions from loudness matches

The BELRH was studied because it was one of the assumptions of a method used to derive loudness functions from BLDEL data. This method is fully described in Whilby *et al.*, 2006 and Marozeau *et al.*, 2006. Instead of the classical power function, the logarithm of loudness was modeled with a third-order polynomial

$$\log(F_m) = a_m L^3 + b_m L^2 + c_m L + d_m, \quad (4)$$

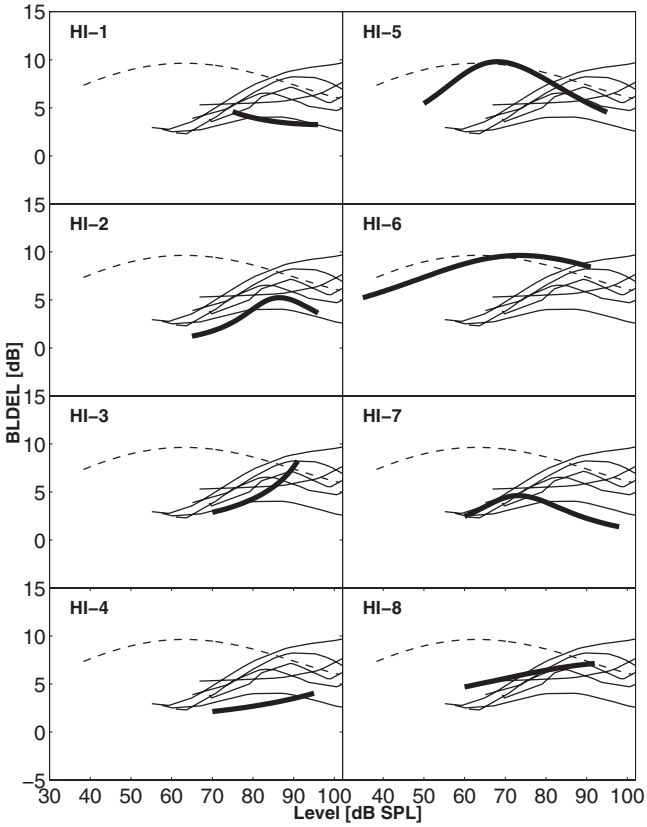


FIG. 4. The BLDEL is plotted as a function of level. The polynomial fits to individual data from the present study (thick lines) are compared with the data of Whilby *et al.* (2006): individual data from eight HI listeners (thin lines) and averaged data from eight normal-hearing listeners (dashed lines). Each panel represents a polynomial fit for one listener from the present study compared with all the data from Whilby *et al.* (2006).

$$\log(F_b) = a_b L^3 + b_b L^2 + c_b L + d_b, \tag{5}$$

where F_m and F_b are the monaural and binaural loudness functions, respectively, at the level L; a_m, b_m, c_m, d_m, a_b, b_b, c_b, and d_b are the free parameters of the polynomial fit. The BELRH assumes that the slopes of the two functions are parallel. Therefore, it implies that the first three coefficients of both functions are the same: $a_m = a_b$, $b_m = b_b$, and $c_m = c_b$. It also implies that the difference of the last parameter is equal to the log of the binaural-to-monaural ratio, K: $dm\text{-}db = \log(K)$. The data of Whilby *et al.* (2006) indicate how the BLDEL for each individual listener will vary with level. For a fixed selected level of the monaural stimulus, L_m, the BLDEL estimates the level of the binaural stimulus, L_b, at which the loudnesses of the monaural and binaural stimuli were equal,

$$\log(F_b(L_b)) = \log(F_m(L_m)), \tag{6}$$

$$a({L_m}^3 - {L_b}^3) + b({L_m}^2 - {L_b}^2) + c(L_m - L_b) = db - dm = \log(k). \tag{7}$$

The BLDEL data from the study of Whilby *et al.* (2006) were fitted to extract the coefficients of the model. Then, by using a least-squares fit, the three free parameters (a, b, and c) were selected to minimize the error of the fit between the model and the BLDEL data. Marozeau *et al.* (2006) showed

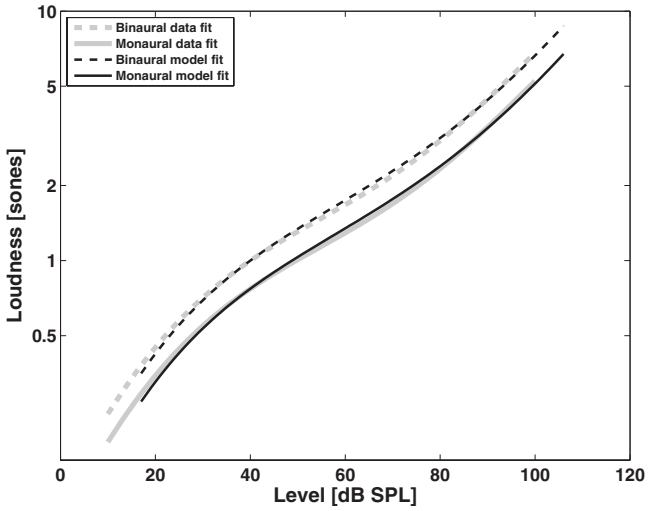


FIG. 5. Modification of the method described in Marozeau *et al.*, 2006. The thick gray lines represent the binaural (dashed line) and average monaural (continuous) loudness functions of eight normal-hearing listeners (Marozeau *et al.*, 2006); the thin dark lines represent the binaural (dashed line) and monaural (continuous line) loudness-growth functions constructed for individual listeners from the binaural level difference data for equal loudness (Whilby *et al.*, 2006). The function is constructed here with a fourth-order polynomial, instead of the third order used in Whilby *et al.*, 2006.

that the method was valid by comparing the averaged loudness function to the derived loudness function from the BLDEL data. The result indicates that the method accurately predicts the data from low-to-moderate levels. A review of literature on individual loudness functions (Marozeau and Florentine, 2007) shows that the slope as a function of level is not a symmetrical function. In other words, the slope changes more rapidly at low levels than high levels. Therefore, a third-order polynomial was required to fit the slope, and by extension a fourth order for the loudness function (i.e., the integral of the slope). Figure 5 shows the same data as in Fig. 4 of Marozeau *et al.* (2006), except that now the function has been fitted with a fourth-order polynomial, which fits the data better. The method to derive loudness functions from loudness matches now appears valid throughout the entire audible range of levels, not just low-to-moderate levels, at least for normal-hearing listeners.

V. CONCLUSION

The results of this study indicate that the BELRH holds for most HI listeners with symmetrical sensorineural hearing losses of primary cochlear origin. Furthermore, the results show that monaural loudness functions for right and left ears of an individual HI listener are quite similar. However, important individual differences are found among HI listeners. Some listeners have loudness functions that show rapid growth (also known as recruitment), some show softness imperception, and some an intermediate behavior. The BLDEL data extracted from the fitted monaural and binaural loudness functions are within the range found by Whilby *et al.* (2006). This last result—combined with the BELRH data—indicates that loudness-growth functions can be constructed for individual listeners from BLDEL data.

ACKNOWLEDGMENTS

The authors wish to thank Lauren Chase and Becky Daley, Au.D., for assistance in data acquisition. Professor Michael Epstein provided helpful comments. This research was supported by NIH-NIDCD Grant No. R01DC02241 and a fellowship for advanced researchers from the Swiss National Science Foundation.

[1]Moore (2004) modeled loudness-growth functions of HI listeners as follows: (1) HI and normal listeners have the same loudness at threshold, (2) both groups show the same loudness growth from 0- to 5-dB SL, and (3) HI listeners show a steeper slope after 5-dB SL.

ANSI (**2004**). "American National Standard Specification for Audiometers," ANSI S3.6–2004.

Brand, T., and Hohmann, V. (**2001**). "Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling," Audiology **40**, 92–103.

Dix, M. R., Hallpike, C. S., and Hood, J. D. (**1948**). "Observations upon the loudness recruitment phenomenon, with especial reference to the differential diagnosis of disorders of the internal ear and VIII nerve," Proc. R. Soc. Med. **41**, 516–526.

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

Epstein, M., and Florentine, M. (**2006**). "Loudness of brief tones measured by magnitude estimation and loudness matching," J. Acoust. Soc. Am. **119**, 1943–1945.

Fletcher, H. (**1953**). *Speech and Hearing in Communication*, 2nd ed., Bell Telephone Laboratories Series (Van Nostrand, Princeton, NJ).

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Florentine, M., Buus, S., and Hellman, R. P. (**1997**). "A model of loudness summation applied to high-frequency hearing loss," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, New York), pp. 187–198.

Harrell, R. W. (**2002**). "Puretone evaluation," in *Handbook of Clinical Audiology*, 5th ed., edited by J. Katz (Lippincott Williams & Wilkins, Philadelphia, PA), pp. 71–87.

Hellman, R. (**1991**). in *Ratio Scaling of Psychological Magnitude*, edited by S. J. Bolanowski and G. A. Gescheider (Erlbaum, Hillsdale, NJ), pp. 215–228.

Hellman, R. P. (**1994**). "Relation between the growth of loudness and high-frequency excitation," J. Acoust. Soc. Am. **96**, 2655–2663.

Hellman, R. P., and Meiselman, C. H. (**1990**). "Loudness relations for individuals and groups in normal and impaired hearing," J. Acoust. Soc. Am. **88**, 2596–2606.

Horst, J. W., Wit, H. P., and Albers, F. W. J. (**2003**). "Quantification of audiogram fine-structure as a function of hearing threshold," Hear. Res. **176**, 105–112.

Knight, K. K., and Margolis, R. H. (**1984**). "Magnitude estimation of loudness. II: Loudness perception in presbycusic listeners," J. Speech Hear. Res. **27**, 28–32.

Kochkin, S. (**2005**). "Marketrak VII: Hearing loss population tops 31 million people," Hear. Rev. **12**, 16–29.

Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am. **64**, 107–113.

Marks, L. E. (**1980**). "Binaural summation of loudness: noise and two-tone complexes," Percept. Psychophys. **27**, 489–498.

Marozeau, J., Epstein, M., Florentine, M., and Daley, B. (**2006**). "A test of the binaural equal-loudness-ratio hypothesis for tones," J. Acoust. Soc. Am. **120**, 3870–3877.

Marozeau, J., and Florentine, M. (**2007**). "Loudness growth in individual listeners with hearing losses: A review," J. Acoust. Soc. Am. **122**, EL81–EL87.

Moore, B. C. (**2004**). "Testing the concept of softness imperception: Loudness near threshold for hearing-impaired ears," J. Acoust. Soc. Am. **115**, 3103–3111.

Stevens, S. S. (**1955**). "The measurement of loudness," J. Acoust. Soc. Am. **27**, 815–829.

Whilby, S., Florentine, M., Wagner, E., and Marozeau, J. (**2006**). "Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing," J. Acoust. Soc. Am. **119**, 3931–3939.

Investigating the effects of stimulus duration and context on pitch perception by cochlear implant users

Joshua S. Stohl, Chandra S. Throckmorton, and Leslie M. Collins[a)]
Department of Electrical and Computer Engineering, Duke University, 129 Hudson Hall, Box 90291, Durham, North Carolina 27708-0291

(Received 28 June 2008; revised 21 April 2009; accepted 22 April 2009)

Cochlear implant sound processing strategies that use time-varying pulse rates to transmit fine structure information are one proposed method for improving the spectral representation of a sound with the eventual goal of improving speech recognition in noisy conditions, speech recognition in tonal languages, and music identification and appreciation. However, many of the perceptual phenomena associated with time-varying rates are not well understood. In this study, the effects of stimulus duration on both the place and rate-pitch percepts were investigated via psychophysical experiments. Four Nucleus CI24 cochlear implant users participated in these experiments, which included a short-duration pitch ranking task and three adaptive pulse rate discrimination tasks. When duration was fixed from trial-to-trial and rate was varied adaptively, results suggested that both the place-pitch and rate-pitch percepts may be independent of duration for durations above 10 and 20 ms, respectively. When duration was varied and pulse rates were fixed, performance was highly variable within and across subjects. Implications for multi-rate sound processing strategies are discussed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3133246]

PACS number(s): 43.66.Ts, 43.66.Hg, 43.66.Fe [BCM] Pages: 318–326

I. INTRODUCTION

Individuals with cochlear implants (CIs) often exhibit speech recognition performance in quiet conditions that is comparable to that of normal-hearing individuals; however, there continue to be situations in which discrepancies in performance between implant users and normal-hearing individuals exist. For example, speech recognition in situations where there is environmental noise, or there are multiple talkers, remain a challenge for implant users. In addition, the subtle pitch cues that are used to communicate in tonal languages are not accurately conveyed via current implants (Wei *et al.*, 2004; Kong and Zeng, 2006), and the lack of fine spectral cues also renders music identification and appreciation a challenge (Gfeller *et al.*, 2007; McDermott, 2004).

The majority of clinical speech processing algorithms divide the acoustic input obtained from the microphone of the CI into M subbands that correspond to M electrodes, extract the slowly-varying temporal envelope within each band, and use those envelopes to modulate fixed-rate biphasic pulse trains that are presented on N of the M implanted electrodes, $N \leq M$, (i.e., Wilson *et al.*, 1991; Kiefer *et al.*, 2001; Koch *et al.*, 2004). While many subjects are able to use envelope cues to achieve high speech recognition scores in quiet conditions, it has been hypothesized that representing the fine structure (FS) of the input signal may be necessary to obtain further improvements in CI user performance on more challenging tasks such as those mentioned above (Wilson *et al.*, 2004). The envelope and FS may be defined in the following way: A signal may be represented analytically as $s(t)=A(t)\cos\phi(t)$, where $A(t)$ may be referred to as the envelope, and $\cos\phi(t)$ may be referred to as the FS, with $\phi(t)=\arctan(x_i(t)/x_r(t))$, and $x_i(t)$ being the Hilbert transform of the original signal, $x_r(t)$. It is possible that some of the FS information is contained in the envelope of the signal if this information is below the cutoff frequency used during envelope extraction, and the stimulation rate is high enough to adequately represent that information. However, the majority of clinically available speech processing algorithms do not explicitly encode FS (Wilson *et al.*, 2004; Zeng, 2004).

One proposed method of increasing the amount of FS information that is available to CI users via a CI speech processing strategy is to encode the FS by using time-varying stimulation rates (Grayden *et al.*, 2004; Büchner *et al.*, 2005; van Hoesel and Tyler, 2003; Lan *et al.*, 2004; Nie *et al.*, 2005; Throckmorton *et al.*, 2006; Arnoldner *et al.*, 2007). Unlike in acoustic hearing, the place- and rate-pitch mechanisms may function independently in electric hearing (Tong *et al.*, 1983; McKay *et al.*, 2000), and these proposed multi-rate strategies use the independence of place and rate in an attempt to improve upon existing algorithms that transmit spectral cues only as a function of electrode placement in the cochlea. By conveying fundamental frequency (F0) (Lan *et al.*, 2004; Büchner *et al.*, 2005), the dominant frequencies in a bank of subbands (Nie *et al.*, 2005; Throckmorton *et al.*, 2006), or fine timing cues (van Hoesel and Tyler, 2003; Grayden *et al.*, 2004; Arnoldner *et al.*, 2007) with time-varying stimulation rates, researchers hope to provide additional information about the spectrum of the acoustic signal within the constraints of existing implant hardware.

The implementation of a CI strategy that incorporates time-varying stimulation rate requires practical knowledge regarding the perception of such stimuli. Specifically, the pulse rate difference and the minimum duration of any single rate stimulus required for distinct percepts across rates may

[a)]Author to whom correspondence should be addressed. Electronic mail: lcollins@ee.duke.edu.

 0001-4966/2009/126(1)/318/9/$25.00

TABLE I. Demographic information for implanted subjects.

Subject ID	Gender	Age (year)	Age at onset of deafness (years)	Age at implantation (years)	Mode of stimulation	Speech recognition (percent correct)
S2	F	72	46	66	MP1+2	92
S5	F	59	26	54	MP1+2	90
S7	M	53	50	50	MP1+2	97
S8	M	55	35	52	MP2	100

be incorporated into a multi-rate strategy. Historically, researchers have investigated the first issue, the ability of CI users to discriminate one pulse rate from another, when each pulse rate is presented in isolation, i.e., with stimuli separated by an inter-stimulus interval (Fearn *et al.*, 1999; McDermott and McKay, 1997; Pijl, 1995; Shannon, 1983; Tong and Clark, 1985; Townshend *et al.*, 1987; Zeng, 2002). These studies demonstrated that this percept saturates somewhere between 300–500 pps (pulses per second) for most implant users. However, proposed multi-rate sound processing strategies typically incorporate time-varying signals in which pulse rate may be changing rapidly and without the relatively long interstimulus gaps that are used in most rate discrimination studies (approximately 200–500 ms).

Rate-based psychophysical experiments in CI subjects have traditionally focused on frequency modulation (FM) detection rather than the investigation of perceptual responses to instantaneous changes in stimulation rate (Chen and Zeng, 2004; Tong *et al.*, 1982; Luo and Fu, 2007). FM difference limens (FMDL) increased with increasing base rate (Chen and Zeng, 2004), and when coupled with amplitude modulation (AM), FMDL increased with an increase in AM depth. This finding is relevant to multi-rate CI strategies in which AM is used to convey the slowly varying envelope information within each band, and simultaneous changes in stimulation rate are used to transmit FS information. Expanding upon previous studies, the experiments presented below were designed to measure pulse rate difference limen (PRDL) for stimuli containing relatively large (i.e., 200 pps), instantaneous changes in rate. These data are applicable to multi-rate strategies that propose quantizing the variations in the FS and mapping those changes on to predefined stimulation rates (e.g., Throckmorton *et al.*, 2006).

In addition to investigating the rate difference required for listeners to detect an instantaneous change in rate, the influence of duration on the rate-pitch percept is of interest for implementation of a multi-rate strategy. Stimulation rate may be updated as often as on a pulse-by-pulse basis, yet there remains some question about whether or not rapid changes in stimulation rate are detectable by CI listeners. Chen and Zeng (2004) observed a significant difference between FMDLs at low sinusoidal FM rates (<80 Hz) and higher FM rates (160 and 320 Hz), suggesting that some threshold may exist at which the ability to identify rapid changes in rate increases in difficulty for CI listeners. Tong *et al.* (1982) investigated frequency discrimination as a function of duration using linear FM sweeps. Decreasing the duration of the FM sweep resulted in a decreased ability to detect the difference between a stimulus with a linear sweep and a stimulus with a constant pulse rate for the single subject tested. These FM studies suggest that both the range of pulse rates and the duration of each rate may affect the ability of strategies that rely on time-varying stimulation rate to convey FS cues. In this study, direct measurement of these two variables was conducted using quantized changes in rate.

The study consisted of three experiments. First, the consistency of the place-pitch percept was measured as a function of duration using a pitch ranking task. The ability to consistently pitch rank electrodes independent of duration would imply that salient percepts remained when presenting single-rate stimuli at short durations. The subsequent experiments were designed to investigate the pulse rate separation and single-rate duration required for distinct percepts on one electrode. PRDLs were measured with an adaptive procedure for stimuli both in isolation and when embedded (i.e., with no interstimulus interval between stimuli with different pulse rates). Repeating the experiment in both conditions allowed for the comparison of PRDLs measured with rapid transitions in pulse rate to traditional PRDLs. This also provided a direct measurement of the ability to detect rate changes as they would be presented in a multi-rate sound processing strategy. In the third experiment, embedded PRDLs were used to set pulse rates, and an adaptive procedure was implemented to determine the minimum duration required for detection of a change between two fixed pulse rates with no interstimulus interval between them. The goal of this experiment was to determine how often rate may be changed while still evoking a perceptual change. These experiments may provide insight regarding the ways in which time-varying pulse rates may best be implemented in a CI speech processing algorithm to provide maximum benefit to the individual user.

II. METHODS

A. Subjects and stimuli

Four postlingually deafened subjects participated in the experiments described below. Demographic information for the subjects is presented in Table I. All subjects were implanted with a version of Cochlear Corporation's CI24 implant and had a minimum of 3 years experience with their device prior to testing. Testing occurred over six to nine sessions, and each session lasted between 2 and 4 h. All subjects were paid for their time except for subject S7 who elected to volunteer his time. These experiments were ap-

proved by the Institutional Review Board at Duke University, as was subject S7's voluntary participation.

Pulse trains consisting of biphasic rectangular pulses with 25 μs pulse widths and an 8 μs interphase gap were presented via the SPEAR3 research sound processor (Stohl *et al.*, 2008). Only electrodes available in the subjects' clinical MAP were used for testing, and threshold and comfort levels were measured at the beginning of each test session. While all four subjects use a monopolar 1+2 (MP1+2) mode of stimulation in their clinical devices, only subjects S2, S5, and S7 were stimulated using an MP1+2 mode in this study. Subject S8 was stimulated in Monopolar 2 mode due to safety concerns regarding the use of the Freedom Implant (CI24RE) and MP1+2 stimuli with the SPEAR3.[1]

B. Experiment 1

In general, electrical stimulation via a multichannel CI elicits percepts that follow the tonotopic arrangement of the cochlea (Eddington *et al.*, 1978; Townshend *et al.*, 1987). The first experiment was designed to determine what effects, if any, duration had on pitch ranking due to place of stimulation alone.

1. Stimuli and experimental task

A two-interval, forced-choice, pitch ranking task was used. In each trial, subjects were presented with two intervals and instructed to pick the interval containing the sound with the higher pitch (Townshend *et al.*, 1987). All pulse trains were presented at a rate of 200 pps throughout this experiment. Pulse train durations were 10, 20, 50, 100, and 200 ms, and the interstimulus interval was 300 ms. An experimental block consisted of one comparison of each active electrode to all other active electrodes at a single duration, and a set included the presentation of one block at all of the durations listed above. Electrode pairs were chosen randomly within each block and were presented in a random order during each trial. Blocks were presented from the longest duration to the shortest, and stimuli were presented at all durations before the set of durations was repeated. Each set was repeated three to seven times.

All stimuli were loudness balanced across electrode prior to testing, and subjects S5 and S7 required increased stimulation levels for stimuli with shorter durations. Estimates of equal loudness were obtained via the method of adjustment. Two pulse trains were presented repeatedly in an alternating fashion. The first, or reference, stimulus was a 200 pps pulse train presented on the most apical active electrode at a comfortable level determined by listener before the task, and the second, or target, stimulus was a pulse train presented on the adjacent active electrode in the basal direction. The subject was instructed to use a knob that was connected to the test personal computer (PC) to adjust the volume (current level) of the target stimulus. The initial amplitude of the target stimulus was between 5 and 10 current steps below the amplitude of the first stimulus depending on the dynamic range for that electrode. A repeat measurement was made with the initial amplitude of the target stimulus set above the reference stimulus amplitude but below the maximum comfortable loudness (MCL) for the electrode being tested. If the two estimates of equal loudness were separated by more than four current steps, this procedure was repeated. The resulting amplitudes were arithmetically averaged to determine the level at which the two stimuli were considered to be equally loud. The target electrode became the reference electrode, and the next active electrode in the basal direction became the target electrode for the next trial. All electrodes were loudness balanced to the adjacent apical electrode using this procedure (Throckmorton and Collins, 2001).

2. Results

Pitch ranking data were analyzed using row sum analysis (Collins *et al.*, 1997; David, 1988), and the results are presented in Fig. 1. The horizontal axis shows electrode number, where electrode 1 is the most basal electrode in the array and numbering increases toward the apex. The vertical axis shows each electrode's percent wins. This value is the percentage of time that a given electrode was ranked as higher in pitch than all other active electrodes. Each duration is indicated by a unique line style and is also accompanied by a coefficient of consistence in the legend labeled ζ. The coefficient of consistence, $\zeta \in [0,1]$, is unity when no inconsistencies occur in the ranking of the stimuli throughout the experiment (Kendall and Smith, 1940).

As can be seen in Fig. 1, pitch ranking as a function of place generally remained consistent when pulse rate was fixed, regardless of the stimulus duration. Furthermore, subjects were able to rank electrodes with the same consistency across durations, as is indicated by stable values of ζ for all durations. These data suggest that a pitch percept is available at durations as short as 10 ms, and that this percept remains constant with respect to all other electrodes in the array.

C. Experiment 2

One goal of this experiment was to determine PRDL as a function of duration. While PRDLs are typically measured with an interval between two stimuli with different pulse rates (e.g., Zeng, 2002), this does not accurately reflect the time-varying stimulation rate that would be observed on any single electrode for a multi-rate strategy. Therefore, this experiment was designed to test CI subjects' ability to discriminate pulse rates that change instantaneously, i.e., without an interstimulus interval. Another goal of this experiment was to determine what impact changing rate instantaneously may have on PRDLs, relative to those measured in the traditional fashion. Therefore, PRDLs were measured for both the traditional (isolated) and embedded patterns of stimulation.

1. Stimuli and experimental task

Two rate discrimination tasks were implemented. Both tasks used four-interval, two-alternative, forced-choice, adaptive procedures (Levitt, 1971) in which the subject was instructed to identify the interval that sounded "different." Intervals 2 and 3 contained possible targets, and intervals 1 and 4 were always fixed as reference intervals. Subjects selected the interval that they perceived to be different. In the

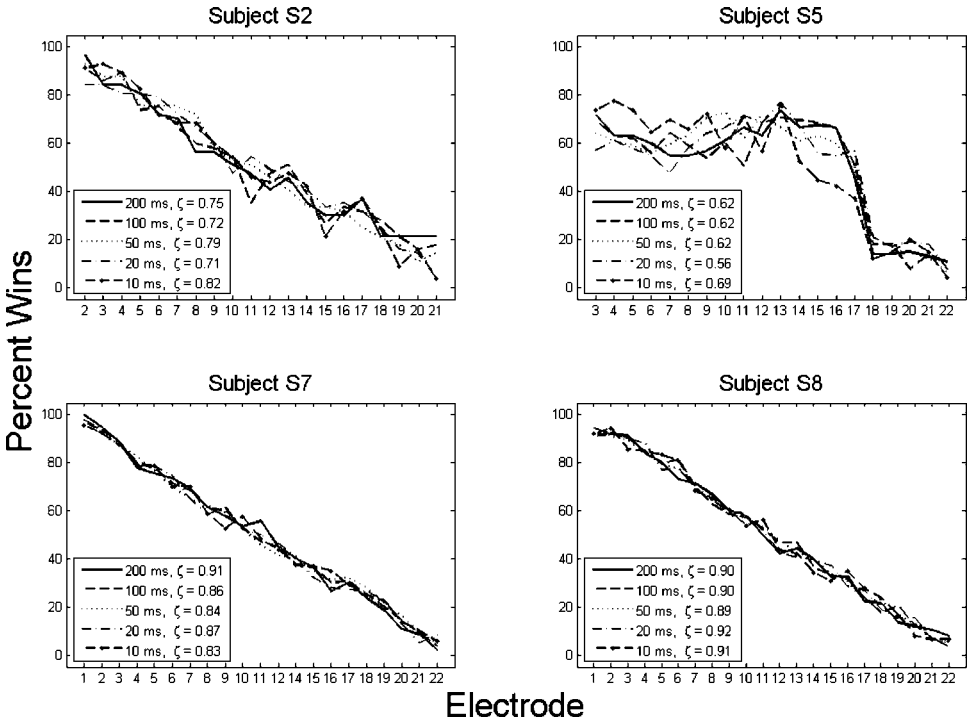


FIG. 1. Row sum analysis results of pitch ranking data are plotted for four CI subjects. The horizontal axis indicates electrode, and the vertical axis indicates the corresponding percent wins. Pulse train duration is indicated by line style, and the mean coefficient of consistence (ζ) for each duration is listed in the legend.

first task, subjects heard 200 ms fixed-rate pulse trains. The base or reference rate was always 200 pps, and the target rate was higher and changed from trial-to-trial. This task will be henceforth referred to as isolated rate discrimination (IRD), given that the stimuli associated with different pulse rates were always isolated from one another by a 500 ms interstimulus interval. During the second task, the target interval was of an ABA format in which the rate of A was always the reference rate of 200 pps, and the rate of B varied from trial-to-trial. Reference intervals were always presented at a constant 200 pps (AAA). Stimuli were approximately 600 ms in duration, and this task was repeated using four durations for rate B (20, 50, 100, and 200 ms). For each trial, the number of rate B pulses was calculated such that the error between the actual duration and the desired duration was minimized, where duration refers to the time between the onset of the first pulse and end the of the period of the final pulse (i.e., 4 pulses at 200 pps has a duration of 20 ms). The number of rate A pulses was then calculated so that the total duration of the target interval was as close to 600 ms as possible, and rate B was always embedded in the middle of the stimulus. The second task will be henceforth referred to as embedded rate discrimination (ERD), as the rate B pulse train may be thought of as embedded in a pulse train of rate A. Figure 2(a) illustrates an example target stimulus with an embedded rate change.

Stimuli were loudness balanced via the method of adjustment prior to testing, and amplitudes were roved an average of ± 4 current steps[2] to minimize loudness cues due to differences in stimulation rate (McKay *et al.*, 2001). Each interval was roved in the IRD task, whereas in the ERD task, each third of each 600 ms interval was roved independently. The durations of each of the three A sections of the reference intervals were equal to the desired target durations (e.g., 275-50-275 ms) to minimize any cues arising from recognition of any residual loudness differences at the rate transitions in the target stimulus that remained after loudness balancing the stimuli across pulse rate. The stimulation levels of each third of the reference and target stimuli were roved to prevent changes in loudness in the target interval from providing an additional cue to the subjects.

Each task was considered complete after 12 reversals or 60 trials, whichever occurred first. For the first four reversals, a one-down, one-up rule was used, and a two-down, one-up rule was applied for the final eight reversals. The step size was a factor of 1.4, and the geometric mean of the DL at the final eight reversals was taken as the DL that corresponds to 70.7% probability of a correct decision (Levitt, 1971). For the cases in which 60 trials preceded 12 reversals, the geometric mean of the DL at the last N reversals was computed and taken as the DL, where N was even and $4 < N < 12$. Sixty trials were reached in less than 3% of all repetitions and were never reached for subject S5.

Difference limens (DLs) were measured for three electrode locations for each subject, one basal, one middle, and

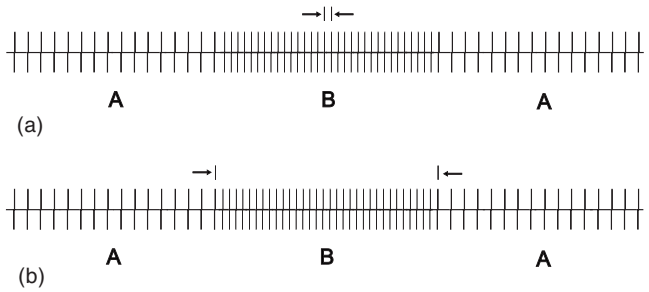


FIG. 2. Examples of the stimuli used in the ERD task and the MDD task, respectively. During the ERD task, the rate of section B is varied while the duration of section B is held constant. During the MDD task, the rate of section B is fixed, and its duration is varied adaptively.

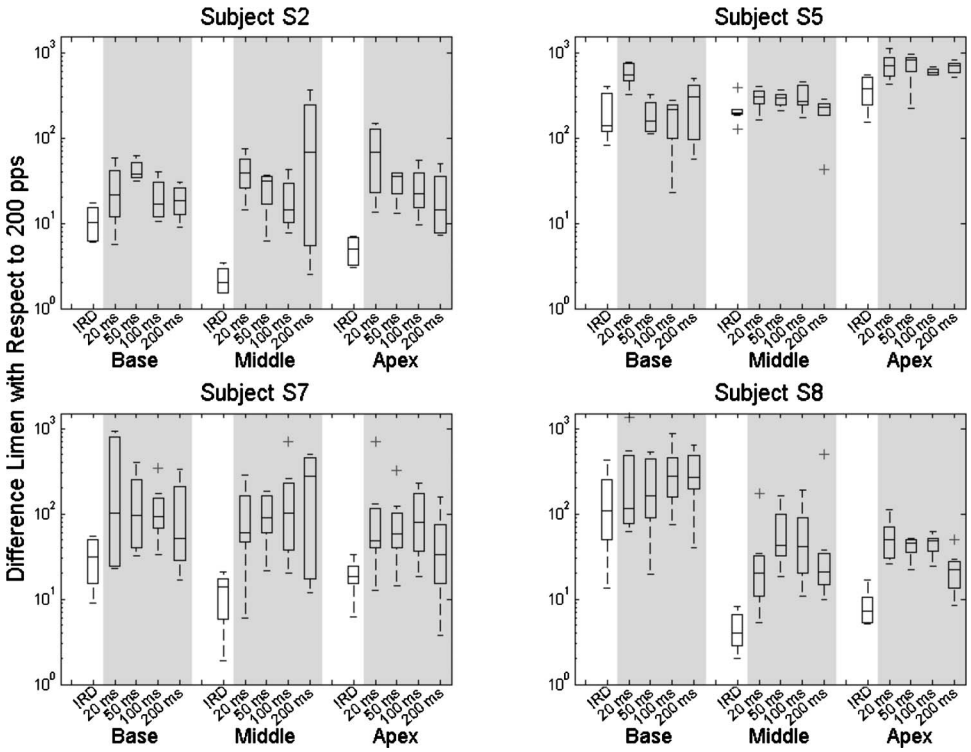


FIG. 3. Isolated and embedded rate DLs for all four subjects as a function of electrode and the duration of B in the ABA target stimulus. ERDDL is marked by a gray background. The median DL, upper and lower quartiles, and 1.5 times the interquartile difference are plotter for each condition. Outliers are indicated by a plus (+) symbol.

one apical. IRDDLs were measured between four and eight times with an average of approximately seven measurements per subject, and ERDDLs were measured between four and ten times with an average of approximately seven measurements per subject. DLs were measured for all three electrodes in a random order for a given task before repeat measurements were taken. The tasks were interleaved randomly as well.

2. Results

IRDDL and ERDDL are plotted as a function of electrode in Fig. 3, with ERDDL further subdivided by the duration of B and indicated by a gray background. The median DL, upper and lower quartiles, 1.5 times the interquartile range, and outliers, as indicated by a plus symbol, are plotted for each DL (McGill *et al.*, 1978). Data for all four subjects are shown, and subjects are grouped by panel. Nonparametric statistics were used to analyze these data. The use of nonparametric statistical analysis eliminates the need to assume that observations of the DL are drawn from some specific underlying distribution and are appropriate for analyzing these data given the relatively small sample size (four to eight observations per DL) (Moses, 1952).

A Kruskal–Wallis test, an extension of the two-class Wilcoxon rank sum test and a nonparametric one-way analysis of variance that tests for equal medians across k classes, $k>2$, was performed for each subject and electrode with duration as factor (Gibbons, 1985). Due to multiple comparisons being made with the same data, Bonferroni adjustment was applied, and in general ERDDLs were not significantly affected by duration ($p<0.05$). Two exceptions were found. The ERDDL for S5's electrode 4, when the duration of B was 20 ms, was significantly higher than the ERDDL measured when B was 50 and 100 ms ($p<0.02$). A significant difference was also found between the ERDDL for S8's electrode 20 when measured at a duration of B equal to 200 ms and compared to measurements taken when the duration of B was equal to 20, 50, and 100 ms ($p<0.02$).

Given that there was generally no significant difference in ERDDL as a function of duration, ERDDLs were collapsed and the same analysis was applied to ERDDL within subject with electrode as factor. ERDDL and IRDDL are plotted for each subject and the group in Fig. 4. The group ERDDL includes all measurements, whereas those ERDDLs that were determined to be significantly different as a function of electrode were removed in the individual data shown.

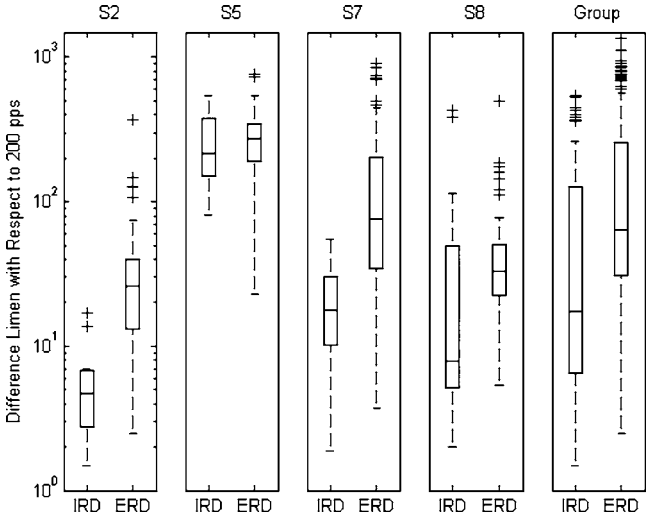


FIG. 4. Isolated and embedded rate DLs for all four subjects and the group. ERDDLs shown in Fig. 3 have been combined for each subject and electrode due to the lack of significant effect of duration. DL medians, upper and lower quartiles, and whiskers indicating 1.5 times the interquartile difference are shown. Outliers are indicated by a plus (+) symbol.

TABLE II. Stimulation rate of B in ABA stimulus, amplitude used for reference rate A (200 pps) and amplitude used for B that was adjusted for equal loudness with respect to A for used in Experiment 3. All amplitudes are listed in current steps (0–255) by subject and electrode.

Subject	S2			S5			S7			S8		
Electrode	3	11	20	4	13	21	3	11	20	3	12	20
Rate of B	400	400	400	400	500	700	500	500	500	600	400	400
Amplitude of A	177	196	193	163	175	176	199	202	198	151	149	146
Amplitude of B	176	190	193	161	172	172	198	200	197	150	145	145

A significant effect of electrode was observed in two cases. ERDDLs on S5's electrode 21 and S8's electrode 3 were significantly higher than the other two electrodes tested for each subject ($p<0.001$), and both ERDDLs were also significantly higher than the IRDDL for the same electrode ($p<0.001$). When removing these two electrodes and comparing IRDDL and ERDDL within subject via the Wilcoxon rank sum test that checks for equal medians between two groups, ERDDL was significantly higher than IRDDL for S4, S7 and S8 ($p<0.002$). S5 suffers from tinnitus and thus had a competing percept that made it difficult to perform tasks involving pitch. This is reflected in elevated IRDDLs, and the resulting lack of significant difference between IRDDLs and ERDDLs. Nonetheless, group ERDDLs were observed to be higher than group IRDDLs ($p<0.001$) when pooling data from all four subjects. The median IRDDL for the group for all electrodes tested was found to be approximately 17.5 pps, and the median ERDDL for the group across electrode and duration was found to be approximately 63.8 pps. These DLs expressed as Weber fractions with respect to 200 pps are 0.09 and 0.31 for IRDDL and ERDDL, respectively. The significant increase in DL suggests that DLs should be measured in the embedded condition when determining discriminable rates for use in a multi-rate sound processing strategy.

D. Experiment 3

The lack of effect of duration on rate discrimination in the embedded case has implications for implementation of a multi-rate CI strategy. However, changing rate over a fixed duration does not necessarily reflect what may happen in a class of multi-rate strategies in which the stimulation rates are predefined (Throckmorton *et al.*, 2006) and the selected rate of stimulation is a time-varying function of the location of spectral peaks in the input signal. Furthermore, there is evidence in the normal-hearing literature that suggests that uncertain time of arrival and uncertain duration may both have a negative impact on signal detection ability (Wright and Fitzgerald, 2004; Wright, 2005). Therefore, in the third experiment the duration of the embedded stimulus was varied adaptively and rate of stimulation was fixed. Results from this experiment may provide insight into the minimum duration required for listeners to detect a change in stimulation rate when pulse rates are fixed. An example of the target stimulus used in the minimum detectable duration (MDD) task can be seen in Fig. 2(b). This is in contrast to the second experiment in which duration was fixed and rate varied adaptively [see Fig. 2(a) for comparison].

1. Stimuli and experimental task

The stimuli used in Experiment 3 were identical to those used in Experiment 2, except that the duration of the middle third of each stimulus was varied adaptively on a trial-by-trial basis instead of the stimulation rate. The stimulation rates of A and B were fixed, and the duration of the first and third sections of the stimuli were adjusted to maintain a total interval duration of approximately 600 ms. Each interval was separated by a 500 ms interstimulus interval. The reference rate was fixed at 200 pps for all subjects; however, the rate of B was selected using the ERDDL data obtained in Experiment 2 and was always at least 400 pps. Stimuli of different pulse rates were loudness balanced prior to testing using the same method as Experiment 2, and amplitudes were roved by an average of 4 current steps above and below the loudness-balanced levels. Subject specific stimulation rates that were used for the middle section of the target stimulus (B of ABA) are listed in Table II along with the corresponding current levels required for equal loudness.

The same procedure that was used in Experiment 2 was also used in Experiment 3. Twelve reversals were reached in all cases but one, in which S5 reached 60 trials on her first attempt at the task. In that case, eleven reversals occurred, and the geometric mean of the MDD at the last six reversals was taken as the MDD for that experimental run. Four to eight MDD measurements were taken using each electrode for each subject. Measurements were taken using all three electrodes of interest in a random order before repeating the task.

2. Results

The MDD for a fixed difference in rate is plotted in Fig. 5 as a function of electrode location for each subject as well as across electrode for the group, with each subject and the group in different panels. As before, each MDD is represented by a horizontal line at the median, lines at the upper and lower quartiles, and whiskers that indicate values within 1.5 times the interquartile range. MDDs outside of the whiskers are marked with a plus symbol and are considered outliers.

As is shown in Fig. 5, there is substantial variability of MDD within and across electrode for each subject and across subject as well. When applying a Kruskal–Wallis test with Bonferonni adjustment, no significant difference was observed between MDD when the factor was electrode for subjects S2 and S5; however, the MDD for S7's electrode 20 with the rate of A fixed at 200 pps and the rate of B fixed at 500 pps was significantly lower than the other two electrodes tested ($p<0.05$). Similarly, the MDD for S8's electrode 20

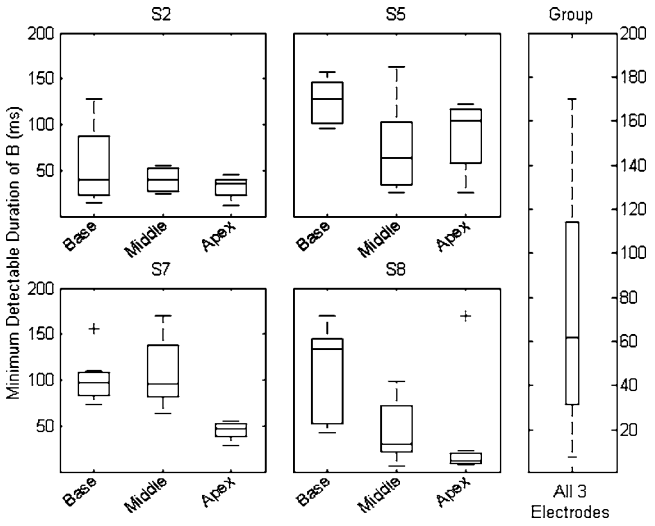


FIG. 5. The minimum duration required to detect an embedded rate change is plotted as a function of electrode location for each subject and for the group irrespective of electrode. Data for each subject are contained in a separate panel. Associated with each electrode for each subject as well as for the group is the MDD median, upper and lower quartiles, whiskers extending to those points inside of 1.5 times the interquartile range, and outliers indicated with a plus (+) symbol.

was significantly lower than the MDD for electrode 3 with the rates of B equal to 400 and 600 pps, respectively ($p < 0.05$).

This variability suggests that the task was extremely challenging for some subject and electrode combinations. Similar inter-electrode trends and intra-electrode variability can be seen when comparing the ERDDL and MDD data. The presence of MDD values of 100 ms and higher suggests that in some cases, sound processing algorithms may need to limit how often stimulation rate changes in order to be effective.

III. DISCUSSION

The results from Experiment 1 suggest that place-pitch is not a function of duration in electric hearing for durations greater than 10 ms. Tong *et al.* (1982) found that the ability to discriminate stimuli containing shifts in electrode position from stimuli with static electrode position was independent of duration for one subject. Tong *et al.* (1982) measured the ability to discriminate place as a function of duration using a same-different task, and the pitch ranking results from Experiment 1 expanded upon that finding by using a larger number of electrodes (22 versus 4) and more subjects. Tong *et al.* (1982) concluded that the ability to detect rapid transitions in electrode position suggested that place of stimulation may be effectively used to transmit rapidly varying (segmental) speech information, and this hypothesis is supported by the pitch ranking data collected in the first experiment.

Comparing the IRDDL and ERDDL data collected in Experiment 2 indicated a significant increase in PRDL in the embedded case. This increase in PRDL may be due in part to the fact that the stimuli used in the ERD task were more complicated as they had longer durations and a higher number of level changes (12 total changes in level in a single trial as opposed to only 4 changes in level in a single trial of the IRD task). The ERD task may have therefore required a greater amount of cognitive effort on the part of the listener. Nonetheless, the stimuli used to collect ERD data better mimic the stimulation patterns that may be used in a multirate startegy, and thus when determining the appropriate rates for use in such a strategy, it may be more appropriate to measure the DLs without an interstimulus interval between rate changes. Results from Experiment 2 also imply that changes in stimulation rate are detectable for durations as brief as approximately 20 ms, and that the ability to detect an embedded change in rate is not significantly impacted by duration when the duration of the embedded stimulus is fixed across observations and greater than 20 ms. These results, in combination with those data obtained in Experiment 1, suggest that in electric hearing the ability to detect changes in both the place and rate of stimulation are independent of duration above some minimum duration. Chen and Zeng (2004) observed a similar lack of significant effect of duration, or FM rate, when measuring sinusoidal FM detection at low FM rates (≤ 80 Hz); however, there was a significant increase in FMDLs at high sinusoidal FM rates (160 and 320 Hz) when compared to low FM rates. The lack of effect of duration on FM detection at longer durations, or lower FM rates, is similar to the trend observed in the data collected in Experiment 2 in which changes in rate were quantized.

In Experiment 3, MDD was measured with one pair of pulse rates for each electrode per subject. Results showed considerably more variability both within and across subject than the ERDDL results from Experiment 2, and no clear lower bound on the ability to detect instantaneous changes in rate was observed in general. In addition to measuring the detectability of shifts in electrode position, Tong *et al.* (1982) measured the ability of one subject to detect linear changes in pulse rate on single electrodes, and found an inability to perform FM sweep detection when the duration of the sweep was less than 25 ms. Note that in that study, the onset of the sweep was consistent across trial, as was the case in Experiment 2 of this study. Although inter-electrode trends were similar when comparing the ERDDL and MDD data from this study, data from Experiment 3 suggest that a time-varying duration that changes from trial-to-trial or observation-to-observation may hinder performance of embedded pulse rate change detection. It has been suggested in the normal-hearing literature that when performing a signal detection task listeners form a template for the target signal that includes both frequency and duration information (Wright, 2005), and it has also been shown that signal detection performance worsens when the onset time of the target stimulus is unknown (Wright and Fitzgerald, 2004). In Experiment 2, only the frequency, or pulse rate, of B changed from trial-to-trial, while the duration and onset time of B remained relatively constant. In Experiment 3, both the duration and the onset of time of B changed from trial-to-trial while the pulse rates for A and B remained constant, and it is possible that changing these two features (duration and onset time) between each trial resulted in a much more challenging task for the subjects that participated in this study.

Tong *et al.* (1982) concluded that while pulse rate cues may be appropriate for transmitting longer duration (supra-

segmental) information contained in the fundamental frequency of speech, the inability to perform FM sweep detection when the duration of the sweep was less than 25 ms may prevent these cues from being used to transmit segmental information. This supposition is supported by the variability in the data collected in Experiment 3 and that many of the MDDs obtained in Experiment 3 were greater than the average duration of an English phoneme (Umeda, 1975, 1977). Thus, using rate to transmit information about changes in a speech spectrum may not be effective for all subjects. In the event that rate update durations are restricted to the values determined by Experiment 3, spectral smearing across phonemes is likely to occur. However, the ability of some subjects to identify changes in rate at durations well below the median MDD (i.e., Subject S8, electrode 20) suggests that further investigation may be necessary to identify a true lower bound on rate change detection. As can be observed by comparing ERDDL from Experiment 2 summarized as the rate of B for each electrode in Table II, the probability of accurately detecting rate B with respect to rate A (200 pps) may have varied according to subject and electrode. Repeating Experiment 3 with multiple pulse rates assigned to B in an ABA stimulus and varying the duration of B (and thus A) adaptively would potentially provide a better representation of the interaction between duration and CI listeners' ability to detect changes in pulse rate.

Given that the ultimate goal of using multiple stimulation rates is to transmit FS cues in a CI speech processing algorithm, it may be most beneficial to determine sets of discriminable rates that allow pulse rate to be updated as often as possible, thus potentially transmitting the most information per unit time to the listener. Furthermore, a multi-rate CI speech processing algorithm will necessarily include AM in addition to time-varying rates, and extending the work of Luo and Fu (2007) to include changes in pulse rate that are quantized instead of slowly varying may provide further insight into the interaction of pulse rate, duration, and amplitude change.

IV. CONCLUSION

Improvement in speech recognition with multi-rate strategies may require subject specific tuning as supported by the variability in the data collected in this study. Specifically, ERDDL and MDD both varied considerably within and across subject, and parameters such as rate of stimulation and minimum duration for a single pulse rate may need to be adjusted according to subject-specific psychophysical data for users to take advantage of the additional information encoded via pulse rate. Based on the results from Experiment 2, PRDLs measured in the embedded case are more likely to reflect detectable pulse rates in a multi-rate strategy, but the duration at which they are measured may not be a significant factor above approximately 20 ms. Results from Experiment 3 suggest that update durations that provide discriminable changes in rate are both subject and electrode position specific. These differences have the potential to limit a multi-rate strategy's ability to transmit rapidly varying spectral information. Even if only for suprasegmental aspects of speech, the use of multi-rate strategies may still provide additional, usable information that when deliberately included may aid in speech recognition in noisy conditions. Inconsistencies observed in the data from Experiment 2 and Experiment 3 indicate the need for further investigation on how to use pulse rate to encode FS information and how to best implement a multi-rate sound processing strategy such that the listener is able to extract the FS information that is available.

ACKNOWLEDGMENTS

The authors would like to thank the subjects who participated in this study for their time and patience, and the anonymous reviewers and Brian Moore for their comments and suggestions on this manuscript. This work was supported by the National Institutes of Health R01-DC007994.

[1]The steps required to safely present stimuli to a Freedom Implant in Monopolar 1+2 mode via the SPEAR3 have been implemented since the completion of this experiment (Personal Communication: Colin Irwin of Cochlear Corporation). See Stohl *et al.*, 2008 for details.

[2]A current step is a unit defined by Cochlear Corporation, and has a logarithmic relationship to current: $I=a175C/255$, where I is current in amperes, $a=10$ μA, and $C\in[0,255]$ is current step.

Arnoldner, C., Riss, D., Brunner, M., Durisin, M., Baumgartner, W.-D., and Hamzavi, J.-S. (**2007**). "Speech and music perception with the new fine structure speech coding strategy: Preliminary results," Acta Oto-Laryngol. **127**, 1298–1303.

Büchner, A., Edler, B., and Nogueira, W. (**2005**). "Fundamental frequency coding in NofM strategies for cochlear implants," in Audio Engineering Society 118th Convention Barcelona, Spain, p. 6515.

Chen, H., and Zeng, F.-G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–455.

David, H. (**1988**). *The Method of Paired Comparisons* (Oxford University Press, New York).

Eddington, D., Dobelle, W., Brackmann, D., Mladejovsky, M., and Parkin, J. (**1978**). "Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers," Trans. Am. Soc. Artif. Intern. Organs **24**, 1–5.

Fearn, R., Carter, P., and Wolfe, J. (**1999**). "The dependence of pitch perception on the rate and place of stimulation of the cochlea: A study using cochlear implants," Ear Hear. **27**, 41–43.

Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., and Olszewski, C. (**2007**). "Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise," Ear Hear. **28**, 412–423.

Gibbons, J. D. (**1985**). *Nonparametric Statistical Inference*, Statistics: Textbooks and Monographs, Vol. **65**, 2nd ed. (McGraw-Hill, Dallas, TX).

Grayden, D., Burkitt, A., Kenny, O., Clarey, J., Paolini, A., and Clark, G. (**2004**). "A cochlear implant speech processing strategy based on an auditory model," in Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004, pp. 491–496.

Kendall, M., and Smith, B. B. (**1940**). "On the method of paired comparisons," Ear Hear. **31**, 324–345.

Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T., and Gstoettner, W. (**2001**). "Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI24M cochlear implant system," Audiology **40**, 32–42.

Koch, D. B., Osberger, M. J., Segel, P., and Kessler, D. (**2004**). "Hiresolution and conventional sound processing in the hiresolutiontm bionic ear: Using appropriate outcome measures to assess speech recognition ability," Audiol. Neuro-Otol. **9**, 214–223.

Kong, Y.-Y., and Zeng, F.-G. (**2006**). "Temporal and spectral cues in man-

darin tone recognition," J. Acoust. Soc. Am. **120**, 2830–2840.
Lan, N., Nie, K., Gao, S., and Zeng, F. (**2004**). "A novel speech-processing strategy incorporating tonal information for cochlear implants," Ear Hear. **51**, 752–760.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Luo, X., and Fu, Q.-J. (**2007**). "Frequency modulation detection with simultaneous amplitude modulation by cochlear implant users," J. Acoust. Soc. Am. **122**, 1046–1054.
McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.
McDermott, H. J., and McKay, C. M. (**1997**). "Musical pitch perception with electrical stimulation of the cochlea," J. Acoust. Soc. Am. **101**, 1622–1630.
McGill, R., Tukey, J. W., and Larsen, W. A. (**1978**). "Variations of box plots," Am. Stat. **32**, 12–16.
McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). "Place and temporal cues in pitch perception: Are they truly independent?," Ear Hear. **1**, 25–30.
McKay, C. M., Remine, M. D., and McDermott, H. J. (**2001**). "Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level and mode of stimulation," J. Acoust. Soc. Am. **110**, 1514–1524.
Moses, L. E. (**1952**). "Non-parametric statistics for psychological research," Psychol. Bull. **49**, 122–143.
Nie, K., Stickney, G., and Zeng, F.-G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.
Pijl, S. (**1995**). "Musical pitch perception with pulsatile stimulation of single electrodes in patients implanted with the nucleus cochlear implant," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 224–227.
Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.
Stohl, J. S., Throckmorton, C. S., and Collins, L. M. (**2008**). "Developing a flexible spear3-based psychophysical research platform for testing cochlear implant users," Technical Report, Duke University, Durham, NC.
Throckmorton, C. S., and Collins, L. M. (**2001**). "A comparison of two loudness balancing tasks in cochlear implant subjects using bipolar stimulation," Ear Hear. **22**, 439–448.
Throckmorton, C. S., Kucukoglu, M. S., Remus, J. J., and Collins, L. M. (**2006**). "Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: Implications for cochlear implants," Hear. Res. **218**, 30–42.
Tong, Y., and Clark, G. (**1985**). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients," J. Acoust. Soc. Am. **77**, 1881–1888.
Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). "Psychophysical studies for two multiple-channel cochlear implant patients," J. Acoust. Soc. Am. **71**, 153–160.
Tong, Y., Dowll, R., Blamey, P., and Clark, G. (**1983**). "Two-component hearing sensations produced by two-electrode stimulation in the cochlea of a deaf patient," Science **219**, 993–994.
Townshend, B., Cotter, N., Compernolle, D. V., and White, R. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–114.
Umeda, N. (**1975**). "Vowel duration in American English," J. Acoust. Soc. Am. **58**, 434–445.
Umeda, N. (**1977**). "Consonant duration in American English," J. Acoust. Soc. Am. **61**, 846–858.
van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
Wei, C.-G., Cao, K., and Zeng, F.-G. (**2004**). "Mandarin tone recognition in cochlear-implant subjects," Hear. Res. **197**, 87–95.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Wilson, B., Sun, X., Schatzer, R., and Wolford, R. (**2004**). "Representation of fine structure or fine frequency information with cochlear implants," in Eighth International Cochlear Implant Conference, edited by R. T. Miyamoto (Elsevier, New York), Vol. **1273**, pp. 3–6.
Wright, B. A. (**2005**). "Combined representations for frequency and duration in detection templates for expected signals," J. Acoust. Soc. Am. **117**, 1299–1304.
Wright, B. A., and Fitzgerald, M. B. (**2004**). "The time course of attention in a simple auditory detection task," Percept. Psychophys. **66**, 508–516.
Zeng, F.-G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.
Zeng, F.-G. (**2004**). "Trends in cochlear implants," Trends Amplif. **8**, T1–T34.

Cantonese tone recognition with enhanced temporal periodicity cues

Meng Yuan[a] and Tan Lee
Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Kevin C. P. Yuen
Department of Otorhinolaryngology Head and Neck Surgery, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Sigfrid D. Soli
House Ear Institute, Los Angeles, California 90057

Charles A. van Hasselt and Michael C. F. Tong
Department of Otorhinolaryngology Head and Neck Surgery, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

(Received 30 June 2008; revised 19 March 2009; accepted 20 March 2009)

This study investigated the contributions of temporal periodicity cues and the effectiveness of enhancing these cues for Cantonese tone recognition in noise. A multichannel noise-excited vocoder was used to simulate speech processing in cochlear implants. Ten normal-hearing listeners were tested. Temporal envelope and periodicity cues (TEPCs) below 500 Hz were extracted from four frequency bands: 60–500, 500–1000, 1000–2000, and 2000–4000 Hz. The test stimuli were obtained by combining TEPC-modulated noise signals from individual bands. For periodicity enhancement, temporal fluctuations in the range 20–500 Hz were replaced by a sinusoid with frequency equal to the fundamental frequency of original speech. Tone identification experiments were carried out using disyllabic word carriers. Results showed that TEPCs from the two high-frequency bands were more important for tone identification than TEPCs from the low-frequency bands. The use of periodicity-enhanced TEPCs led to consistent improvement of tone identification accuracy. The improvement was more significant at low signal-to-noise ratios, and more noticeable for female than for male voices. Analysis of error distributions showed that the enhancement method reduced tone identification errors and did not show any negative effect on the recognition of segmental structures.
 [DOI: 10.1121/1.3117447]

PACS number(s): 43.66.Ts, 43.66.Hg, 43.71.Bp, 43.66.Mk [BCM] Pages: 327–337

I. INTRODUCTION

Cantonese is a major Chinese dialect spoken by tens of millions of people in Southern China, including Hong Kong and Macau (Bauer, 1997). It is a tone language, in which pitch of voice carries lexical information (Pike, 1948). Each Chinese character is spoken as a monosyllable with a specific pitch pattern known as lexical tone. If the tone changes, the syllable corresponds to another character that has a different meaning. For example, the syllable /fu/,[1] when associated with different tones, can represent six different characters: "夫" (husband), "虎" (tiger), "富" (rich), "符" (symbol), "妇" (woman), or "父" (father). Figure 1 depicts the illustrative pitch patterns of the six basic tones of Cantonese.[2] Four of them (tones 1, 3, 4, and 6) have flat or slightly falling pitch contours and the other two (tones 2 and 5) have rising contours. This is unlike Mandarin, in which the tones all have distinctive shapes of pitch contours (Xu, 1997).

Pitch is a subjective attribute which is mainly related to the fundamental frequency ($F0$) of voiced speech signals. $F0$ quantifies time-domain periodicity of the signals. In the spectral domain, periodicity is manifested by the harmonic peaks at multiples of $F0$. Perceptually, pitch is detected by both temporal and spectral cues. For normal-hearing listeners, low-order harmonics can be resolved along the basilar membrane (von Bekesy, 1963). Temporal periodicity is encoded via repetitive fluctuation of the stimulus amplitude and by phase locking to individual harmonics (Moore, 2007). The rate of such repetition reflects the pitch-related periodicity. A person with cochlear hearing loss usually has poor frequency selectivity and hence difficulty in resolving harmonics (Moore and Peters, 1992). It was found that listeners with cochlear damage use relatively more temporal information and less spectral information for pitch perception than normal-hearing people (Rosen, 1992). In existing cochlear implant (CI) devices, the electrical stimuli being transmitted to the patients contain primarily temporal cues for pitch in speech (Geurts and Wouters, 2001; Vandali *et al.*, 2005).

[a]Author to whom correspondence should be addressed. Electronic mail: myuan@ee.cuhk.edu.hk

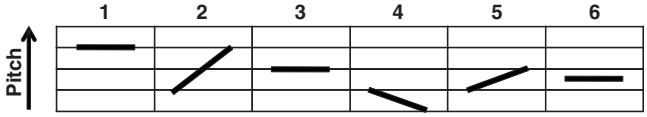


FIG. 1. Pitch patterns of Cantonese tones. The six tones are labeled from 1 to 6, as in the Jyut Ping transcription system (LSHK, 1997).

Temporal cues in speech are roughly divided into two categories, namely, temporal envelope and fine structure. Temporal envelope refers to amplitude variation at a relatively slow rate, e.g., <500 Hz, while fine structure captures the remaining high-frequency fluctuation, e.g., 500–10 000 Hz. The contributions of different types of temporal cues to speech recognition have been studied extensively using acoustic simulations with normal-hearing subjects (Shannon *et al.*, 1995; Dorman *et al.*, 1997; Xu *et al.*, 2002). In these simulations, temporal envelopes extracted from band-pass filtered speech were used to modulate noise signals, so as to generate speech stimuli that contain no fine structure information. For Mandarin tone recognition, high performance levels in quiet were observed with temporal cues below 500 Hz, from four bands (Fu *et al.*, 1998; Xu *et al.*, 2002; Kong and Zeng, 2006). Rosen (1992) distinguished between the periodicity-related temporal fluctuation of 50–500 Hz and the slowly-varying envelope below 50 Hz. Fu *et al.* (1998) and Xu *et al.* (2002) showed that these *F*0-related periodicity cues were very important to Mandarin tone recognition. The recognition accuracy improved noticeably when the cut-off frequency for temporal envelope extraction was increased from 50 to 500 Hz.

In realistic hearing environments, temporal periodicity cues are easily degraded by background noise. To make pitch information more salient against noise, it was suggested that the modulation depth of the periodicity cues should be increased (Lorenzi *et al.*, 1999; Vandali *et al.*, 2005). By expanding amplitude of the temporal envelope using power-*n* expansion, Lorenzi *et al.* (1999) reported a small but consistent performance improvement of speech recognition in noise. Vandali *et al.* (2005) evaluated the pitch ranking performances of CI users with different sound coding strategies. *F*0-related periodicity information was encoded by increasing the modulation depth of *F*0-related fluctuation across all activated electrodes. Among several experimental algorithms, the multi-channel envelope modulation (MEM) strategy was shown to lead to significantly higher scores in pitch perception than the commercially available advanced combinational encoder (ACE) strategy. For speech recognition tests in noise, the performance produced by MEM and by ACE was similar, indicating that the MEM could code segmental features properly. The MEM strategy was recently applied to speech recognition of Cantonese (Wong *et al.*, 2008). The results with nine CI recipients showed that MEM and ACE led to comparable performance. It was expected that the existing MEM algorithm could be further improved to benefit speech recognition of tone languages. In Lan *et al.* (2004) and Luo and Fu (2004b), acoustic simulations were carried out with *F*0-controlled sinusoidal or pulse-train carriers modulated by sub-band temporal envelopes. They demonstrated better Mandarin tone recognition than obtained using systems with fixed-frequency carrier or noise carrier. In Green *et al.* (2004), the complex periodicity cue extracted from the original speech was simplified to a sinusoidal or a sawtooth wave at the same *F*0, and then combined with the intact envelope cue. Improvements of pitch perception were observed in both acoustic simulations with normal-hearing subjects and listening tests with real CI users.

There have been fewer studies of Cantonese speech recognition with temporal information than of English and Mandarin. Perceptual studies of lexical tones and intonation of Cantonese speech were reported by Ma *et al.* (2005, 2006). It was shown that tonal context played an important role in Cantonese tone perception. Six-tone identification accuracy varied from 98.2% for tones in natural sentences to 78.8% for those presented in isolation. Lee *et al.* (2002) showed a significant difference in Cantonese tone identification ability between normal-hearing and CI children (92% vs 64%). Au (2003) tested Cantonese tone identification of a group of post-lingually deafened Cantonese-speaking CI users. The average accuracy was about 69%, and great individual differences were observed. These results indicated that existing CI systems are not effective in delivering tone-related information.

In Yuen *et al.* (2007), we investigated the contributions of temporal envelope and periodicity cues (TEPCs) to Cantonese tone identification in quiet. A noise-excited vocoder was used to simulate the continuous interleaved sampling (CIS) strategy (Green *et al.*, 2002). The original speech band from 0 to 4 kHz was divided into four sub-bands: 60–500, 500–1000, 1000–2000, and 2000–4000 Hz. TEPC below 500 Hz was extracted from each sub-band and used to modulate a noise signal in the same band. Different combinations of frequency bands were evaluated and compared on the task of six-tone identification with monosyllabic stimuli. TEPCs extracted from the two high-frequency bands (1–4 kHz) were found to be more important than those from the two low-frequency bands (<1 kHz). Tone identification accuracy was significantly improved by simply removing the two low-frequency bands. It was also noted that tone identification accuracy of female speech was consistently poorer than for male speech.

With isolated syllables, the task of tone identification is very similar to pitch discrimination since the candidate syllables in each test trial all have the same segmental composition and no linguistic context is provided. In the present study, we aimed at a linguistic test in which tone is used as a contrastive component for word identification. The test materials were disyllabic words. Listening tests with acoustic simulations using normal-hearing subjects were carried out with the disyllabic stimuli to investigate the contributions of TEPCs from different frequency regions to Cantonese tone recognition in quiet and in noise. Another major focus of this study is on the effectiveness of temporal periodicity enhancement for Cantonese tone recognition. We adopted the processing algorithm described in Green *et al.* (2004). A slowly-varying temporal envelope component (TEC) was extracted by full-wave rectification and low-pass filtering at 20 Hz. In each sub-band, the TEC was multiplied with a constant-amplitude sinusoidal wave that followed the *F*0 trajectory of

the original speech. This produced a modified TEPC, in which the temporal periodicity component was simplified. Tone identification tests were carried out to compare the modified CIS strategy with the standard one. We expected that (1) Cantonese tone identification accuracy would be improved by using periodicity-enhanced TEPCs, (2) periodicity enhancement of TEPCs from a high-frequency region (1–4 kHz) would be particularly effective to improve tone identification accuracy, and (3) the effect of periodicity enhancement would be more prominent for noisy speech than for clean speech. In addition, the use of disyllabic word test materials made it possible to assess the effect of the modification of temporal cues on the reception of segmental information.

II. METHODS

A. Subjects

Five male and five female college students participated in the tests. Their ages ranged from 20 to 23. All of them were native Cantonese speakers with normal hearing. Their pure-tone thresholds were better than 20 dB hearing loss (HL) at octave frequencies from 125 to 4000 Hz in both ears.

B. Speech materials

Lexical tones cannot be perceived on their own. They have to be carried by syllables that correspond to meaningful words in the language. Tone identification tests were carried out using monosyllabic words that have the same segmental structure and carry different tones (Ciocca *et al.*, 2002; Wei *et al.*, 2004; Yuen *et al.*, 2007). Due to limited segmental variation in the test materials, the linguistic role of tone was not fully reflected in these tests. In the present study, we aimed at a linguistic test in which tone is used as a contrastive component for word identification. In Chinese, disyllabic words are much more commonly used than monosyllabic words (Chin, 1998), and the former are therefore considered more appropriate for linguistic tests. Ideally, we would need many sets of disyllabic words with each set of words minimally contrasted by the tone of one of the syllables. Given the lexical constraints of Cantonese, it is difficult to find even one set of words that cover all of the six tones with the same segmental structures. Therefore, we decided to include only one pair of contrasting tones in each set of words. There are 15 contrasting tone pairs. The contrasting tone may be on either the first or the second syllable of the disyllabic words. Thus, we needed 30 sets of words.

Table I lists the 30 sets of words used in our study. There are four words in each set. The left two words, denoted by MC_A and MC_B, carry the intended contrasting tones with the same segmental properties. For example, MC_A and MC_B in set 1 are /ging1 lik6/ and /ging2 lik6/, respectively, which differ in the tones carried by the first syllables. The same tone contrast is given in set 16, but on the second syllables in the words. If each test trial involves only two candidate words, the subjects may easily realize that the test is focused on one of the syllables. To minimize the learning effect, two additional words were included in each set, as shown in the right two columns of Table I. These words are referred to as quasi-controlled tone contrasting words and denoted by QC_A and QC_B. They share the same tone contrast as that between MC_A and MC_B, but have slightly different segmental compositions. The whole set of 120 disyllabic words covers nearly 90% of the Cantonese phonemes.

TABLE I. This list of Cantonese disyllabic words. There are in total 30 sets, each containing four candidate words.

Set	MC_A	MC_B	QC_A	QC_B
1	ging1 lik6 經歷	ging2 lik6 警力	gung1 lik6 功力	geng2 lik6 頸力
2	gei1 gin2 機件	gei3 gin2 寄件	gai1 gin2 雞件	gei3 cin2 寄錢
3	jau1 mei5 優美	jau4 mei5 柔美	au1 mei5 歐美	ngau4 mei5 牛尾
4	jau1 ji6 優異	jau5 ji6 有異	jat1 ji6 一二	jau5 si6 有事
5	dak1 ji3 得意	dak6 ji3 特意	hak1 ji3 刻意	dik6 ji3 敵意
6	gu2 jan4 古人	gu3 jan4 故人	gaa2 jan4 假人	go3 jan4 個人
7	jan2 jing4 隱形	jan4 jing4 人形	jan2 cing4 隱情	jan4 cing4 人情
8	waan2 gau3 玩夠	waan5 gau3 挽救	gaan2 gau3 揀夠	laan5 gau3 懶夠
9	gau2 paai4 狗牌	gau6 paai4 舊牌	zau2 paai4 酒牌	hau6 paai4 後排
10	paa3 gou1 怕高	paa4 gou1 爬高	gwaa3 gou1 掛高	ngaa4 gou1 牙膏
11	jau3 ji4 幼兒	jau5 ji4 友誼	jau3 si4 幼時	jau5 si4 有時
12	ngoi3 gwok3 愛國	ngoi6 gwok3 外國	ngoi3 gwo3 愛過	hoi6 gwok3 害國
13	mei4 miu6 微妙	mei5 miu6 美妙	kei4 miu6 奇妙	mei5 maau6 美貌
14	jyun4 ji3 原意	jyun6 ji3 願意	cyun4 ji3 傳意	gyun6 ji3 倦意
15	lou5 jan4 老人	lou6 jan4 路人	mou5 jan4 冇人	zou6 jan4 做人
16	daai6 baan1 大班	daai6 baan2 大阪	daai6 caan1 大餐	daai6 daan2 大蛋
17	daai6 ji1 大衣	daai6 ji3 大意	daai6 si1 大師	daai6 si3 大使
18	do1 jyu1 多於	do1 jyu4 多餘	do1 syu1 多書	do1 ji4 多疑
19	daai6 jyu1 大於	daai6 jyu5 大雨	daai6 zyu1 大豬	daai6 ji5 大耳
20	jau5 jik1 有益	jau5 jik6 有翼	jau5 sik1 有色	jau5 lik6 有力
21	jat1 bun2 一本	jat1 bun3 一半	jat1 wun2 一碗	jat1 gun3 一罐
22	jyu4 leon2 魚卵	jyu4 leon4 魚鱗	jyu4 ceon2 愚蠢	jyu4 seon4 魚唇
23	daai6 jyu2 大魚	daai6 jyu5 大雨	daai6 jyun2 大丸	daai6 jau5 大有
24	gaau1 doi2 膠袋	gaau1 doi6 交待	gaau1 daai2 膠帶	gaau1 ngoi6 郊外
25	daa2 ping3 打拼	daa2 ping4 打平	daa2 ting3 打聽	daa2 sing4 打成
26	kyut3 ji3 決意	kyut3 ji5 決議	kyut3 zi3 決志	syut3 ji5 雪耳
27	daai6 bou3 大埔	daai6 bou6 大步	daai6 ngou3 大澳	daai6 dou6 大盜
28	dou6 jau4 導遊	dou6 jau5 道友	ngou6 jau4 遨遊	deoi6 jau5 隊友
29	juk6 maa4 肉麻	juk6 maa6 辱罵	juk6 ngaa4 肉芽	juk6 baa6 欲罷
30	jyun4 mei5 完美	jyun4 mei6 原味	jyun4 lei5 原理	jyun4 bei6 完備

In each test trial, one disyllabic word was presented as the stimulus. The subject was asked to identify it from the four candidate words in the same set. Table II lists all possible outcomes of the test. T and $\bar{T}$ are used to represent the cases of correct and wrong identifications of tone, respectively, while S and $\bar{S}$ refer to correct and wrong identifications of segmental structures, respectively. For example, $(T,\bar{S})$ means that tone is correctly identified but segmental identification is wrong. Although this study is focused primarily on tone recognition, the results also facilitate the investigation on the importance of TEPCs in conveying segmental information.

The speech materials were recorded from a female and a male native speaker of Hong Kong Cantonese. Recordings were made in a sound-treated booth with an Etymotic Research ER-11 microphone connected to a desktop computer with an external sound card. The recorded signals were digitized with a sampling frequency of 44 100 Hz and 16-bit resolution. To maintain a consistent pitch level of the speakers' voice, all words were recorded with the same carrier sentence: "這個詞是 _ _" (This word is_). Five repetitions

TABLE II. Possible outcomes of the word identification tests. T and $\bar{T}$ denote correct and wrong identifications of tone, respectively. S and $\bar{S}$ denote correct and wrong identifications of segmental structures, respectively.

		Recognized word			
		MC_B	MC_B	QC_A	QC_B
Presented word	MC_A	(T,S)	$(\bar{T},S)$	$(T,\bar{S})$	$(\bar{T},\bar{S})$
	MC_B	$(\bar{T},S)$	(T,S)	$(\bar{T},\bar{S})$	$(T,\bar{S})$
	QC_A	$(T,\bar{S})$	$(\bar{T},\bar{S})$	(T,S)	$(\bar{T},\bar{S})$
	QC_B	$(\bar{T},\bar{S})$	$(T,\bar{S})$	$(\bar{T},\bar{S})$	(T,S)

were recorded for each test word. Another two native speakers assessed the quality and naturalness of these recordings and selected the best one for our experiments. The test words were manually excised from the carrier sentences.

Several sets of noisy speech materials were generated by adding noise to clean speech at different signal-to-noise ratios (SNRs). For each of the speakers, a noise was generated by shaping the spectrum of white noise to follow the average spectrum of all test words spoken by that speaker. The noise signal was then added to the clean utterances at SNRs of 0, 10, and 20 dB. The signal amplitude was normalized between −1 and 1. The SNR was controlled by fixing the root-mean-square (rms) intensity level of speech signals at −25 dB and varying the noise level. Both clean and noisy speech materials were low-pass filtered at 4 kHz.

C. Speech processing and stimuli

Figure 2 depicts the standard implementation of a noise-excited vocoder. It simulates the CIS speech processing strategy, which is commonly used in CI devices (Wilson *et al.*, 1991; Wilson, 2000; Vandali *et al.*, 2005). In our implementation, the input speech went through a set of band-pass filters. The TEPCs of individual bands were extracted by full-wave rectification and low-pass filtering at the cut-off frequency of 500 Hz. They were then used to amplitude modulate noise carriers in the respective bands. The modulated noise signals were band-passed again to remove undesirable frequency components generated in the modulation process. This was done with the same band-pass filters as for the input speech. To achieve across-channel synchrony, the re-filtering was applied in the reverse direction in time such that the entire processing had zero-phase distortion (Oppenheim and Schafer, 1989; Geurts and Wouters, 2001). All band-pass filters were sixth-order elliptical infinite impulse response (IIR) filters with a roll-off of 40 dB/octave. The low-pass filters for TEPC and TEC extraction were seventh-order elliptical IIR filters with a roll-off of 50 dB/octave. In each band, the intensity of the re-filtered signal was adjusted to retain the same rms value as that of the input signal in the respective band. Thus the relative energy levels of different frequency bands were retained. Finally, acoustic stimuli were generated by combining the sub-band modulated signals.

Figure 3 illustrates the modified speech processing strategy that incorporates enhanced $F0$-related periodicity information. It differs from the standard strategy in that the TEPC used to modulate the noise carrier was not directly derived from the input speech. For each band, a slowly-varying TEC was extracted with a low cut-off frequency of 20 Hz. The TECs were multiplied with a sinusoidal wave that followed the $F0$ trajectory of the clean input speech. In other words, the original complex periodicity cues were replaced by a simplified periodicity pattern (Green *et al.*, 2004). As an example, Fig. 4 compares the original TEPC and the periodicity-enhanced one of a Cantonese syllable with tone 2 (rising tone). Panel A shows the original TEPC, superimposed by the TEC below 20 Hz. Panel B plots a constant-amplitude sinusoidal wave that follows the $F0$ contour of the syllable. Panel C gives the modified TEPC which is given by multiplying the TEC in panel A with the sinusoidal wave in panel B. The modified TEPC shows a relatively simple $F0$ modulation pattern, i.e., in each $F0$ cycle, only one primary peak is retained and all secondary peaks are removed. At the same time, the modulation depth is increased to 100%. After

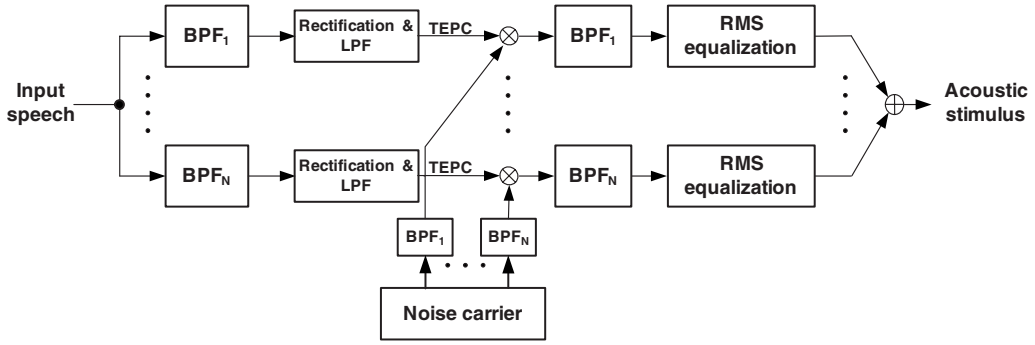


FIG. 2. Speech processing of the "standard" noise-excited vocoder. BPF and LPF indicate band-pass filter and low-pass filter, respectively.

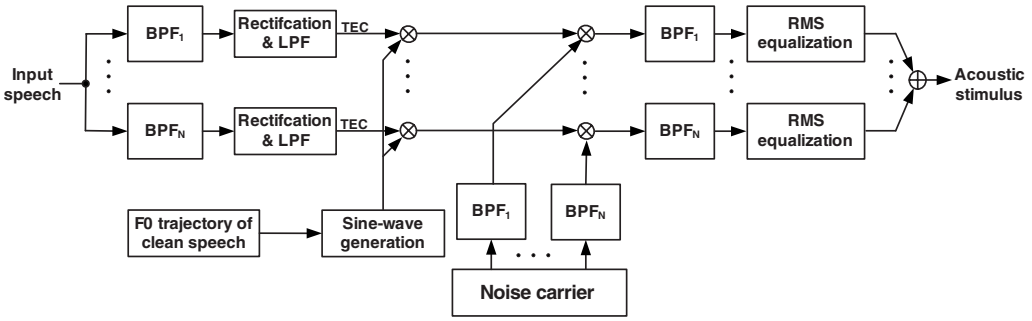


FIG. 3. The modified speech processing strategy with enhanced periodicity cues.

the modification of TEPC, the subsequent steps for acoustic stimuli generation were the same as in the standard strategy in Fig. 2.

In our study, a four-channel vocoder was used. The frequency ranges of the four bands were 60–500, 500–1000, 1000–2000, and 2000–4000 Hz, which, except for the lowest band, roughly follow the octave scale (Yuen *et al.*, 2007). This band structure is slightly different from those used in some previous studies (Dorman *et al.*, 1997; Fu *et al.*, 1998). Shannon *et al.* (1998) also used a four-channel vocoder and compared different spacings of the band-pass cutoff frequencies. It was found that the cutoff frequency values were not critical to the results of their speech recognition tests.

The *F*0 trajectories of the periodicity-enhanced test stimuli were pre-computed from full-band clean speech signals. This was done with the pitch estimation algorithm implemented in the PRAAT software.[3] The *F*0 values were manually checked and errors were corrected.

To investigate the contributions of envelope and periodicity information from different frequency regions, several sets of test stimuli were generated with different combinations of sub-bands. Based on the findings of Yuen *et al.* (2007) and Yuan *et al.* (2007), three different sub-band combinations were used:

LOW: 60–1000 Hz (only the two low-frequency bands were included and the two high-frequency bands were not used).

HIGH: 1000–4000 Hz (only the two high-frequency bands were included and the two low-frequency bands were not used).

ALL: 60–4000 Hz (all of the four bands were included).

With the three frequency regions and the two processing strategies, there were six different test conditions, as shown in Table III. The standard and modified strategies are abbreviated as STD and MOD, respectively. For clean speech materials and noisy speech at 10 dB SNR, all of the six conditions were tested. For noisy speech at 0 and 20 dB SNRs, only the HIGH conditions were tested. Including the unprocessed natural speech, there were 17 sets of test stimuli for each speaker.

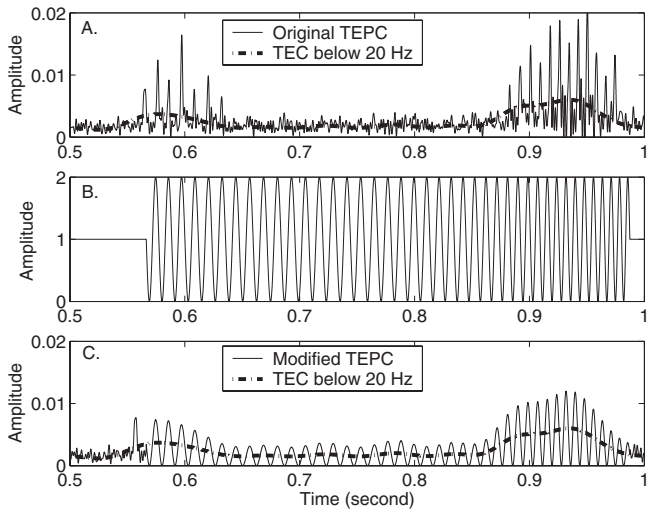


FIG. 4. Comparison of the original TEPC and the periodicity-enhanced one. The speech segment contains a Cantonese syllable with additive noise at 10 dB SNR. The TEPCs were extracted from the frequency band of 2–4 kHz.

D. Psychophysical procedures

The equipment included a laptop computer with a high-quality external audio interface (TASCAM US-122). Acoustic stimuli were presented to the subject via a Paired E.A.R. Tone 3A Insert Earphone (50 Ω). Computer software with a graphical user interface was developed to control the presentation of test stimuli and collect responses from subjects.

Each subject was required to attend two test sessions on different days. Each session involved all test stimuli from one of the speakers. The presentation order of the two speakers was balanced over all subjects. In each test session, the unprocessed clean speech was presented at the beginning so that the subjects could familiarize themselves with the process and materials. Subsequently the 16 sets of processed stimuli were presented in randomized order.

Each set of stimuli included the 120 disyllabic words, as described in Sec. II B. They were presented in randomized order without repetition. A four-alternative forced-choice procedure was adopted. Each presented stimulus was accompanied by a test screen, as shown in Fig. 5. The four choices were displayed in the form of Chinese characters, and the display positions were randomly assigned. After the presentation of a stimulus item, the subject was asked to select by mouse clicking the word that he/she had heard. The time allowed for responding to each test item was limited to 5 s. The subjects were encouraged to make a guess if they were not sure about the correct answer. If there was no response after 5 s, the test item would be regarded as "incorrectly

TABLE III. The 16 sets of processed test stimuli used in this study. Each row represents a specific processing condition for TEPC. The four columns correspond to different noise levels.

	CLEAN	20 dB	10 dB	0 dB	
all_{STD}	√		√		Original TEPCs from 60 to 4000 Hz
all_{MOD}	√		√		Modified TEPCs from 60 to 4000 Hz
low_{STD}	√		√		Original TEPCs from 60 to 1000 Hz
low_{MOD}	√		√		Modified TEPCs from 60 to 1000 Hz
$high_{STD}$	√	√	√	√	Original TEPCs from 1000 to 4000 Hz
$high_{MOD}$	√	√	√	√	Modified TEPCs from 1000 to 4000 Hz

recognized" and the system proceeded to the next item automatically. No feedback was given to the subjects.

The whole test session (for one speaker) took about 3 h. For each set, the 120 test items were presented without a break. Between two test sets, the subjects were allowed to take an optional break of about 2 min. In addition, a mandatory break of 5 min was required every 1 h.

E. Method of analysis

For each set of test stimuli in Table III, the percent correct tone identification and word identification were evaluated over all subjects. The tone score was based on all responses with correct tone identification, i.e., the recognized word carried the same tone as the presented word. This included the items (T,S) and $(T,\bar{S})$ in the confusion matrix of Table II. The chance level for tone identification is 50%. The word score was based on the answers that exactly matched the presented words, i.e., the item (T,S). The chance level for word identification is 25%.

Two primary factors that affect the test results are frequency region (LOW, HIGH, ALL) and processing strategy (STD, MOD). Statistical analysis and comparison were performed using repeated-measures analysis of variance (ANOVA) and Tukey honestly significant difference *post-hoc* tests. All scores were arcsine transformed before analysis (Studebaker, 1985). We also examined the effect of noise level and the difference between male and female voices.

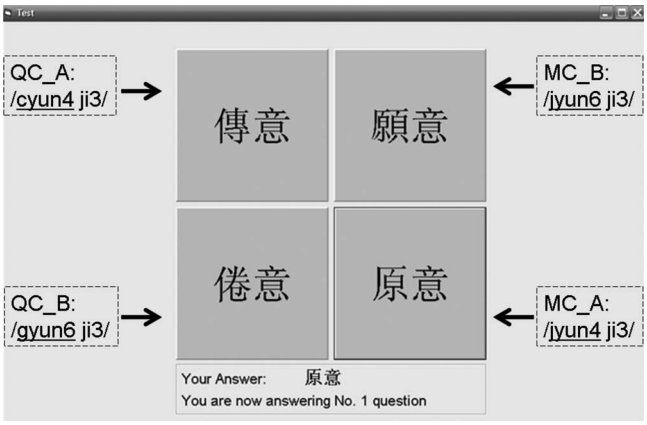


FIG. 5. The computer interface used in the listening tests. The four candidate words being displayed are from one of the 30 word sets in Table I. The notations and phonemic transcriptions in the dashed-line boxes were not shown to the subjects.

As seen in Table II, there were three different types of errors: $(T,\bar{S})$, $(\bar{T},S)$, and $(\bar{T},\bar{S})$. The percentage distributions of these errors were analyzed to reveal the effects of the above condition factors on perception of lexical tones and segmental structures.

III. RESULTS

A. Contributions from different frequency regions

1. Tone identification

Figure 6 shows the percent correct tone identification for clean speech and noisy speech at 10 dB SNR. The results for male and female voices are displayed separately. The tone scores attained with the six processing conditions in Table III are compared. The tone scores under the HIGH condition were consistently higher than those under the LOW condition. The scores under the ALL condition were between those for HIGH and LOW in most cases. The MOD strategy produced better tone identification than STD in most cases, especially for female voice. The improvement was more no-

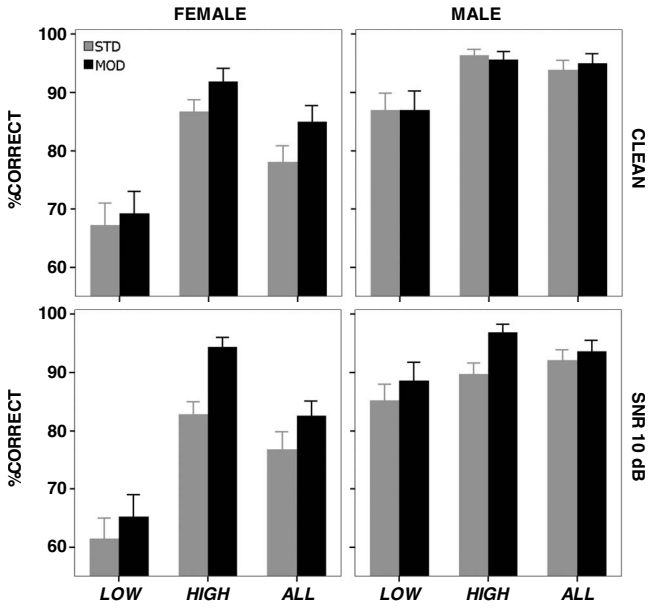


FIG. 6. Percent correct tone identification. The results for clean speech and noisy speech (10 dB SNR) and for female and male voices are shown in separate panels. Each panel shows the scores from the six processing conditions. The error bars indicate 95% confidence intervals of the mean scores over all subjects. Chance level is 50%.

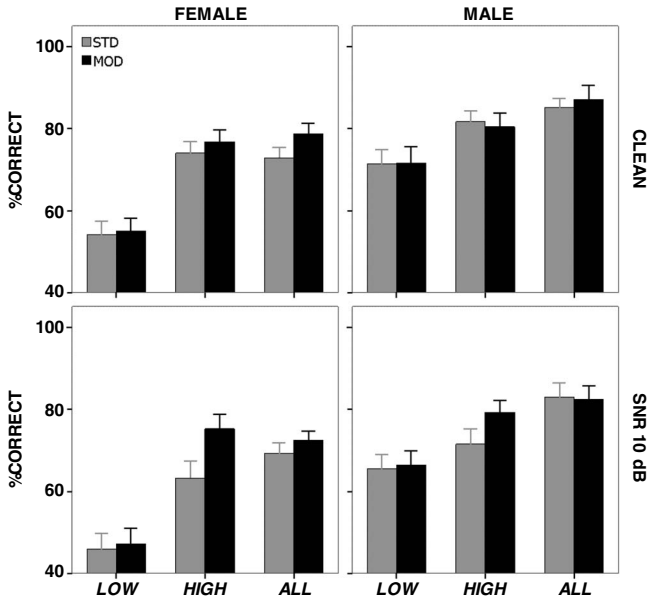


FIG. 7. Percent correct word identification. Chance level is 25%. Otherwise as Fig. 6.

ticeable for noisy speech than for clean speech. The scores for male voice were significantly higher than those for female voice for a given condition.

The results were analyzed using two-way ANOVA with the factors of processing strategy and frequency region. The ANOVA was done separately for different noise conditions (CLEAN and SNR 10 dB) and different speakers (MALE and FEMALE). The analyses revealed significant main effects of both factors in most cases ($p<0.05$). There was an exception for CLEAN-MALE, where the main effect of processing strategy was not significant [$F(1,9)=0.102$, $p=0.757$]. This might be related to the high level of performance attained with the standard strategy (85%–96% accuracy). This would limit any further improvement of the scores. A significant main effect was found for the two-way interaction between processing strategy and frequency region, reflecting that the effect of MOD was larger for HIGH than for LOW and ALL, as seen from Fig. 6.

Post-hoc tests confirmed that, for female voice, the scores differed across frequency regions, with HIGH giving the highest scores and LOW the lowest ($p<0.05$). The same trend was observed for male voice but the difference was not significant ($p>0.05$). *Post-hoc* comparison also showed that, for the HIGH condition, the tone score with MOD was significantly higher than that with STD. For clean male speech, the modified processing strategy did not improve tone identification performance. This again might be due to the high performance attained with the standard strategy ($\sim 95\%$).

2. Word identification

Figure 7 shows the percent correct word identification under different test conditions. Word identification under the HIGH condition was consistently higher than under the LOW condition. In contrast to the tone identification results, ALL gave higher scores than HIGH in most cases. The MOD strategy improved word identification accuracy over STD, especially for female voice and noisy speech. The word scores for male voice were higher than those for female voice for a given condition.

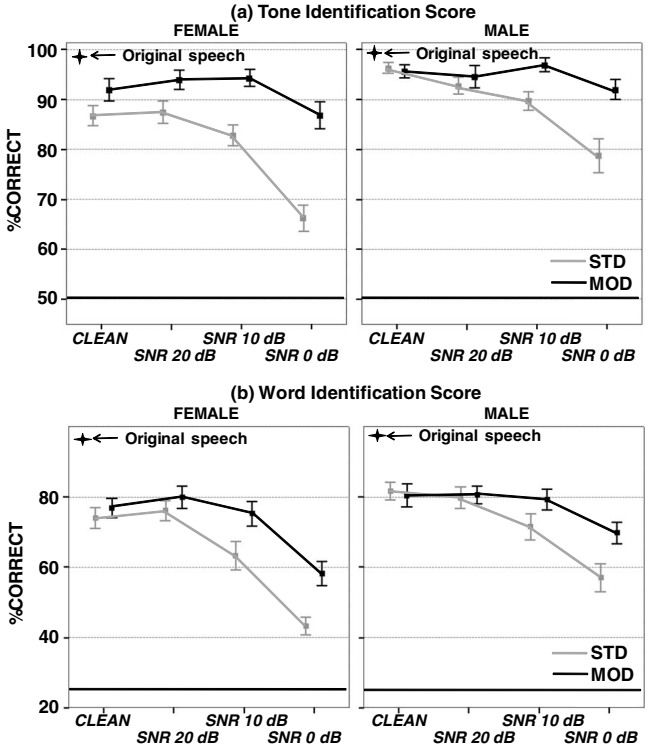


FIG. 8. Percent correct (a) tone identification and (b) word identification as a function of SNR. The error bars indicate the 95% confidence intervals of the mean scores over all subjects.

A two-way ANOVA was carried out to analyze the effect of frequency region and processing strategy. Significant main effects of both factors were found for different noise levels and different speakers ($p<0.05$). Similar to the tone identification results, the main effect of processing strategy for CLEAN-MALE was not significant [$F(1,9)=0.415$, $p=0.535$]. The interaction between processing strategy and frequency region for SNR 10 dB female was significant. *Post-hoc* comparison revealed no significant difference between STD and MOD except for SNR 10 dB female ($p<0.05$), indicating that the periodicity enhancement method may not be beneficial to word recognition. The word scores for HIGH and ALL were significantly higher than those for LOW in most cases ($p<0.05$). For both processing strategies, there was no significant difference between HIGH and ALL ($p>0.05$), except for SNR 10 dB male ($p=0.030$). Overall, the periodicity-enhanced processing method did not show any negative effect on word recognition but did show a positive effect on tone identification, especially for noisy speech.

B. Effect of noise

Figure 8 shows the test results as a function of SNR: clean, 20, 10, and 0 dB under the HIGH condition. The scores for unprocessed clean speech are also given for reference. In general, both tone identification and word identification declined as the SNR decreased. MOD led to a consis-

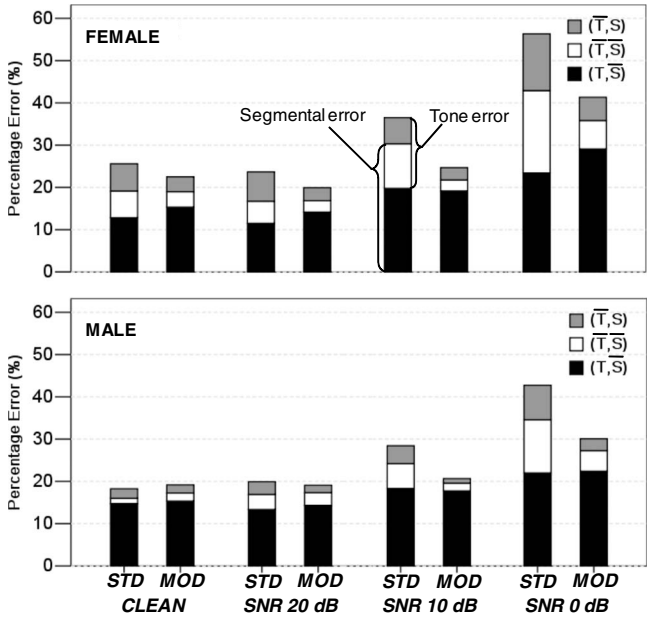


FIG. 9. Comparison of percentage error distributions between the standard and the modified processing strategies at different SNRs. The percentage error is computed as the ratio of the number of the respective type of errors to the total number of test items for all subjects.

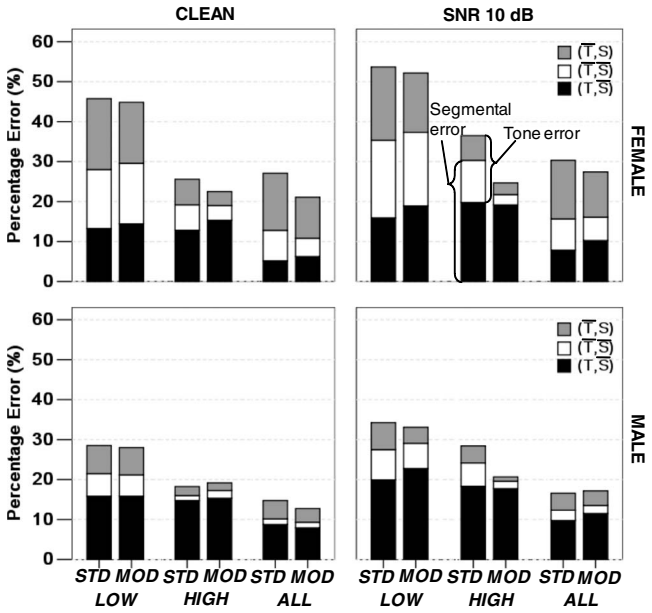


FIG. 10. Comparison of percentage error distributions between the standard and the modified processing strategies under different frequency region conditions.

tently higher performance than STD. For tone identification, MOD led to an accuracy of about 90% across all noise conditions.

A two-way ANOVA was used to analyze the effects of noise level and processing strategy. Both factors and their interaction were shown to have significant effects on tone identification and word identification ($p<0.05$). The interaction between processing strategy and noise level reflects the result that the performance difference between STD and MOD depends on the noise level. The effectiveness of MOD was more noticeable at low SNR than at high SNR. *Post-hoc* tests indicated that the tone scores and word scores for MOD were significantly higher than for STD for SNR 10 dB and SNR 0 dB ($p<0.05$). No significant difference was found between the scores for CLEAN and SNR 20 dB ($p>0.05$). With MOD, there was no significant difference among CLEAN, SNR 20 dB, and SNR 10 dB ($p>0.05$).

C. Analysis of error distributions

Figure 8 shows that, under the HIGH condition, the use of periodicity-enhanced TEPCs leads to improvement of both tone identification and word identification accuracies, especially at SNR of 10 dB or below. A word identification error may be caused by tone error, segmental error, or both. Figure 9 shows the percentage distributions of different types of errors in the test results. It is noted that tone errors, which include $(\bar{T},S)$ and $(\bar{T},\bar{S})$, were substantially reduced by the periodicity-enhancement processing strategy, especially at low SNR. Meanwhile, the number of the $(T,\bar{S})$ errors increased because the improved tone identification made some of the $(\bar{T},\bar{S})$ errors become $(T,\bar{S})$. The total number of segmental errors, which include $(T,\bar{S})$ and $(\bar{T},\bar{S})$, decreased at low SNR and remained intact at high SNR. In other words, although the modified processing strategy removes the details of periodicity cues, it does not seem to affect the delivery of segmental information.

Figure 10 compares the distributions of different types of errors among the LOW, HIGH, and ALL conditions. The number of the $(T,\bar{S})$ errors was similar between LOW and HIGH, but considerably smaller for ALL. This indicates that TEPCs from all frequency bands contain useful information for identifying segmental cues. On the other hand, there were much fewer $(\bar{T},S)$ errors for HIGH than for LOW and ALL. The use of TEPCs from the low-frequency region seems to negatively affect tone recognition. The number of the $(\bar{T},\bar{S})$ errors for LOW was much greater than for HIGH and ALL. Apparently, the superiority of HIGH over LOW in word identification was due to the improved tone identification.

IV. GENERAL DISCUSSION AND CONCLUSION

A. Effectiveness of periodicity enhancement

Our experimental results showed that Cantonese tone identification in quiet could reach a high performance level without using any fine structure cues. When speech was masked by noise, performance deteriorated drastically. The presence of noise leads to many spurious temporal peaks, which contaminate the representation of periodicity in the extracted TEPCs and hence adversely affect pitch perception.

We expected that using a better representation of temporal periodicity would lead to improvement on Cantonese tone recognition performance. We used a speech processing strategy modified from the conventional multi-channel noise-excited vocoder. As shown in Fig. 4, temporal periodicity was made more salient in two different ways. First, a simple periodicity pattern was used to replace the complex periodicity cues in the original speech. Second, the periodicity-related modulation depth was increased. This method of pe-

riodicity enhancement was first proposed and explored by Green *et al.* (2004). The test stimuli were synthesized English diphthong segments with gliding *F*0. The subjects were asked to distinguish between "rising" and "falling" pitch contours. By using *F*0-related sinusoidal or sawtooth waveforms to replace complex periodicity cues in speech signals, pitch discrimination capabilities of both normal-hearing subjects and CI recipients were improved noticeably. Laneau *et al.* (2006b) designed a new sound processing strategy to improve pitch perception. In this strategy, slowly-varying temporal envelopes in individual frequency bands were used to modulate sinusoidal waves that carried the *F*0 contour of the input signals. Temporal envelope pitch cues were maximized with 100% modulation depth and across-channel synchronization. The test results showed that the performance of CI recipients on music and pitch perception could be improved.

In our study, the benefit of using simplified periodicity cues was evaluated in a linguistic task for a specific language. The experimental results not only confirmed that pitch perception had been improved but also showed that speech recognition could be improved in both quiet and noisy conditions. The tone scores attained using the modified strategy at SNRs of 20 and 10 dB were very close to that for clean speech. Even for very noisy speech with a 0 dB SNR, the tone score was maintained at a very high level. In human sound perception, the peaks of a temporal envelope stimulus are translated into neural impulses. The intervals between successive impulses correspond approximately to the period of the sound or its integer multiples (Moore, 2007). The complex periodicity cues, especially those extracted from noisy speech, may contain many pitch-irrelevant fluctuations such that the true *F*0 cannot be clearly represented in the neural firing pattern. It is believed that a simplified periodicity pattern provides a better representation of *F*0 in the neural firing pattern (Green *et al.*, 2004).

Temporal periodicity enhancement by increasing the modulation depth in TEPC has been widely studied (McKay *et al.*, 1995; Lorenzi *et al.*, 1999; Geurts and Wouters, 2001; Vandali *et al.*, 2005). McKay *et al.* (1995) found that, if the modulation depth was too small, the perceived pitch would correspond to the frequency of the pulse-train carrier instead of the *F*0-related modulating frequency. Geurts and Wouters (2001) used sinusoidally amplitude-modulated pulse trains. Their results showed that pitch discrimination performance of CI users degraded when modulation depth was decreased. In Lorenzi *et al.* (1999), noticeable performance improvement on speech recognition in noise was achieved with power-2 expansion of TEPC. In the present study, the modulation depth of the periodicity-enhanced TEPCs was set to be 100%. The test results showed the effectiveness of increasing modulation depth for tone recognition, which is consistent with the observations of other studies on pitch perception.

B. Contributions from different frequency regions

We have shown that TEPCs from different frequency regions do not contribute equally to Cantonese tone identification. The high-frequency region (1–4 kHz) is more important than the low-frequency region (<1 kHz). The same conclusion was reached in our previous study (Yuen *et al.*, 2007), where monosyllabic test stimuli were used. Luo and Fu (2004a) reported that periodicity cues from high-frequency bands were more important than those from low-frequency bands for Mandarin tone recognition. This was explained in two ways: (1) characteristics of periodic waveforms in different frequency channels and (2) sensitivities of different cochlear positions of normal-hearing listeners to periodic fluctuations. The periodic waveform in the highest frequency band is the simplest, which might be helpful for detecting pitch and its changes. For normal-hearing listeners, the ability to use amplitude modulation cues for pitch detection is much better in the high-frequency region than in the low-frequency region (Moore, 2007). On the other hand, in the noise-excited vocoder, the TEPC from each frequency band was used to modulate a noise carrier in the same band. For a low-frequency channel, the carrier frequency is very close to the frequency of the TEPC, which is below 500 Hz. The modulation process may generate undesirable components that interfere with pitch extraction. This makes the TEPCs from the low-frequency region less useful in delivering pitch information. McKay *et al.* (1994) suggested that the carrier frequency in a CI should be at least 4 times the *F*0. Otherwise pitch perception would be affected by the undersampling of the F0-related periodicity.

C. Effect of *F*0 range

Our experimental results showed that tone identification performance was consistently much better for male voice than for female voice. This indicates that the relatively high *F*0 modulation frequency in the TEPC of female speech might not be represented as well as the low *F*0 in male speech. Qin and Oxenham (2005) reported that pitch perception with TEPC was poorer with high *F*0 modulation than with low *F*0 modulations. Fu and Zeng (2000) investigated temporal envelope cues for Mandarin tone recognition with single-vowel syllable stimuli. They found that the tone recognition score for male voice was 10% higher than that for female voice.

In contrast, Xu *et al.* (2002) found no significant difference between Mandarin tone recognition performance for male and female voices. In their study, a long training session (4–5 h) was conducted to familiarize each subject with the test materials. There was a possibility that the subjects learnt to utilize acoustic cues other than *F*0 to distinguish the tones. Indeed, vowel duration and amplitude contour were found to be useful cues for Mandarin tone identification (Lin, 1988; Whalen and Xu, 1992; Fu and Shannon, 2000). For Cantonese, tone identification replies mostly on *F*0, while duration and amplitude contour are not considered to be reliable discriminative features (Qian *et al.*, 2007).

D. Practical implications

Ciocca *et al.* (2002) investigated Cantonese tone perception of a group of early-deafened CI users. They found that the children had great difficulty in extracting pitch information from temporal cues. There is a strong need to improve

existing CI speech processing strategies for better pitch perception and tone recognition. The results of our study suggested that Cantonese tone recognition in noise could be improved by (1) adjusting the balance between temporal cues from high- and low-frequency regions and (2) simplifying the complex periodicity cues of input speech. These observations can be considered in designing CI processing strategies for Cantonese-speaking users. Although acoustic simulations with normal-hearing subjects may not directly reflect the performances of real CI users (Laneau *et al.*, 2006a), they allow assessment of the acoustic information that is available in the processed signal.

For the implementation of this method, one major practical problem is real-time *F*0 estimation in realistic acoustic environments. Although there have been many attempts to address the problem of robust *F*0 estimation, most existing algorithms can handle only a limited range of noise conditions (Kinjo and Funaki, 2006; Krini and Schmidt, 2007). In our study, the *F*0 trajectory of the speech signal was extracted from the original clean speech. Therefore the modified TEPCs optimally represent the *F*0 information in the original speech. In practice, it is difficult to perform accurate real-time estimation of *F*0 with the available computational power in CI processors, especially for noisy speech. However, it is noted that very precise *F*0 contours may not be needed for tone perception. In Li and Lee (2007), it was shown that Cantonese tone contours could be approximated by simple linear movements without producing noticeable perceptual difference. Thus the *F*0 estimation problem may be alleviated by using coarsely predicted tone contours.

ACKNOWLEDGMENTS

This research was partially supported by the Earmarked Research Grants (Contract Nos. CUHK 413405 and CUHK 413507) from the Hong Kong Research Grants Council. The authors would like to thank Brian Moore and the two anonymous reviewers for their valuable suggestions and constructive comments.

[1]In this article, the Jyut Ping system is used for transcribing Cantonese syllables and tones. Jyut Ping was devised by the Linguistic Society of Hong Kong (LSHK, 1997).

[2]Traditionally, Cantonese is said to have nine tones (Hashimoto, 1972). Three of them are the so-called "entering tones," which have contrastively shorter duration than the "non-entering" tones. If duration difference is not considered, each entering tone can be combined with one of the non-entering tones that has a similar pitch pattern (Bauer, 1997; Cutler and Chen, 1997).

[3]PRAAT 5.0.20, Copyright © 1992–2008 by P. Boersma and D. Weenink, www.praat.org (Last viewed January, 2009).

Au, D. K. K. (**2003**). "Effects of stimulation rates on Cantonese lexical tone perception by cochlear implant users in Hong Kong," Clin. Otolaryngol. **28**, 533–538.

Bauer, R. S. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, New York), Vol. **103**, Chap. 2, p. 109.

Chin, A. C. (**1998**). *Quantitative and Computational Studies on the Chinese Language* (Language Information Sciences Research Centre, City University of Hong Kong, Hong Kong).

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). "The perception of Cantonese lexical tones by early-deafened cochlear implantees," J. Acoust. Soc. Am. **111**, 2250–2256.

Cutler, A., and Chen, H. C. (**1997**). "Lexical tone in Cantonese spoken-word processing," Percept. Psychophys. **59**, 165–179.

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.

Fu, Q.-J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition in cochlear implants," J. Acoust. Soc. Am. **107**, 1889–1900.

Fu, Q.-J., and Zeng, F.-G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pacific J. Speech, Lang. Hearing **5**, 45–57.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

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. (**2002**). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," J. Acoust. Soc. Am. **112**, 2155–2164.

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.

Hashimoto, O.-K. Y. (**1972**). *Phonology of Cantonese* (Cambridge University Press, Cambridge).

Kinjo, T., and Funaki, K. (**2006**). "F0 estimation of noisy speech based on complex speech analysis," in Proceedings of the Digital Signal Processing Workshop, 12th Signal Processing Education Workshop, pp. 434–437.

Kong, Y.-Y., and Zeng, F.-G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**, 2830–2840.

Krini, M., and Schmidt, G. (**2007**). "Spectral refinement and its application to fundamental frequency estimation," in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 251–254.

Lan, N., Nie, K.-B., Gao, S.-K., and Zeng, F.-G. (**2004**). "A novel speech processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.

Laneau, J., Moonen, M., and Wouters, J. (**2006a**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.

Laneau, J., Wouters, J., and Moonen, M. (**2006b**). "Improved music perception with explicit pitch coding in cochlear implants," Audiol. Neuro-Otol. **11**, 38–52.

Lee, K. Y. S., Chiu, S. N., and van Hasselt, C. A. (**2002**). "Tone perception ability of Cantonese-speaking children," Lang Speech **45**, 387–406.

Li, Y. J., and Lee, T. (**2007**). "Perceptual equivalence of approximated Cantonese tone contours," in Proceedings of the ISCA Interspeech, pp. 2677–2680.

Lin, M. C. (**1988**). "The acoustic characteristics and perceptual cues of tones in standard Chinese," Chinese Yuwen **204**, 182–193.

Lorenzi, C., Berthommier, F., Apoux, F., and Bacri, N. (**1999**). "Effects of envelope expansion on speech recognition," Hear. Res. **136**, 131–138.

LSHK (**1997**). *Hong Kong Jyut Ping Characters Table* (Linguistic Society of Hong Kong Press, Hong Kong).

Luo, X., and Fu, Q.-J. (**2004a**). "Contributions of periodicity fluctuation cues in individual frequency channels to Chinese speech recognition," in Proceedings of the International Symposium on Chinese Spoken Language Processing, pp. 133–136.

Luo, X., and Fu, Q.-J. (**2004b**). "Importance of pitch and periodicity to Chinese-speaking cochlear implant patients," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. **4**, pp. 1–4.

Ma, J. K.-Y., Ciocca, V., and Whitehill, T. (**2005**). "Contextual effect on perception of lexical tones in Cantonese," in Proceedings of the Eurospeech, pp. 401–404.

Ma, J. K.-Y., Ciocca, V., and Whitehill, T. (**2006**). "Effect of intonation on Cantonese lexical tones," J. Acoust. Soc. Am. **102**, 3978–3987.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). "Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: The effect of modulation depth," J. Acoust. Soc. Am. **97**, 1777–1785.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd ed. (Wiley, Chichester).

Moore, B. C. J., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.
Oppenheim, A., and Schafer, R. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ.).
Pike, K. L. (**1948**). *Tone Languages* (University of Michigan Press, Ann Harbor, MI).
Qian, Y., Lee, T., and Soong, F. K. (**2007**). "Tone recognition in continuous Cantonese speech using supratone models," J. Acoust. Soc. Am. **121**, 2936–2945.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., and McDermott, H. J. (**2005**). "Pitch ranking ability of CI recipients: A comparison of sound processing strategies," J. Acoust. Soc. Am. **117**, 3126–3138.
von Bekesy, G. (**1963**). "Hearing theories and complex sounds," J. Acoust. Soc. Am. **35**, 588–601.
Wei, C. G., Cao, K. L., and Zeng, F.-G. (**2004**). "Mandarin tone recognition in cochlear-implant subjects," Hear. Res. **197**, 87–95.
Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.
Wilson, B. S. (**2000**). "Strategies for representing speech information with cochlear implants," in *Cochlear Implants: Principles and Practices*, edited by J. K. Niparko (Lippincott Williams and Wilkins, Philadelphia), Chap. 7, pp. 129–170.
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.
Wong, L. L. N., Vandali, A. E., Ciocca, V., Luk, B., Ip, V. W. K., Murray, B., Yu, H. C., and Chung, I. (**2008**). "New cochlear implant coding strategy for tonal language speakers," Int. J. Audiol. **47**, 337–347.
Xu, Y. (**1997**). "Contextual tonal variations in Mandarin," J. Phonetics **25**, 61–83.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.
Yuan, M., Lee, T., Yuen, K. C. P., Soli, S. D., Tong, M. C. F., and van Hasselt, C. A. (**2007**). "Band-specific temporal periodicity enhancement for Cantonese tone perception with noise-excited vocoder," in Proceedings of the Annual International Conference of IEEE-EMBC, pp. 694–697.
Yuen, K. C. P., Yuan, M., Lee, T., Soli, S. D., Tong, M. C. F., and van Hasselt, C. A. (**2007**). "Frequency-specific temporal envelope and periodicity components for lexical tone identification in Cantonese," Ear Hear. **28**, 107s–113s.

Factors affecting masking release in cochlear-implant vocoded speech

Ning Li and Philipos C. Loizou[a)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

(Received 27 August 2008; revised 3 February 2009; accepted 22 April 2009)

Cochlear-implant (CI) listeners generally perform better when listening to speech in steady-state noise than in fluctuating maskers, and the reasons for that are unclear. The present study presents a new hypothesis for the observed absence of release from masking. When listening to speech in fluctuating maskers (e.g., competing talkers), CI users cannot fuse the pieces of the message over temporal gaps because they are not able to perceive reliably the acoustic landmarks introduced by obstruent consonants (e.g., stops). These landmarks are evident in spectral discontinuities associated with consonant closures and releases and are posited to aid listeners determine word/syllable boundaries. To test this hypothesis, normal-hearing (NH) listeners were presented with vocoded (6–22 channels) sentences containing clean obstruent segments, but corrupted (by steady noise or fluctuating maskers) sonorant segments (e.g., vowels). Results indicated that NH listeners performed better with fluctuating maskers than with steady noise even when speech was vocoded into six channels. This outcome suggests that having access to the acoustic landmarks provided by the obstruent consonants enables listeners to integrate effectively pieces of the message glimpsed over temporal gaps into one coherent speech stream.
 [DOI: 10.1121/1.3133702]

PACS number(s): 43.66.Ts, 43.71.Ky [JCM] Pages: 338–346

I. INTRODUCTION

It is generally accepted that normal-hearing (NH) listeners are able to recognize speech in modulated or fluctuating maskers with higher accuracy than in continuous (steady-state) noise (e.g., Festen and Plomp, 1990). The benefit received when listening to speech in fluctuating maskers compared to steady maskers is often called "release of masking." This benefit can be quite substantial and can range from less than 5 to near 10 dB (e.g., Festen and Plomp, 1990; Peters *et al.*, 1998), depending on the temporal/spectral characteristics of the masker. Several factors contribute to the masking release (see review in Assmann and Summerfield, 2004) including segregation of the target on the basis of F0 differences (between the target and masker) and the ability to glimpse the target during the portions of the mixture in which the signal-to-noise ratio (SNR) is favorable, i.e., during periods in which the temporal envelope of the masker reaches a dip.

Unlike NH listeners who benefit greatly from "listening in the dips," cochlear-implant (CI) listeners are not able to receive masking release when listening to speech in fluctuating maskers. This was confirmed in studies involving CI users (Nelson *et al.*, 2003; Fu and Nogaki, 2004; Nelson and Jin, 2004; Stickney *et al.*, 2004; Cullington and Zeng, 2008) and in studies involving NH listeners listening to CI simulations, i.e., vocoded speech (Qin and Oxenham, 2003, 2005; Stickney *et al.*, 2004). Stickney *et al.* (2004) assessed speech recognition by CI users at SNR levels ranging from 0 to 20 dB using as maskers single talkers (male or female) and steady-state noise. Results showed no release from masking. In fact, performance with single talker maskers was lower than performance with steady-state noise.

The reasons for the lack of masking release are not clear, and several hypotheses have been proposed. One hypothesis suggests that CI users are not able to effectively use F0 cues to segregate the target even when a large number of channels is available (Stickney *et al.*, 2007; Qin and Oxenham, 2003, 2005). Qin and Oxenham (2005) demonstrated that NH listeners are unable to benefit from F0 differences between competing vowels in a concurrent-vowel paradigm despite the good F0 difference limens (<1 semitone) obtained with 8- and 24-channel vocoder processings. A similar outcome was noted by Stickney *et al.* (2007) with CI users listening to target and competing sentences with an F0 separation ranging from 0 to 15 semitones. Others hypothesized (Nelson *et al.*, 2003) that the fluctuating maskers may cause modulation interference particularly when the signal spectral representation is poor, as is the case with current implant systems. Nelson *et al.* (2003) tested CI users with sentences embedded in modulated (gated) maskers with modulation rates varying from 1 to 32 Hz. No release of masking was observed for rates of 2–8 Hz. In fact, lower performance was observed at the syllabic rates (2–4-Hz gating) and that was attributed to the possibility that the modulated maskers were actually a distraction or interference rather than a benefit. Most CI users received benefit with the 1-Hz modulation rate, which assumes unrealistically long (500-ms) silent intervals of opportunity to glimpse the target. As argued in Nelson *et al.* (2003), the lack of masking release could not have been due to lack of audibility in the "dips" since in their study the signal level exceeded the masker level by 8 and 16 dB. Stickney *et al.* (2004) observed greater masking with

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

single-talker than noise maskers, and they attributed that to a stronger influence of informational masking compared to energetic masking. They argued that even though the single-talker maskers are spectrally degraded, it is possible that they retain some phonetic properties of natural speech which may be easily confused with those of the target.

Overall, the outcomes from the above studies do not paint a clear, or overly convincing, picture as to why CI users do not receive release from masking. In the present paper, we investigate an alternative, and new, hypothesis that explains prior findings. As argued in most of the above studies, it is very likely that CI users are not able to integrate the pieces of the message which are glimpsed across temporal gaps to a single auditory image. We then ask the following question: Which pieces (or phonetic segments) in the noisy speech stream are difficult to perceive due to noise masking and/or perhaps CI processing? Put differently, what characteristics or features of the speech signal are more susceptible to noise? As shown by Munson and Nelson (2005) not all phonetic features/segments are affected the same way in noise. Sounds, for instance, with rapidly changing spectral patterns were found to be most vulnerable to misperception in noise by CI users (Munson and Nelson, 2005). From the NH literature we know that the obstruent consonants (stops, fricatives, and affricates) are more susceptible to noise masking than the more-intense sonorant sounds (vowels, semivowels, and nasals). Phatak and Allen (2007), for instance, showed that aside from a small subset of consonants, the vowel-to-consonant recognition ratio is well above unity for a large range of SNR levels (−20–0 dB), suggesting that vowels are easier to recognize than consonants in speech-weighted noise. The study by Parikh and Loizou (2005) showed that the information contained in the first two vowel formants is preserved to some degree even at low SNR levels. In contrast, both the spectral tilt and burst frequency of stop consonants, which are known to convey place of articulation information (e.g., Blumstein and Stevens, 1979), were significantly altered by noise.

If we accept that the obstruent consonants are heavily masked by noise, the question arises as to why better perception of the obstruent consonants (occurring roughly 33% of the time, Mines *et al.*, 1978) would help listeners identify more words in the noisy speech stream. For one, the obstruent consonants are characterized by spectral discontinuities, such as those introduced by the closure and release of stop consonants. These discontinuities manifest themselves as acoustic landmarks, which are posited to be crucial in the segmentation stage of lexical-access models (Stevens, 2002). There is evidence (see Li and Loizou, 2008) that suggests that NH listeners can receive substantial improvements in speech recognition in noise when presented with sentences containing clean obstruent consonants but noise-corrupted voiced sounds, e.g., vowels. The study by Li and Loizou (2008) focused on assessing the contribution of acoustic landmarks to the recognition of speech corrupted by steady-state maskers rather than fluctuating maskers. The present study extends the scope of the study by Li and Loizou (2008) and examines the effect of fluctuating maskers on the perception of vocoded speech. In the context of CIs, the proposed study tests the hypothesis that listeners cannot integrate the pieces of the message across temporal gaps because they cannot perceive reliably the obstruent consonants and associated acoustic landmarks. Restoring the obstruent consonants (and associated landmarks) ought to aid listeners identify more words and allow them to receive release from masking. To test this hypothesis, we present to NH listeners vocoded noisy sentences containing clean obstruent consonants but corrupted sonorant sounds. In doing so, we will assess the contribution of information carried by obstruent consonants (and associated landmarks) to masking release. Vocoded speech with varying spectral resolution and NH listeners will be used in the present paper to study masking release in the absence of confounding factors (e.g., electrode insertion depth) associated with CI users.

II. EXPERIMENT: CONTRIBUTION OF OBSTRUENT CONSONANTS TO RECOGNITION OF VOCODED SPEECH IN NOISE

A. Methods

1. Subjects

Seven NH listeners participated in this experiment. All subjects were native speakers of American English and were paid for their participation. Subject's age ranged from 18 to 40 yrs, with the majority being graduate students at the University of Texas at Dallas.

2. Stimuli

The speech material consisted of sentences taken from the IEEE database (IEEE, 1969). All sentences were produced by a male speaker. The sentences were recorded in a sound-proof booth (Acoustic Systems, Inc.) in our laboratory at a 25-kHz sampling rate. Details about the recording setup and copies of the recordings are available in Loizou (2007). Two types of maskers were used. The first was continuous (steady-state) noise, which had the same long-term spectrum as the test sentences in the IEEE corpus. The second masker was a two-talker competing speech (female) recorded in our laboratory. Two long sentences, produced by a female talker, were used from the IEEE database. This was done to ensure that the target signal was always shorter (in duration) than the masker.

The IEEE sentences were manually segmented into two broad phonetic classes: (a) the obstruent consonants which included the stops, fricatives, and affricates and (b) the sonorant sounds which included vowels, semivowels, and nasals. The segmentation was done in two steps. In the first step, a highly accurate F0 detector, taken from the STRAIGHT algorithm (Kawahara *et al.*, 1999), was used to provide the initial classification of voiced and unvoiced segments. The stop closures were classified as belonging to the unvoiced segments. The F0 detection algorithm was applied every 1 ms to the stimuli using a high-resolution fast Fourier transform to provide for accurate temporal resolution of voiced/unvoiced boundaries. Segments with non-zero F0 values are initially classified as voiced and segments with zero F0 value (as determined by the STRAIGHT algorithm) are classified as unvoiced. In the second step, the voiced and

TABLE I. Cutoff frequencies of the bandpass filters used in the vocoder simulations.

	6 channels		12 channels		22 channels	
Channel	Low (kHz)	High (kHz)	Low (kHz)	High (kHz)	Low (kHz)	High (kHz)
1	0.300	0.487	0.300	0.382	0.300	0.390
2	0.487	0.791	0.382	0.487	0.390	0.489
3	0.791	1.284	0.487	0.620	0.489	0.595
4	1.284	2.085	0.620	0.791	0.595	0.711
5	2.085	3.387	0.791	1.008	0.711	0.835
6	3.387	5.500	1.008	1.284	0.835	0.970
7			1.284	1.636	0.970	1.117
8			1.636	2.085	1.117	1.275
9			2.085	2.658	1.275	1.446
10			2.658	3.387	1.446	1.631
11			3.387	4.316	1.631	1.832
12			4.316	5.500	1.832	2.049
13					2.049	2.228
14					2.228	2.539
15					2.539	2.815
16					2.815	3.114
17					3.114	3.437
18					3.437	3.787
19					3.787	4.165
20					4.165	4.575
21					4.575	5.019
22					5.019	5.500

unvoiced decisions are inspected for errors, and the detected errors are manually corrected. Segments belonging to voiced stops with pre-voicing (e.g., /b/) as well as segments belonging to voiced fricatives (e.g., /z/) are classified as obstruent consonants. Waveform and time-aligned spectrograms were used to refine the voiced/unvoiced boundaries. Criteria for identifying a segment belonging to voiced sounds (sonorant sounds) included the presence of voicing, a clear formant pattern, and absence of signs of a vocal-tract constriction. For the special boundary that separates a prevocalic stop from a following semivowel (as in *truck*), we adopted the rule used in the phonetic segmentation of the TIMIT corpus (Seneff and Zue, 1988). More precisely, the unvoiced portion of the following semivowel or vowel was absorbed in the stop release and was thus classified as an obstruent consonant. The two-class segmentation of all IEEE sentences was saved in text files (same format as the TIMIT phn files) and is available from a CD ROM in Loizou (2007).

3. Signal processing

Signals were first processed through a pre-emphasis filter (2000-Hz cutoff), with a 3-dB/octave rolloff, and then bandpass filtered into 6, 12, or 22 channels using sixth-order Butterworth filters. Logarithmic filter spacing was used to allocate the 6 and 12 channels across a 300–5500-Hz bandwidth, and mel filter spacing was used for the 22-channel condition (see Table I). The envelope of the signal was extracted by full-wave rectification and low-pass filtering (second-order Butterworth) with a 400-Hz cutoff frequency. The envelopes in each channel were modulated by white noise and re-filtered with the same analysis filters. The filtered waveforms of each band were finally summed, and the level of the synthesized speech segment was adjusted to have the same rms value as the original (clean)speech waveform of each band.

The speech stimuli were vocoded using the above algorithm in two different conditions. In the first control condition, the corrupted speech stimuli were left unaltered. That is, the obstruent consonants remained corrupted by the maskers. In the second condition, the speech stimuli contained clean obstruent segments but corrupted sonorant segments. The same level normalization factor was applied to the synthesized waveforms in both conditions. Figure 1 shows an example sentence embedded in steady noise (5-dB SNR) and processed in the two conditions. The top panel shows the spectrogram of the corrupted sentence vocoded into six channels, and the bottom panel shows the same sentence containing clean (vocoded) obstruent segments but corrupted (vocoded) sonorant segments. As shown in Fig. 1 (top panel), two of the fricative segments (at $t=0.6$ s and $t=1.9$ s) are vaguely visible in the corrupted sentence; however, the majority of the stop consonant segments are not easily discernible. The closure and release of the stop /p/, for instance (see bottom panel at $t=2.3-2.5$ s) is completely masked by noise.

4. Procedure

The experiments were performed in a sound-proof room (Acoustic Systems, Inc.) using a PC connected to a Tucker-Davis system 3. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones at a comfortable listening level. Prior to the test,

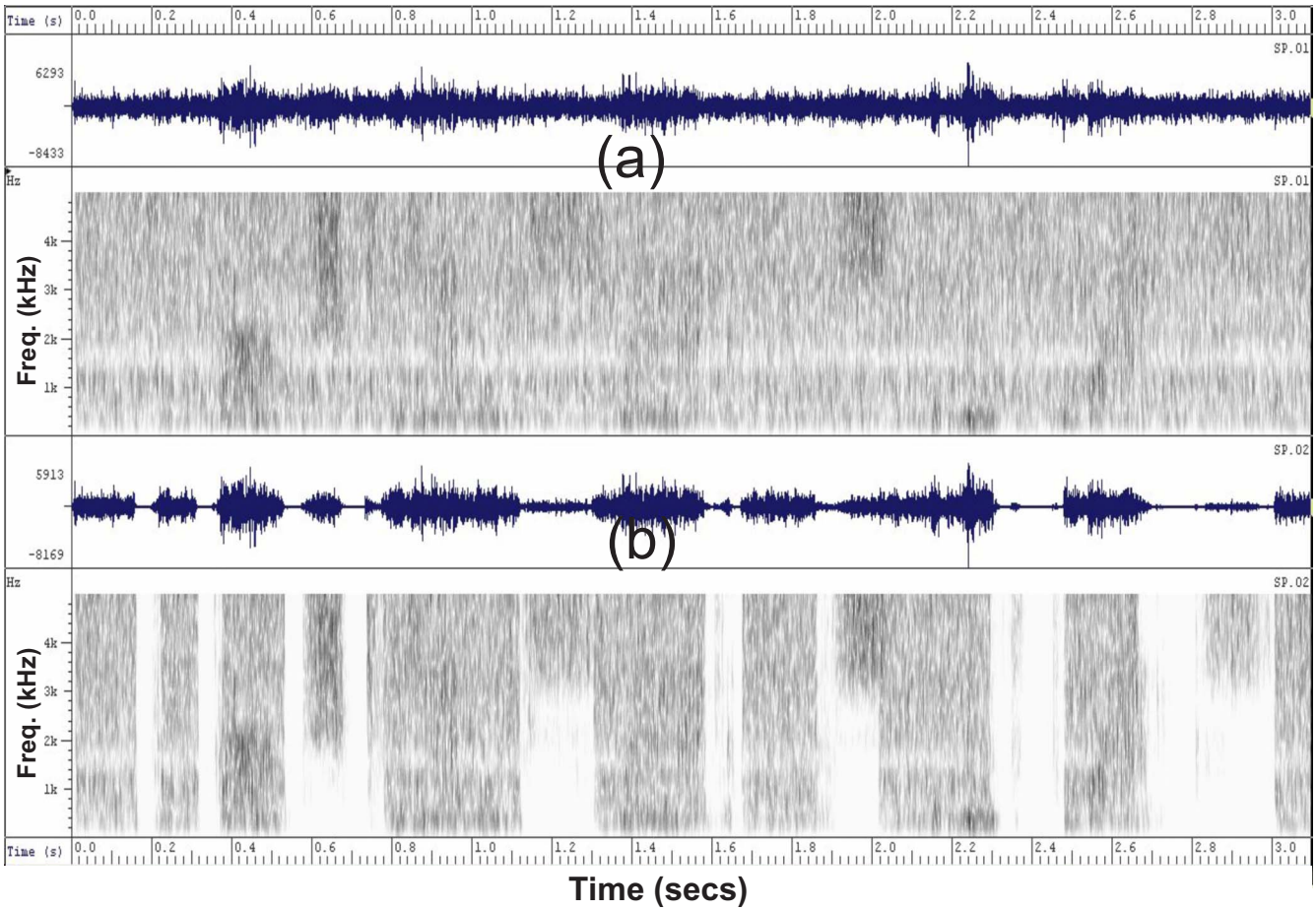


FIG. 1. (Color online) (a) Top panel shows time waveform and wide-band spectrogram of a sentence in 5-dB steady noise vocoded into six channels. (b) Bottom panel shows the same sentence containing clean (vocoded) obstruent segments but corrupted sonorant segments.

subjects listened to vocoded (6, 12, and 22 channels) sentences to become familiar with the processed stimuli. Sentences taken from the H.I.N.T. corpus (Nilsson *et al.*, 1994) were used for the training session. The training session lasted for about 20–30 min. During the test, the subjects were asked to write down the words they heard. Subjects participated in a total of 36 conditions (=3 SNR levels $\times$2 algorithms $\times$2 maskers $\times$3 channels). Two lists of IEEE sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. Sentences were presented to the listeners in blocks, with 20 sentences/block for each condition. The different conditions were run in random order for each listener.

B. Results

The mean scores for all conditions are shown in Fig. 2. Performance was measured in terms of the percentage of words identified correctly (all words were scored). Results are divided into three panels according to the number of channels used, and the individual panels are plotted as a function of SNR level. The performance obtained in quiet (denoted as Q) is also shown for comparative purposes. Two-way analysis of variance (ANOVA) with repeated measures was used to assess effects of masker type. The control noisy stimuli (shown in Fig. 2 with open symbols) processed via six channels showed no significant effect of masker type [$F(1,6)=3.8$, $p=0.098$]. No significant interaction [$F(2,12)=3.3$, $p=0.07$] was found between SNR level and masker type. Similarly, vocoded speech processed via 12 channels showed no significant effect of masker type [$F(1,6)=0.9$, $p=0.379$] and non-significant interaction [$F(2,12)=0.03$, $p=0.96$] between SNR level and masker type. Finally, vocoded speech processed via 22 channels showed significant effect [$F(1,6)=40.1$, $p=0.001$] of masker type and non-significant interaction [$F(2,12)=3.24$, $p=0.075$] between SNR level and masker type. Performance with steady noise was significantly better than performance with the two-talker masker, consistent with findings reported in CI studies (e.g., Stickney *et al.*, 2004). The data obtained in the 6- and 12-channel conditions are partially consistent with that obtained in the studies by Stickney *et al.* (2004) and Nelson and Jin (2004). Performance obtained in the present study with 6 and 12 channels was limited to some degree by flooring effects, at least in the low SNR levels (−5 and 0 dB), and therefore failed to show a masker effect.

A different pattern in performance emerged with the vocoded stimuli in which the obstruent consonants were clean and the remaining sonorant sounds were left corrupted (shown in Fig. 2 with filled symbols). Vocoded speech processed through six channels showed a significant effect [$F(1,6)=10.9$, $p=0.016$] of masker type. Performance obtained with the two-talker masker was significantly higher than performance obtained with the steady noise masker. That is, subjects benefited from the masker fluctuation. Vo-

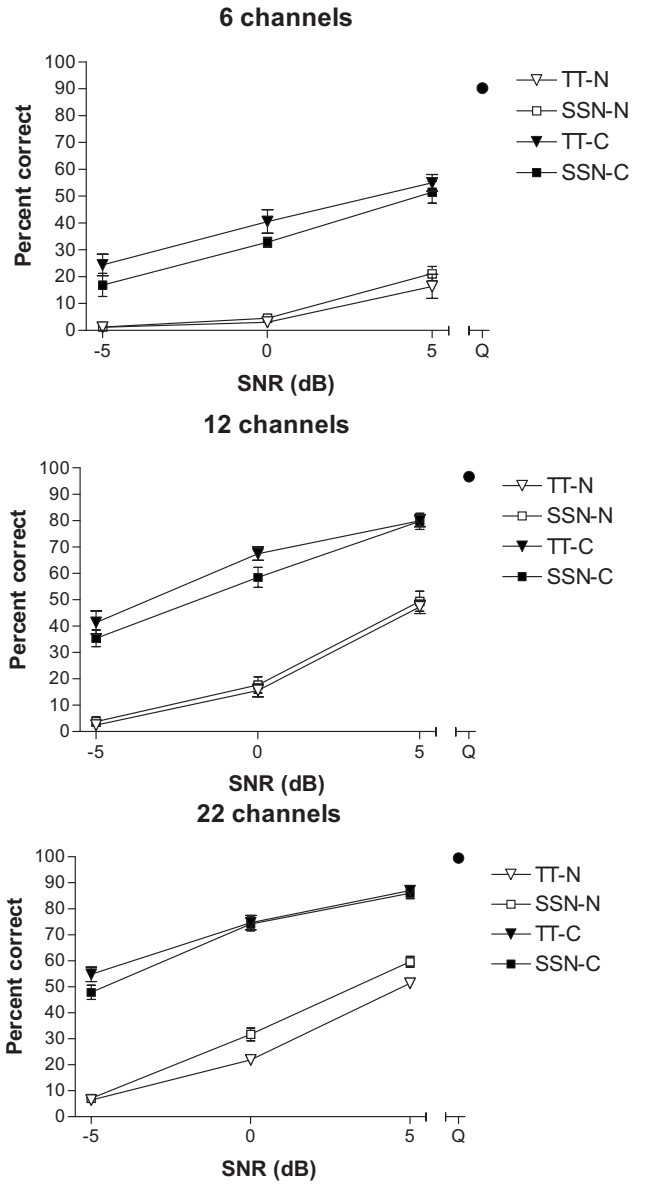


FIG. 2. Mean speech recognition scores as a function of SNR level for the various masker (TT=two-talker and SSN=steady noise) and channel conditions. Filled symbols denote scores obtained with stimuli containing clean obstruent consonants, and open symbols denote scores obtained with the control stimuli containing corrupted obstruent consonants. The performance obtained in quiet (Q) is also shown for comparative purposes. Error bars denote standard errors of the mean.

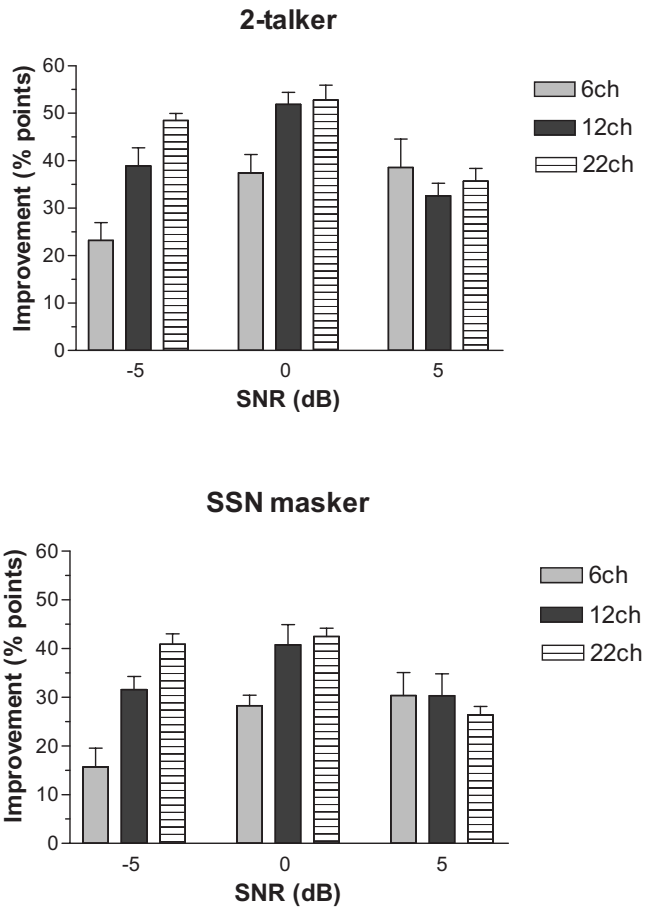


FIG. 3. Improvement in performance (in terms of overall percentage points) obtained when listeners had access to clean obstruent consonants in the various channel and SNR conditions. Error bars denote standard errors of the mean.

coded speech processed through 12 channels also showed a significant effect [$F(1,6)=14.3$, $p=0.009$] of masker type. The interaction between SNR level and masker type was not significant ($p>0.2$) in either 6-channel or 12-channel conditions. The 22-channel condition showed a non-significant effect [$F(1,6)=5.43$, $p=0.059$] of masker type and a significant [$F(2,12)=4.0$, $p=0.047$] interaction. *Post-hoc* tests indicated no significant ($p>0.05$) difference in performance between the two masker types at 0- and 5-dB SNR, but a significant ($p=0.008$) difference at −5-dB SNR. Performance obtained with 22 channels at −5-dB SNR with the two-talker masker was significantly higher than performance obtained with steady noise.

Introducing the clean obstruent consonants in the corrupted vocoded stimuli produced a substantial improvement in performance in all SNR and channel conditions (Fig. 2). For better clarity, Fig. 3 depicts the improvement in performance (in terms of overall percentage points) in the various channel and SNR conditions. The improvement ranged from a low of 15 percentage points (with the noise masker and with 6 channels) to a high of 50 percentage points (with the two-talker masker and with 12 or 22 channels). The estimated speech reception threshold improvement in performance (as assessed by interpolating the scores in Fig. 2) for the 12- and 22-channel conditions was near 10 dB. A steady improvement of 30–40 percentage points was observed in the 5-dB SNR conditions, independent of the number of channels used. A mild dependency on the number of channels was noted in the −5 and 0-dB SNR conditions, with 12 and 22 channels producing larger improvement than 6 channels. ANOVA analysis confirmed that the improvement in performance was highly significant in all channel and SNR conditions. Performance, for instance, with 6-channel vocoded speech containing clean obstruent consonants was significantly higher in both the two-talker masker [$F(1,6)=160.6$, $p<0.0005$] and steady masker [$F(1,6)=444.9$, $p<0.0005$] conditions, compared to the corresponding 6-channel control vocoded conditions with the noisy obstruent consonants. There was no significant interaction ($p>0.05$) between SNR level and processing (clean vs. corrupted obstruent consonants). In summary, the magnitude of

the improvement in performance obtained by listeners when they had access to information provided by the clean obstruent consonants seems to depend on both the SNR level and number of channels. As it will be discussed next, this dependency is probably due to the different sets of acoustic cues (and reliability of those cues) available to the listeners when presented with speech vocoded into a small number (e.g., 6) vs. a large number (e.g., 22) of channels.

III. DISCUSSION

Performance obtained by NH subjects when listening to spectrally degraded speech containing clean obstruent sounds but noisy sonorant sounds was significantly higher in conditions in which speech was corrupted by a two-talker masker than by a steady-noise masker (Fig. 2). That is, subjects benefited from release of masking when they had access to information carried by the clean obstruent consonants. We contend that the obstruent consonants carry information about the location of acoustic landmarks that are present in the signal. Knowing the location of these landmarks is crucial as it enables listeners to identify word boundaries and fuse the pieces of the underlying message across temporal gaps. The listeners were able to do this even in conditions wherein speech was degraded to six channels. The importance and contribution of acoustic landmarks to speech recognition in noise are cast in a lexical-access framework and are discussed next along with the implications of the present findings in CIs.

A. Contribution of obstruent consonants and acoustic landmarks to masking release

Many speech recognition models (e.g., Stevens, 2002; Pisoni and Sawusch, 1975; Cutler and Norris, 1988) assume that speech is first segmented at reliable points ("islands of reliability") of the acoustic signal followed by a classification of the segmented units into a sequence of phonetic units (e.g., syllables and words). The identified phonetic units are then matched against the items in the lexicon to select the best word sequence intended by the speaker. Stevens (2002) proposed a lexical-access model based on distinct acoustic landmarks partitioning vowels and consonants. The prelexical stage of this model consists of three steps. In the first step, the signal is segmented into acoustic landmarks based on detection of peaks and spectral discontinuities in the signal. These landmarks define the boundaries of the vowels, consonants, and glide segments. The second step involves extraction of acoustic cues from the vicinity of the landmarks signifying which articulators are active when the vowel, glide, or consonant landmarks are created and how these articulators are shaped or positioned. The third step consolidates, taking context into account, all the cues collected in step 2 to derive a sequence of features for each of the landmarks detected in step 1. In this speech perception model, each word in the lexicon is represented by a set of features that specify the manner of articulation, the position and shape of the articulators, as well as information about the syllable structure.

It is clear that the first step in Stevens' (2002) model (i.e., segmentation into acoustic landmarks) is crucial to the lexical-access model. If the acoustic landmarks are not detected accurately by the listeners or if the landmarks are perceptually not clear or distinct owing to corruption of the signal by external noise, this will affect the subsequent stages of the model in the following ways. First, errors will be made in step 2 in terms of identifying the shape and position of the articulators used in each landmark. Second, the absence of reliable landmarks can disrupt the syllable structure, which is known to be important for determining word boundaries in fluent speech. This is so because the onset of a word is always the onset of a syllable. Therefore, not knowing when the syllable starts makes word boundary determination very difficult (e.g., Gow *et al.*, 1996). In summary, external noise can degrade the salient cues present in syllable-initial consonants. In the context of CIs, envelope compression can also degrade these cues (see discussion in Sec. III B). These cues are present in the vicinity of the acoustic landmarks, hence identifying or somehow enhancing access to these landmarks ought to aid in identifying word boundaries and consequently improving word recognition.

When comparing the two types of maskers used in the present study, it is reasonable to expect that the two-talker masker provides more visible and perceptually more reliable acoustic landmarks than the noise masker. These landmarks include not only the ones associated with the obstruent consonants occupying the low/high frequency regions of the spectrum but also the vowel-to-glide landmarks occupying the mid-frequency region of the spectrum. Consequently, we expect higher performance with the two-talker masker than the steady noise masker. Indeed, subjects received significant release from masking (Fig. 2) when speech was processed through 6 and 12 channels and had access to acoustic landmarks present in the clean obstruent consonants. No significant release of masking was noted with 22 channels, at least for SNR $\geq$0 dB, and we believe that was because of a trading relationship between spectral resolution and importance of acoustic landmarks. When the spectral resolution is poor (e.g., six channels), then speech redundancy is greatly reduced and listeners have to rely on an alternative set of cues, such as those provided by acoustic landmarks to identify word boundaries. Hence, the importance of acoustic landmarks is greatly amplified when the spectral resolution is poor. However, when the spectral resolution is fine (e.g., 22 channels) and the SNR level is relatively high listeners can use other cues in addition to those introduced by acoustic landmarks. For one, listeners could use F1/F2 transition information which is adequately represented with 22 channels (e.g., Qin and Oxenham, 2003). In contrast, F1/F2 information is poorly represented in speech vocoded into a small number of channels (e.g., six channels) as the formant transitions might fall in the same band. The study by Munson and Nelson (2005), for instance, demonstrated that CI users have difficulty discriminating synthetic speech stimuli on the basis of formant transitions. Hence, when the spectral resolution is fine and the SNR is sufficiently high ($\geq$0 dB), then the acoustic landmarks play a comparatively minor role on speech recognition as the listeners have access to other cues (e.g., F2 transitions). It is for this reason that we believe that no masking release was noted with 22 channels in most con-

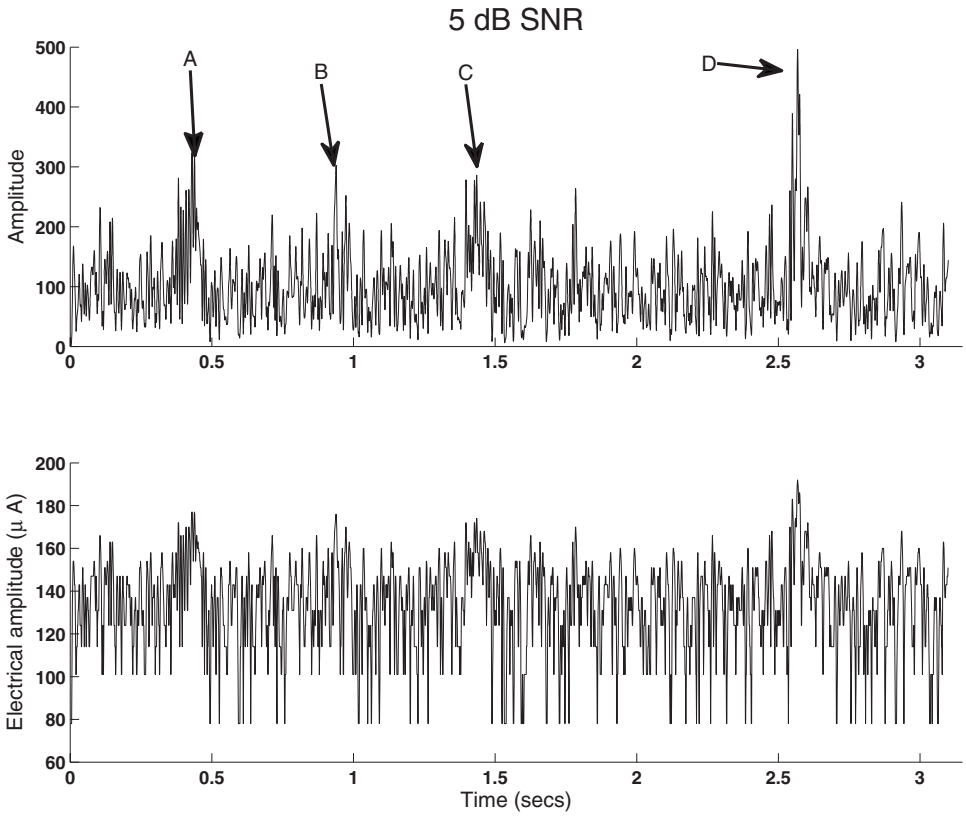


FIG. 4. Envelope (fourth channel with center frequency of 600 Hz) shown before (upper panel) and after (bottom panel) applying log-type compression to a sentence embedded in speech-shaped noise at 5-dB SNR. Arrows show some of the dominant vowel peaks present.

ditions, except in the SNR=−5-dB condition, wherein the acoustic landmarks were severely smeared by external noise. It is also for this reason that we believe that a larger improvement in performance (Fig. 3) was obtained with 12–22 channels compared to 6 channels (at least in the −5- and 0-dB SNR conditions) since the listeners had access to more cues.

The present study focused on the contribution of acoustic landmarks introduced by obstruent sounds on speech recognition in steady and fluctuating noise conditions. The sonorants (e.g., vowels and glides) also introduce landmarks in the signal (Stevens, 2002), but were not studied in this paper. We cannot exclude the possibility that the masking release observed with clean obstruent segments would have been just as large if clean sonorant segments were introduced or if equally long clean segments were placed randomly across the sentence. Introducing clean sonorant segments, however, does not reflect realistic noisy conditions since the acoustic noise does not degrade the sonorant segments to the same degree as the obstruent segments. As mentioned in the Introduction, we restricted our attention to the landmarks introduced by obstruent sounds for the main reason that these sounds are more susceptible to noise (i.e., easily masked) than the sonorant sounds (Phatak and Allen, 2007; Parikh and Loizou, 2005).

B. Implications for CIs

As mentioned in the Introduction, CI users do not receive release of masking when listening to speech in modulated interference (e.g., Nelson *et al.*, 2003; Fu and Nogaki, 2004; Stickney *et al.*, 2004; Cullington and Zeng, 2008). Based on the findings from the present study, we believe that two interrelated factors contribute to that. The first factor is envelope compression and reduced dynamic range. In current CI systems, the envelopes extracted from each band are compressed with a logarithmic function in order to map the wide acoustic dynamic range to the small (5–15-dB) electrical dynamic range. This envelope compression smears the acoustic landmarks a great deal (more so in noise) making it extremely difficult for CI users to identify word boundaries. Figure 4 shows an example envelope (fourth channel) obtained before and after applying log-type compression to a sentence embedded in speech-shaped noise at 5-dB SNR. Figure 5 shows the same sentence in quiet. Arrows shown and labeled as A–D (Fig. 4) point to some of the vowel landmarks and arrows labeled as E–G (Fig. 5) point to some of the obstruent landmarks. It is clear from Fig. 4 (bottom panel) that the obstruent landmarks are not easily discernible and probably not easily perceptible. As an indirect measure of assessing the presence of obstruent landmarks, one can compute the peak-to-trough ratio for the peaks labeled as A–D in Fig. 4. For instance, for peak A and trough E (Fig. 5), the peak-to-trough ratio for the linearly processed (i.e., no compression) envelope is 10 dB, where the trough level is set to the mean noise floor level of that channel. The corresponding peak-to-trough ratio for the compressed envelope is 2.4 dB. As shown in a previous study in our laboratory (Loizou and Poroy, 2001), it is unlikely that patients will be able to perceive such a small (~2-dB) peak-to-trough ratio and consequently perceive the acoustic landmarks introduced by the obstruent consonants.

The second factor, which is a direct consequence of the first, is poor access to the location of the acoustic landmarks needed to determine word or syllable boundaries. Poor spectral resolution further exacerbates the situation as it reduces

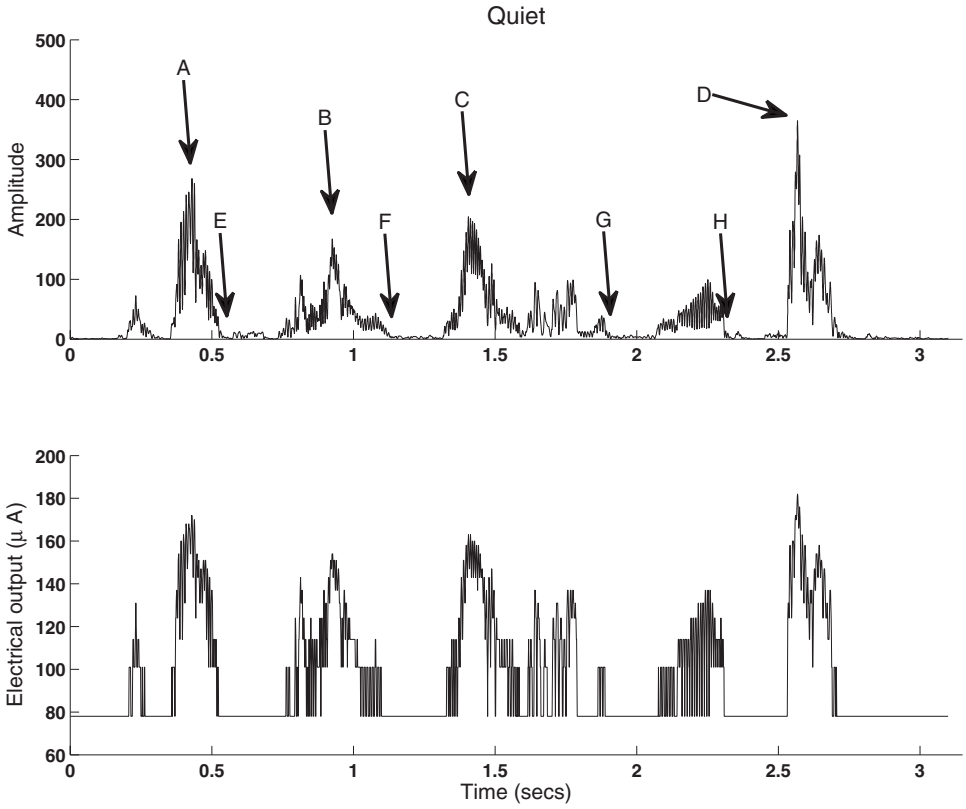


FIG. 5. Envelope (fourth channel with center frequency of 600 Hz) shown before (upper panel) and after (bottom panel) applying log-type compression to a sentence (same as in Fig. 4) in quiet. Arrows labeled E–G show some of the obstruent landmarks present, and arrows labeled A–D show some of the vowel landmarks.

speech redundancy and forces listeners to rely more on information carried by acoustic landmarks to identify word or syllable boundaries. Without good and accurate knowledge of the location of the acoustic landmarks, it becomes extremely difficult for users to first identify the pieces (based perhaps on their delineating boundaries) of the underlying message and then integrate those pieces together. Hence, the second factor is quite detrimental as it limits the CI user's ability to integrate information across temporal gaps into one coherent speech stream.

In terms of (indirectly) addressing the second factor, it is extremely challenging to improve spectral resolution, at least with existing technology and knowledge. Increasing the number of electrode contacts, for instance, would not necessarily increase the number of independent channels of information (e.g., Friesen *et al.*, 2001). Fortunately, there are ways to address the first factor. One can use a dynamically changing compression function that is less compressive during the obstruent consonant segments and more compressive during the sonorant sound segments (current CIs use the same shape compression function for all phonetic segments). Such a strategy would require the use of automatic landmark detection algorithms that would identify the locations of acoustic landmarks in the signal. Fortunately, several such algorithms exist in the literature and have been found to work quite well, at least in quiet (e.g., Liu, 1996; Junega and Espy-Wilson, 2008). An alternative approach was proposed by Kasturi and Loizou (2007) based on the use of s-shaped input-output functions which are expansive for low input levels, up to a knee point level, and compressive thereafter. The knee points of the s-shaped functions changed dynamically and were set proportional to the estimated noise floor level. For the most part, the expansive (i.e., less compressive) part of the s-shaped functions operated on obstruent segments, which generally have lower intensity compared to that of sonorant segments. The main advantage of using s-shaped functions to map the acoustic signal to electrical output values is that these functions do not require landmark detection algorithms as they are applied to all phonetic segments. Replacing the conventional log mapping functions with the s-shaped functions yielded significant improvements in speech intelligibility in noise by nine CI users (Kasturi and Loizou, 2007). In summary, one research direction that warrants further investigation is the development of dynamically changing compression functions that would maintain or enhance the acoustic landmarks present in the signal. In doing so, and according to the present data (Fig. 2) as well as the data from Kasturi and Loizou (2007), it is reasonable to expect that CI users would obtain large gains in intelligibility in noisy backgrounds and receive release of masking.

Finally, the data from Fig. 2 suggest that the obstruent consonants contribute significantly to speech recognition in noise. Large improvements in performance were observed when listeners had access to the clean obstruent spectra even when the spectral resolution was poor (see Figs. 2 and 3). The improvement was quite substantial and amounted to 30–50 percentage points for speech vocoded in 12–22 channels and to 15–30 percentage points for speech vocoded in 6 channels (Fig. 3). In the context of CI processing, the data in Figs. 2 and 3 also suggest that the obstruent consonants need to be processed, or at least treated, differently than the sonorant sounds. One possibility is to apply, as mentioned above, different shape compression functions to the obstruent segments, and another possibility is to enhance and/or clean

selectively the noisy obstruent spectra. Such techniques have the potential of preserving the acoustic landmarks in the signal and consequently improving speech recognition in noise by CI users.

IV. CONCLUSIONS

It is well established that CI users do not receive release of masking when listening to speech in fluctuating maskers (e.g., Nelson *et al.*, 2003; Stickney *et al.*, 2004). The present study tested a new hypothesis for the observed absence of release of masking using vocoded speech and NH listeners as subjects. The proposed hypothesis is that the CI user's ability to fuse information across temporal gaps is limited by their ability to perceive acoustic landmarks such as those introduced by obstruent consonants. These landmarks play an important role in models of lexical access (Stevens, 2002) and are needed to identify word/syllable boundaries. Results indicated that when listeners were presented with vocoded speech (6–22 channels) with corrupted sonorant sounds (e.g., vowels) but clean obstruent sounds (e.g., stops), they were able to receive release of masking, even with 6 channels of stimulation. That is, listeners performed better in fluctuating maskers than in steady-noise maskers. Dramatic improvements in performance were observed when listeners had access to the clean obstruent spectra even when the spectral resolution was poor. The improvement amounted to 30–50 percentage points for speech vocoded in 12–22 channels and to 15–30 percentage points for speech vocoded in 6 channels. The present data suggest that the obstruent consonants need to be treated differently in noisy conditions so as to preserve the acoustic landmarks present in the signal. One possibility suggested is to apply a different shape compression function (e.g., less compressive) during the obstruent segments since the log-compressive function tends to smear and, in some cases, abolish critical obstruent landmarks (see example in Fig. 4).

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC007527 from the National Institute of Deafness and other Communication Disorders, NIH.

Assmann, P., and Summerfield, Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. Ainsworth, A. Popper, and R. Fay (Springer, New York), pp. 231–308.

Blumstein, S., and Stevens, K. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1017

Cullington, H., and Zeng, F.-G. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Festen, J., 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.

Friesen, L. M., Shannon, R. Y., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q., and Nogaki, G. (**2004**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Gow, D. W., Melvold, J., and Manuel, S. (**1996**). "How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing," Proc. Int. Conf. Spoken Lang. Proc., Philadelphia, PA, pp. 66–69.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Junega, A., and Espy-Wilson, C. (**2008**). "A probabilistic framework for landmark detection based on phonetic features for automatic speech recognition," J. Acoust. Soc. Am. **123**, 1154–1168.

Kasturi, K., and Loizou, P. (**2007**). "Use of s-shaped input-output functions for noise suppression in cochlear implants," Ear Hear. **28**, 402–411.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.

Li, N., and Loizou, P. (**2008**). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise," J. Acoust. Soc. Am. **124**, 3947–3958.

Liu, S. (**1996**). "Landmark detection for distinctive feature-based speech recognition," J. Acoust. Soc. Am. **100**, 3417–3430.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC, Boca Raton, FL).

Loizou, P., and Poroy, O. (**2001**). "Minimum spectral contrast needed for vowel identification by normal-hearing and cochlear implant listeners," J. Acoust. Soc. Am. **110**, 1619–1627.

Mines, M., Hanson, B., and Shoup, J. (**1978**). "Frequency of occurrence of phonemes in conversational English," Lang Speech **21**, 221–241.

Munson, B., and Nelson, P. (**2005**). "Phonetic identification in quiet and in noise by listeners with cochlear implants," J. Acoust. Soc. Am. **118**, 2607–2617.

Nelson, P., and Jin, S. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.

Nelson, P., Jin, S., Carney, A., and Nelson, D. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Nilsson, M., Soli, S., and Sullivan, J. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.

Peters, R., Moore, B., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Phatak, S., and Allen, J. (**2007**). "Consonants and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.

Pisoni, D., and Sawusch, J. (**1975**). "Some stages of processing in speech perception," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. Nooteboom (Springer-Verlag, Berlin), pp. 16–34.

Qin, M., and Oxenham, A. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Qin, M., and Oxenham, A. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Seneff, S., and Zue, V. (**1988**). "Transcription and alignment of the TIMIT database," in *Proceedings of the Second Symposium on Advanced Man-Machine Interface Through Spoken Language*, Oahu, HI (20–22 November 1988).

Stevens, K. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Stickney, G., Assmann, P., Chang, J., and Zeng, F.-G. (**2007**). "Effects of implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Stickney, G., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Multiband product rule and consonant identification

Feipeng Li[a] and Jont B. Allen
Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

(Received 31 July 2008; revised 5 February 2009; accepted 6 May 2009)

The multiband product rule, also known as band-independence, is a basic assumption of articulation index and its extension, the speech intelligibility index. Previously Fletcher showed its validity for a balanced mix of 20% consonant-vowel (CV), 20% vowel-consonant (VC), and 60% consonant-vowel-consonant (CVC) sounds. This study repeats Miller and Nicely's version of the hi-/lo-pass experiment with minor changes to study band-independence for the 16 Miller–Nicely consonants. The cut-off frequencies are chosen such that the basilar membrane is evenly divided into 12 segments from 250 to 8000 Hz with the high-pass and low-pass filters sharing the same six cut-off frequencies in the middle. Results show that the multiband product rule is statistically valid for consonants on average. It also applies to subgroups of consonants, such as stops and fricatives, which are characterized by a flat distribution of speech cues along the frequency. It fails for individual consonants. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3143785]

PACS number(s): 43.71.An, 43.71.Es, 43.71.Gv [MSS] Pages: 347–353

I. INTRODUCTION

A fundamental problem of human speech perception is how the human auditory system integrates speech cues across frequency. The most relevant study on this topic dates back to the 1920s, when Fletcher[1] at Bell Laboratories investigated speech articulation over voice communication systems. Low-pass and high-pass filtered "nonsense syllables" were used for the study of phone recognition. They found that the average phone error e of the full-band stimuli is equal to the product of the error of the low-pass filtered stimuli e_L and the error of the complimentary high-pass filtered stimuli e_H, that is,

$$e = e_L \times e_H. \tag{1}$$

In other words, the low-pass band and the high-pass band are consistent with the assumption that the low band and high band are independent. Equation (1) was then generalized, by assumption, into a multiple band form[1–3]

$$e = e_1 e_2 \cdots e_K. \tag{2}$$

The number of independent articulation bands is generally taken to be $K=20$, which makes each band correspond to about 1 mm along the basilar membrane.[2]

Let s denote the average phone articulation (i.e., the probability of the nonsense phones being correctly recognized), then the articulation error $e=1-s$, and the articulation band error $e_1=1-s_1$, etc. Given Eq. (2),

$$\log(1-s) = \sum_{k=1}^{K} \log(1-s_k). \tag{3}$$

Notice that $\log(1-s_k)$ is similar to the definition of entropy,[4] thus may be interpreted as the information carried by the kth band.[2,3] Equation (3) implies that the human speech recognition system consists of at least K parallel channels and that the total information is equal to the sum over the information in the K articulation bands. This relation may also be called the *additivity law of frequency integration*. It is the foundation of the two ANSI standards: articulation index (AI) (Ref. 5) and more recently speech intelligibility index (SII).[6]

Based on the assumption of independent articulation bands, French and Steinberg[7] developed a method for calculation of AI based on the intensity of the long-term average speech and noise. Following the verification by Beranek[8] and Kryter,[9] French and Steinberg's method[7] became an ANSI standard in 1969. Then in 1970–1980 Steeneken and Houtgast[10] extended the AI to the speech transmission index (STI) by introducing a modulation transfer function to account for reverberation and peak clipping. The original AI was developed for the use of normal hearing listeners. Later it was extended to estimate the speech intelligibility for the hearing-impaired listener,[11–14] resulting in a new ANSI standard named the SII. All the three models, AI, STI, and SII, are based on the same Fletcher–Galt assumption[3] that the total articulation is the sum of the contribution from multiple independent narrow bands.

Despite its importance to the widely used articulation models, the validity of the multiband product rule [Eq. (2)] has actually been a key open question.[15] For example, Kryter[16] showed that AI was a valid predictor of the intelligibility of speech under a wide variety of conditions of noise masking and speech distortion except for the cases of three non-contiguous pass bands at 0–600, 1200–2400, and 4800–9600 Hz. Grant and Braida[17] found that the predicted AI based on the sum of the AIs from individual bands was greater than the observed AI by approximately 18% for adjacent 1/3-octave bands, while the AI predicted for combinations of non-adjacent bands was less than the observed AI by approximately 41%. Lippmann[18] also found that the stop-band data did not agree with AI calculation. In 2001 Müsch and Buus[19,20] coined two new terms *synergistic* and *redun-*

[a]Author to whom correspondence should be addressed. Electronic mail: fli2@illinois.edu

dant interactions between neighboring bands to explain why the AI under, or over, estimates the wide-band error, compared to the product of the errors associated with the narrow bands. It has been conjectured that a revised model, which accounts for the mutual dependency between adjacent bands, might give a better prediction.[21] In a recent study, Ronan *et al.*[22] compared several frequency integration models for the prediction of individual consonant articulation score, for narrow-band cases. Results indicated that Fletcher's product rule[1] [Eq. (2)] made satisfactory predictions under various combinations of adjacent and non-adjacent narrow-band speech, except for the case of multiple high-frequency narrow bands, for which none of the evaluated methods are satisfactory. Investigation of SII (Ref. 23) also found that it greatly over-predicted performance at high sensation levels, and under-predicted performance at low sensation levels for many hearing-impaired listeners. The information contained in each frequency band is not strictly additive.

In 1955, Miller and Nicely[24] (MN55) repeated Fletcher and Galt's high-pass and low-pass filtering experiment[3] for the analysis of perceptual confusion. The speech stimuli includes 16 consonant sounds, /p,t,k,f,θ,s,ʃ,b,d,g,v,ð,z,ʒ,m,n/ spoken initially before the vowel /a/. Using the data from experiment MN55, we checked the validity of Fletcher's product rule [Eq. (1)].[1] Results[24] show that the model applies to the consonants on average, despite that it over-predicts the full-band error by 10%. We then plotted the product of e_L and e_H against the full-band error e for each of the 16 consonant sounds [see Fig. 2(b)]. To our surprise, more than half of the consonant sounds, specifically, /p,k,f,ʃ,b,d,g,ʒ,m,n/, show only small discrepancy.

Designed for the purpose of confusion analysis, the MN55 data are unsuitable for the study of the multiband product rule, for several reasons. First, the frequency samples are limited. Only six low-pass and five high-pass condition are included, in contrast Fletcher[1] and French and Steinberg[7] suggested $K=20$ frequency points. Second, the cut-off frequencies are not evenly distributed along the effective range of speech communication. Four out of six low-pass samples are below 1.5 kHz, with only one high-pass sample within the same frequency range. Interpolation between data points introduces significant error.

In the present study we investigate the validity of the multiband product rule for consonant sounds. The product rule is evaluated on three levels: (1) 16 consonants on average, (2) subgroups such as stops and fricatives, and (3) individual consonants. A computer-based high-pass and low-pass experiment, named HL07, is designed for this purpose (see Fig. 1). The new experiment utilizes the same 16 consonant sounds as experiment MN55. To address the problems listed above, the cut-off frequencies were chosen such that the basilar membrane is evenly divided into 12 bands over the frequency range from 0.25–8 kHz, with the low-pass and high-pass filters sharing the same six cut-off frequencies in the mid-frequency range.

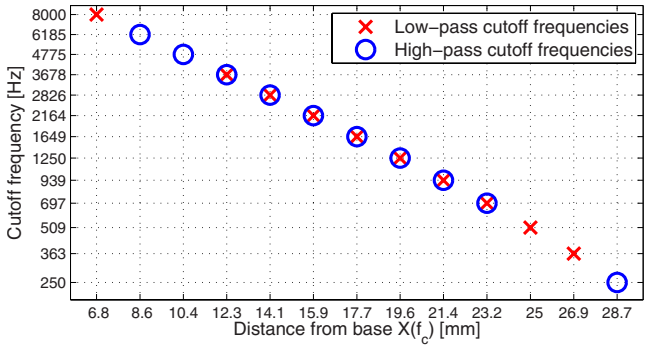


FIG. 1. (Color online) High-pass and low-pass cut-off frequencies of experiment HL07.

II. METHODS

A. Subjects

19 normal hearing subjects were enrolled in the experiment, of which 6 male and 12 female listeners completed. Except for one subject in her 40s, all the subjects were college students in their 20s. The subjects were born in the United States with English being their first language. All subjects were paid for their participation. IRB approval was obtained for the experiment. In order to make sure that all the data are of high quality, the performance of the listeners was assessed by their average recognition score. Those who had abnormally low scores will be excluded for further analysis. In experiment HL07, no subject has been removed for that reason.

B. Speech stimuli

The same 16 nonsense consonant-vowels (CVs) used by Miller and Nicely[24] were chosen. A subset of wide-band syllables sampled at 16 kHz were taken from the LDC-2005S22 corpus. Each CV was spoken by 20 talkers, among which only 6 utterances, half male and half female, were finally chosen for the test, to reduce the total duration of the experiment. The six utterances were selected such that they were representative of the speech material in terms of confusion patterns and articulation score based on the results of similar speech perception experiment.[25] The speech sounds were presented to both ears of the subjects at the listener's most comfortable level, but always less than 80 dB SPL.

C. Conditions

The subjects were tested under 19 filtering conditions, including 1 full-band 0.25–8 kHz, 9 high-pass, and 9 low-pass conditions. The cut-off frequencies were calculated from Greenwood's inverse cochlear map function[26] such that the full-band frequency range (0.25–8 kHz) was divided into 12 bands, corresponding to equal length along the basilar membrane. Figure 1 illustrates the frequency samples and the correspondent distances from the base on the human basilar membrane. The cut-off frequencies of the high-pass filtering were 6185, 4775, 3678, 2826, 2164, 1649, 1250, 939, and 697 Hz, with the upper-limit at 8000 Hz. The cut-off frequencies of the low-pass filter were 3678, 2826, 2164, 1649, 1250, 939, 697, 509, and 363 Hz, with a lower-limit at 250

Hz. The high-pass and low-pass filtering shared the same cut-off frequencies over the middle frequency range that contains most of the speech information. The filters were sixth order elliptical filter with 0.02 dB of peak-to-peak ripple and a stop-band attenuation of −60 dB. To make the filtered speech sound more natural and to mask the stop bands, white noise was used to mask the stimuli at the signal-to-noise ratio (SNR) of 12 dB, based on the average speech spectra of the 96 nonsense syllables.

D. Procedure

The speech perception experiment was conducted in a sound-proof booth. A MATLAB code was developed for the collection of the data. Speech stimuli were presented to the listeners through Sennheisser HD 280-pro headphones. Subjects responded by clicking on the button labeled with the CV that they heard. In case the speech was completely masked by the noise, or the processed token did not sound like any of the 16 consonants, the subjects were instructed to click on a "noise only" button. A total of 2208 tokens were randomized and divided into 16 sessions, each of which lasted for about 15 min. A mandatory practice session of 60 tokens was given at the beginning of the experiment. To prevent fatigue the subjects were instructed to take frequent breaks. The subjects were allowed to play each token for up to three times. At the end of each session, the subject's test score, together with the average score of all listeners, was shown to the listener to provide feedback on their relative progress, as motivation.

E. Difference between HL07 and MN55

Although experiment HL07 can be regarded as a repeat of the MN55 study, the two experiments are distinguished in several important aspects. First, the subjects differ in gender and proficiency. In MN55 five extensively-trained female subjects served as both talkers and listening crew. This introduced a "coupling" effect between the talkers and the listeners, as well as an awareness of the relative difficulty of the sounds. In HL07 we use recorded speech prepared by ten male and eight female talkers from the LDC database. All the 18 subjects (6 male and 12 female) are naive listeners without any experience in speech perception tests. Second, the noise levels are different. Both experiments use white noise at 12 dB SNR. However, in experiment MN55, the speech level was controlled by a volume unit (VU) meter,[27] which measures the speech peaks, while in experiment HL07 the noisy speech were created by setting the rms level of the speech and noise. Thus 12 dB SNR in MN55 is about the same as 14 dB SNR in HL07.[27] As a consequence, the full-band error of MN55 is about 12% lower than that of HL07. Third, the filtering conditions are different. In MN55 the full-band speech was created by a wide-band filter of 0.2–6.5 kHz, and then the distorted speech were created by filtering the full-band speech with low-pass cut-off frequencies of 0.3, 0.4, 0.6, 1.2, 2.5, and 5 kHz and high-pass cut-off frequencies of 0.2, 1.0, 2.0, 2.5, 3.0, and 4.5 kHz. In contrast, the full-band speech in HL07 goes to 8 kHz. The loss of information from 6.5 to 8 kHz accounts well for the over-prediction of MN55 in the high frequency. Fourth, the test platforms are different. Data collection in MN55 was paper-based. The listeners were told to choose a response from the 16 nonsense CVs and write it down on the answer sheet within seconds following the presentation. The HL07 experiment is computer-based. No limit is applied for the responding time. Subjects were allowed to play each sound up to three times. In case the subjects could not tell which sound is presented, a noise only button was added.

F. Data analysis

The validity of Fletcher's product rule[1] [Eq. (1)] is investigated for average speech and individual consonants. The probability of error of a token (an utterance filtered at a frequency) is defined as the number of mis-labeled responses divided by the total number of presentations. The mean error of a consonant is the average over the six tokens pronounced by different talkers. Similarly, the total error of average speech can be calculated by averaging the errors of the 16 consonants. For both average speech and individual consonants, the fitness of the model to the data is evaluated in terms of average bias $B(f_c)$ and $\chi^2(f_c)$ computed from the error of all listeners. The average bias is given by

$$B(f_c) = e - e_L \times e_H, \tag{4}$$

where $e_L \times e_H$ and e are the model error and observed error at a cut-off frequency f_c. The chi-square statistic is

$$\chi^2(f_c) = N\frac{[(1-e_L\times e_H)-(1-e)]^2}{1-e_L\times e_H} + N\frac{[e-e_L\times e_H]^2}{e_L\times e_H}, \tag{5}$$

where N is the total number of presentations for the particular condition. The quantities $(1-e_L\times e_H)$ and $(1-e)$ are the predicted and observed scores. A significance level (the probability of this result not being due to chance) of 0.05 is chosen as the threshold of the chi-square test. A value of χ^2 greater than the threshold indicates that the measurements do not satisfy Eq. (1) at that condition, whereas when χ^2 is less than the threshold of significance, Fletcher's product rule[1] can be regarded as true.

The above analysis is carried out by treating the 18 listeners as average normal listeners. In order to determine if the same conclusion applies to any individual listeners, a one-way analysis of variance (ANOVA) test is applied to the $e-e_L\times e_H$ of different listeners following each χ^2 test. Due to the small number of responses, the 16 sessions are combined into 4 repeats, 4 sessions each. Let B_i denote the bias of $e_L\times e_H$ against e for subject i, and B_{ij} denote the bias of repeat j from subject i. Assuming that B_i has a Gaussian distribution $N(b_i,\sigma)$, where b_i is the mean of B_i, we can compare the mean of the various listeners by testing the hypothesis that they all have the same bias, against the general alternative that they are not all the same. If no two listeners are significantly different, we may conclude that the conclusion based on the average normal listeners is applicable to any individual listeners.

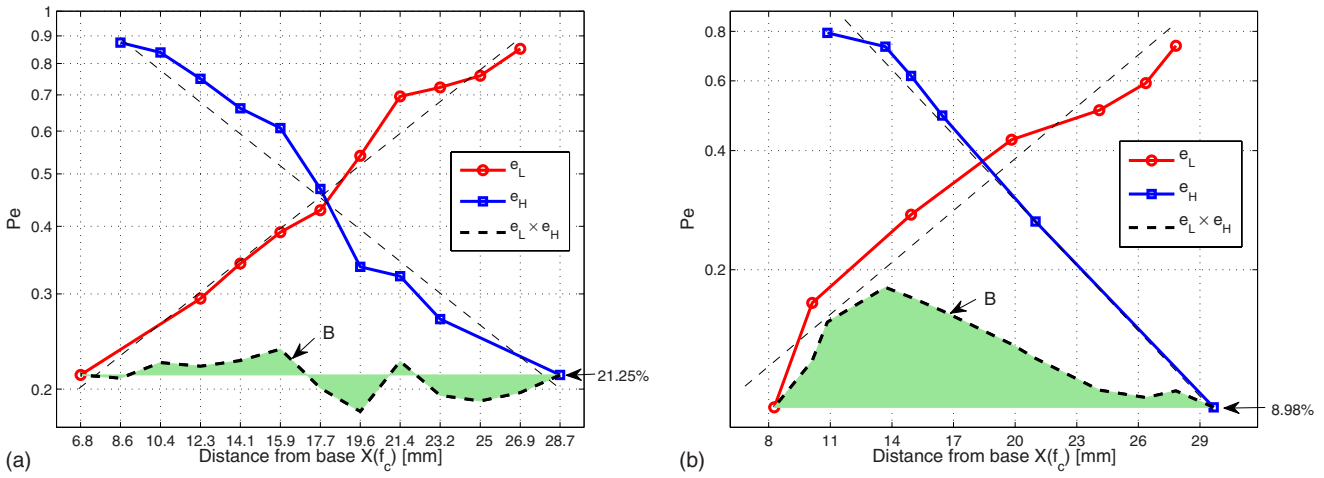


FIG. 2. (Color online) Grand probability of error and the average bias $B=e-e_L\times e_H$ for 16 consonants as a function of cut-off frequency. (a) shows the average low-pass error e_L (circles), the average high-pass error e_H (squares), and the product of the two, $e_L\times e_H$ (thick dashed) for experiment HL07. The full-band error e is defined as $e_L(f_c=8000\ \text{Hz})$ or $e_H(f_c=250\ \text{Hz})$. The average bias B is depicted by the shaded area. (b) shows the same data from experiment MN55, in which the full-band error e is defined as $e_L(f_c=6500\ \text{Hz})$ or $e_H(f_c=200\ \text{Hz})$. Note the log ordinate scale, which makes the figures easily read, actually magnifies the bias visually.

III. RESULTS

A. Multiband product rule for 16 consonants on average

Results indicate that the multiband product rule closely fits the recognition scores averaged over the 16 consonants. Figure 2(a) depicts the low-pass error e_L, the high-pass error e_H, and their product as a function of cut-off frequency. The full-band error e is equal to the low-pass error e_L at 8000 Hz and the high-pass error e_H at 250 Hz. The missing points of the low-pass error at 4775 and 6185 Hz, and the high-pass error at 363 and 509 Hz, are linearly interpolated from the nearest neighboring points. The average bias $B=e-e_L\times e_H$ is depicted by the shaded area. Suppose that the product rule is true, the shaded area would be zero. It is shown in Fig. 2(a) that the difference between $e_L\times e_H$ and the full-band error e is typically less than 3%, which is very close to zero.

Figure 2(b) depicts the results of experiment MN55.[24] Fletcher's product rule[1] over-predicts the full-band error over most frequencies for MN55, but still the measurements fit the model with reasonable accuracy. Since the low-pass and the high-pass conditions do not use the same set of cut-off frequencies, the low-pass error e_L and high-pass error e_H are linearly interpolated along the frequency to create the $e_L\times e_H$ curve, which introduces extra error in the prediction.

For both experiments, the intersection points of the low-pass and high-pass curves that divide the full band into two parts of equal information are about the same (1.5 kHz or 18 mm). The log low-pass error e_L and high-pass error e_H have been fitted by two straight lines that are symmetrical at the intersection point. This means the speech information is evenly distributed across frequency. A significant difference between the results of MN55 and HL07 lies in that the former has a maximum average bias B of 8.02%, which is considerably smaller than that of HL07 (21.25%). This might be due to the aforementioned coupling effect between the talkers and the listeners in experiment MN55, which makes the task relatively easier. Apart from that, the results of the two experiments are generally consistent. Due to the experimental design, experiment HL07 has a better precision (smaller bias) than experiment MN55, as we seen in Fig. 2. Therefore, in the remaining part of Sec. III, we will focus on analyzing the perceptual data of our experiment HL07.

Table I lists the average bias of the predicted score [the same data are depicted in Fig. 2(a) as the shaded area]. The results of the χ^2 tests indicate that e_L, e_H, and e are consistent with Fletcher's product rule[1] at all frequencies. An ANOVA test indicates that the difference between the 18 listeners is too small to be statistically significant at the level of 0.05. The discrepancy between the biases of any individual listeners and the overall average bias is generally less than 5%. Therefore the 18 listeners of normal hearing can be regarded as having the same bias $e-e_L\times e_H$ independent of cut-off frequencies. Thus Fletcher's product rule[1] may be applied to any individual normal hearing listener.

B. Multiband product rule for stops and fricatives

Analysis of the perceptual data indicates that the multiband product rule applies to the stops and fricatives as well. Figure 3(a) depicts the average low-pass error e_L, average high-pass error e_H, and the product of the two $e_L\times e_H$ for the six stop consonants (/pa, ka, ta, ba, ga, da/). The average bias $B=e-e_L\times e_H$, as depicted by the shaded area, is rather small. The high-pass error and the low-pass error cross each other at about 1.5 kHz, which is about the same position (18 mm)

TABLE I. The average bias of 16 consonants on average in experiment HL07 for various cut-off frequencies.

Frequency (Hz)	363	509	697	939	1250	1649	2164	2826	3678	4775	6185
$B=e-e_L\times e_H$	−1.9	−2.6	−1.8	1.3	−3.1	−1.2	2.5	1.3	0.8	1.7	0.3

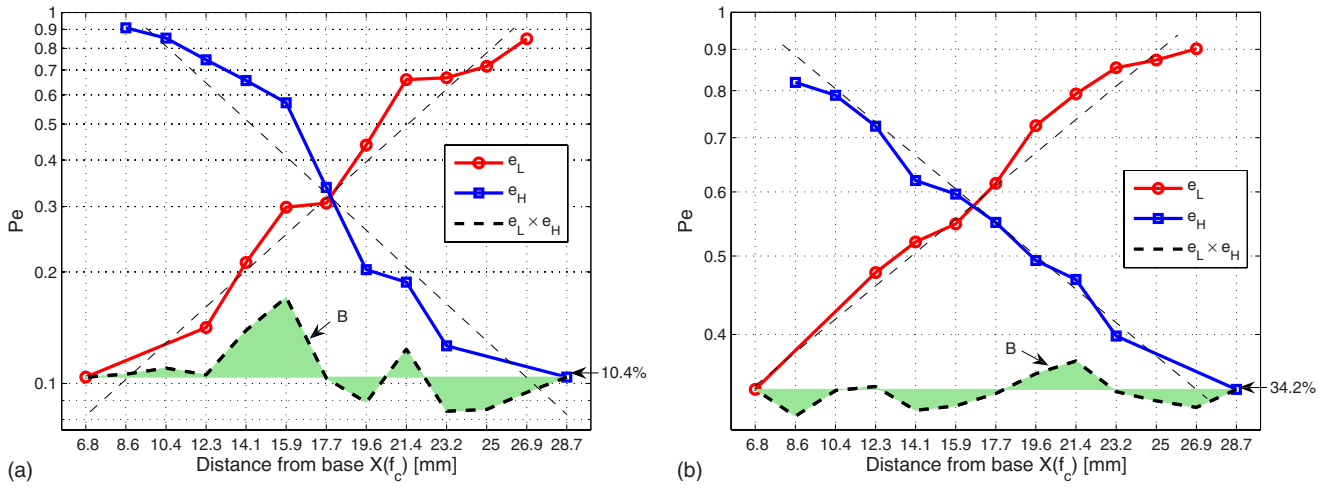


FIG. 3. (Color online) Average probability of error and the average bias $B=e-e_L\times e_H$ for stops (/pa, ka, ta, ba, ga, da/) and fricatives (/fa, θa, sa, ʃa, va, ða, za, ʒa/) as a function of cut-off frequency. (a) shows the average low-pass error e_L (circles), the grand high-pass error e_H (squares), and the product of the two, $e_L\times e_H$ (thick dashed), for stops. The average bias $B=e-e_L\times e_H$ is the shaded area. (b) shows the same results for the fricatives.

as the weight of the 16 consonants on average. The logarithms of e_L and e_H are well approximated by straight lines having complementary but identical slopes.

The results for the eight fricative consonants (/fa, θa, sa, ʃa, va, ða, za, ʒa/) are depicted in Fig. 3(b). The average bias B is almost flat with the maximum prediction error being less than 3%. Like the case of average consonants, $e_L(f_c)$ and $e_H(f_c)$ have near constant equal slopes of opposite signs when the two curves are plotted on log scales, suggesting that the fricative information is evenly distributed across the frequency range.

Table II lists the average bias B for the two sound groups at various cut-off frequencies. All values satisfy the χ^2 test at a significance level of 0.05. An ANOVA test shows no significant difference between the results of the 18 listeners.

C. Multiband product rule for individual consonants

Analysis of our HL07 data reveals that Fletcher's product rule[1] applies to the 16 consonants over limited frequencies for about 80% of the cases (CVs × frequencies). Figure 4 depicts the low-pass error e_L, high-pass error e_H, and the product of the two $e_L\times e_H$ for the 16 consonants. Based on the shape of $e_L\times e_H$, the 16 consonants can be roughly classified into flat and non-flat groups. The flat group includes /pa, ka/ and /ʃa, da, ma, na, za, ga, sa, fa, va/, for which the prediction error $e_L\times e_H-e$ is less than 5% over all frequencies, or less than 5% for most of the cut-off frequencies. The rest of the consonant sounds, /ta, ba, ʒa, θa, ða/, form the biased (non-flat) group.

Table II lists the average bias of the predicted score (the same data are depicted in Fig. 4 as the shaded area). A χ^2 test of significance level 0.05 was applied to each of the 16 consonants. A total of 136 out of 176 cases (16 CVs × 11 frequencies) statistically satisfy Fletcher's product rule[1] at a significance level of 0.05. Only two consonants /pa, ka/ passed the χ^2 test over all frequencies. Most of the unsatisfied cases come from the biased group, such as /ta, ba, ða, ʒa/, for which the fail rate is 50%.

An ANOVA test was used to investigate the listener's dependence. Since the number of tokens per CV × frequency for each listener is only 6, a number too small for a useful statistical test, the 18 listeners are ranked according to their speech recognition scores and artificially divided into three groups. The top six are attributed to the H group. The middle six are attributed to the M group. The lower six are classified as the L group. For 173 out of 176 combinations (16 CVs × 11 frequencies) ANOVA tests produce the same result that the H, M, and L groups are not significantly different in terms of the average bias per CV × frequency. In other words, the three groups of listeners are close to each other in terms of the fitness to the multiband product rule (see Table III).

The perceptual data provide important information on the perceptual cues for the initial consonants. Usually the primary cue of a consonant is located around the intersection point of e_L and e_H, which divides the full band into two parts having equal information (e.g., score). When the primary speech cue is removed, the error climbs dramatically.[28]

TABLE II. The average bias of stops and fricatives in experiment HL07 for various cut-off frequencies.

	Frequency (Hz)										
Subgroup	363	509	697	939	1250	1649	2164	2826	3678	4775	6185
Stop	−1.1	−2.0	−2.0	2.0	−1.5	−0.1	6.6	3.5	0.1	0.9	0.5
Fricatives	−2.1	−1.5	−0.2	2.9	1.5	−0.4	−1.6	−2.0	0.3	0.8	−1.5

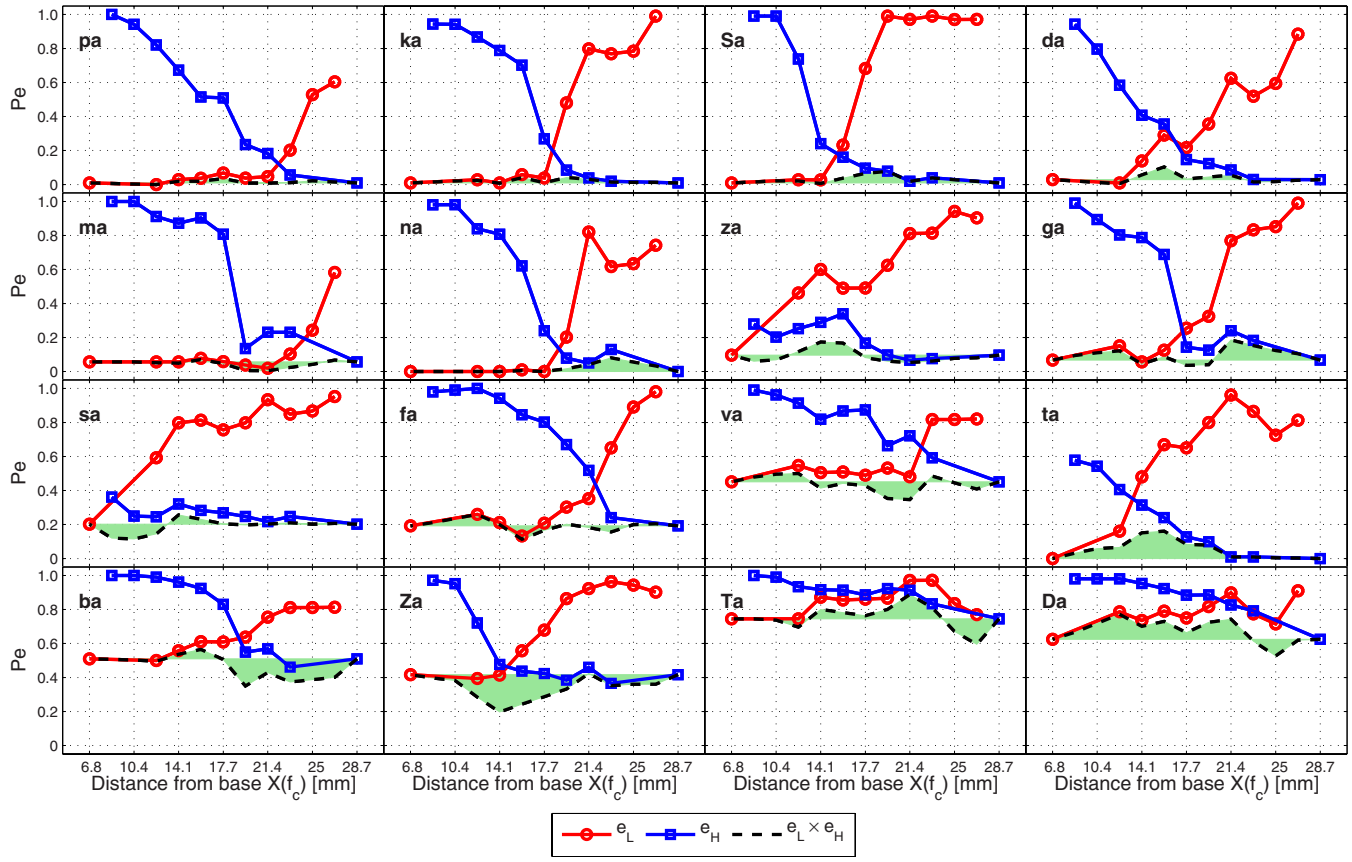


FIG. 4. (Color online) Probability of error for 16 consonants as a function of cut-off frequency. The low-pass error $e_L(f_c)$ and the high-pass error $e_H(f_c)$ are marked by circles and squares, respectively. The dashed curve depicts the product of the two $e_L \times e_H$. The full-band error e is equal to $e_L(f_c=8000$ Hz) or $e_H(f_c=250$ Hz). The bias $B(f_c)=e-e_L \times e_H$ is illustrated by the shaded area. The International Phonetic Alphabet (IPA) symbols for Ta, Sa, Da, and Za are /θa, ʃa, ðа, ʒа/, respectively.

IV. GENERAL DISCUSSION

In Sec. III A, we demonstrated that Fletcher's product rule[1] [Eq. (1)] is true for the average consonants at all cut-off frequencies. This can be regarded as a significant verification of the multiband product rule of frequency integration [Eq. (2)]. Suppose that Eq. (2) is a consequence of the fact that the frequency bands b_k, associated with e_k, are independent in terms of speech perception. A strict proof would require a

TABLE III. The average biases of 16 consonant sounds in experiment HL07 for various cut-off frequencies. Cases for which the χ^2 test was statistically significant at the 0.05 level are marked with an asterisk.

	Frequency (Hz)										
CV	363	509	697	939	1250	1649	2164	2826	3678	4775	6185
pa	0.3	1.0	0.2	−0.1	−0.1	2.5	0.9	1.0	−1.0	−0.7	−0.4
ka	0.2	0.2	0.5	2.1	3.1	0.1	3.1	−0.2	1.5	1.2	0.7
ʃa	0.7	1.7	3.0	0.9	6.8*	5.6*	2.8	−0.3	1.2	1.4	0.8
da	−0.3	−1.1	−1.3	2.5	1.5	0.3	7.5*	2.8	−2.3	−1.7	−0.9
ma	0.2	−1.8	−3.3*	−5.2	−5.1	−1.0	1.4	−0.7	−0.5	0.0	0.0
na	2.4	4.8*	8.0*	4.0*	1.6	0.0	0.6	0.0	0.0	0.0	0.0
za	−1.4	−1.7	−3.5	−4.3*	−3.5	−1.4	7.0*	7.7*	2.0	−2.2	−2.6
ga	2.8	4.7	8.6*	11.8*	−2.6	−3.0	1.9	−2.2	5.5	4.9	3.4
sa	0.1	−0.4	0.8	0.1	−0.4	0.2	2.8	5.4	−5.7	−7.9*	−7.0*
fa	0.8	0.4	−3.6	−1.0	1.0	−2.6	−7.9*	0.7	6.7	4.8	2.4
va	−5.1	−1.5	3.3	−10.4*	−9.8*	−2.3	−0.9	−3.7	4.9	5.2	3.6
ta	0.2	0.4	0.8	0.9	7.8*	8.3*	16.1*	15.1*	6.5*	6.5*	3.9*
ba	−10.5*	−11.9*	−13.6*	−8.1	−16.0*	−0.4	5.5	2.6	−1.4	−0.7	−0.4
ʒа	−5.3	−5.2	−6.5	0.7	−8.4	−12.9*	−17.3*	−22.0*	−13.2*	−3.6	−2.1
θа	−15.5*	−8.0	6.5	14.2*	5.5	1.8	3.7	5.5	−4.9	−0.7	0.0
ðа	−1.7	−10.8*	−1.1	11.8*	9.9*	3.8	10.5*	7.6	14.7*	10.6*	5.5

speech perception test that actually measures the 20 narrowband recognition scores. This is totally impractical for $K=20$, as it would require $20!=2.5\times10^{18}$ tests.

If we look at the real perceptual data [Fig. 2(a)], it actually provides much more information. The logarithms of both e_L and e_H can be closely fitted by two lines symmetrical across the intersection point of the two curves. This clearly indicates that (1) the speech information is uniformly distributed across the basilar membrane, as independently measured by both low-pass and high-pass tests; and (2) the articulation bands are additive in log error in speech perception. Similar results are observed for the two groups of stops and fricatives [Figs. 3(a) and 3(b)].

Based on the observation, it is conjectured that the multiband product rule is a combined property of the peripheral auditory system that has multiple independent parallel channels, and that the input speech stimuli are characterized by a uniform distribution of speech cues along the basilar membrane. It does not apply to individual consonants because the distribution of individual consonant speech cues is not flat. Due to the priori dependence between the speech cues, sometimes the high-pass and low-pass errors do not fit the model. For example, when the primary cue of a sound covers more than one band, the product of the low-pass and high-pass error $e_L\times e_H$ may be lower or higher than full-band error e, due to the fact that the bands neighboring the cut-off frequency are not really independent. To fully understand the interactions between the speech cues and explain why the multiband product rule fails at certain points necessitates knowledge of the speech features.[29]

V. CONCLUSION

The multiband product rule of frequency integration is an empirical formula justified by the two properties about speech and hearing, specifically, (1) the speech information is evenly distributed across the frequency, and (2) the auditory critical bands are independent in terms of speech perception. Results of our experiment HL07 show that the multiband product rule is statistically valid for consonants on average. It may also apply to subgroups of consonant sounds, such as stops and fricatives, which are characterized by a flat distribution of speech cues along the frequency. It fails for individual consonants, as expected.[30,31]

ACKNOWLEDGMENTS

The authors are grateful for thoughtful discussions with Bryce E. Lobdell, Michael Kramer, and members of the HSR group at University of Illinois, Urbana.

[1]H. Fletcher, *Speech and Hearing in Communication*, ASA edition (Acoustical Society of America, Woodbury, NY, 1995).
[2]J. B. Allen, "How do humans process and recognize speech?," IEEE Trans. Speech Audio Process. **2**, 567–577 (1994).
[3]H. Fletcher and R. Galt, "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151 (1950).
[4]C. E. Shannon, "The mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948); "The mathematical theory of communication," **27**, 623–656 (1948).
[5]American National Standard methods for the calculation of the articulation index, A.S3.5-1969 (American National Standards Institute, New York, NY, 1969).
[6]Methods for calculation of the speech intelligibility index (SII-97), A.S3.5-1997 (American National Standards Institute, New York, NY, 1997).
[7]N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119 (1947).
[8]L. L. Beranek, "The design of speech communication systems," Proc. IRE **35**, 880–890 (1947).
[9]K. D. Kryter, "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697 (1962).
[10]H. Steeneken and T. Houtgast, "A physical method for measuring speech transmission quality," J. Acoust. Soc. Am. **67**, 318–326 (1980).
[11]V. Duggirala, G. A. Studebaker, C. V. Pavlovic, and R. L. Sherbecoe, "Frequency importance functions for a feature recognition test material," J. Acoust. Soc. Am. **83**, 2372–2382 (1988).
[12]C. V. Pavlovic, "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258 (1984).
[13]C. V. Pavlovic, G. A. Studebaker, and R. L. Sherbecoe, "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals," J. Acoust. Soc. Am. **80**, 50–57 (1986).
[14]G. A. Studebaker, C. V. Pavlovic, and R. L. Sherbecoe, "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130–1138 (1987).
[15]J. B. Allen, *Articulation and Intelligibility* (Morgan and Claypool, Princeton, NJ, 2005).
[16]K. D. Kryter, "Validation of the articulation index," J. Acoust. Soc. Am. **34**, 1698–1702 (1962).
[17]K. W. Grant and L. D. Braida, "Evaluating the articulation index for auditory visual input," J. Acoust. Soc. Am. **89**, 2952–2960 (1991).
[18]R. P. Lippmann, "Accurate consonant perception without mid-frequency speech energy," IEEE Trans. Speech Audio Process. **4**, 66–69 (1996).
[19]H. Müsch and S. Buus, "Using statistical decision theory to predict speech intelligibility I. Model structure," J. Acoust. Soc. Am. **109**, 2896–2909 (2001).
[20]H. Müsch and S. Buus, "Using statistical decision theory to predict speech intelligibility II. Measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**, 2910–2920 (2001).
[21]H. Steeneken and T. Houtgast, "Mutual dependence of octave-band weights in predicting speech intelligibility," Speech Commun. **28**, 109–123 (1999).
[22]D. Ronan, A. K. Dix, P. Shah, and L. D. Braida, "Integration across frequency bands for consonant identification," J. Acoust. Soc. Am. **116**, 1749–1762 (2004).
[23]T. Y. Ching, H.Dillon, and D. Byrne, "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency," J. Acoust. Soc. Am. **103**, 1128–1140 (1998).
[24]G. A. Miller and P. E. Nicely, "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352 (1955).
[25]S. Phatak and J. Allen, "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326 (2007).
[26]D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605 (1990).
[27]B. Lobdell and J. Allen, "A model of the vu (volume-unit) meter, with speech applications," J. Acoust. Soc. Am. **121**, 279–285 (2007).
[28]M. S. Régnier and J. B. Allen, "A method to identify noise-robust perceptual features: Application for consonant /t/," J. Acoust. Soc. Am. **123**, 2801–2814 (2008).
[29]J. B. Allen, "Consonant recognition and the articulation index," J. Acoust. Soc. Am. **117**, 2212–2223 (2005).
[30]P. Heil, H. Neubauer, A. Tiefenau, and H. von Specht, "Comparison of absolute thresholds derived from an adaptive forced-choice procedure and from reaction probabilities and reaction times in a simple reaction time paradigm," J. Assoc. Res. Otolaryngol. **7**(3), 279–298 (2006).
[31]C. E. Shannon, "Communication in the presence of noise," Proc. IRE **37**, 10–21 (1949).

Acoustic profiles of distinct emotional expressions in laughter

Diana P. Szameitat[a)]
Department of Psychiatry and Psychotherapy, University of Tübingen, Osianderstrasse 24, 72076 Tübingen, Germany and Department of Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

Kai Alter
Institute of Neuroscience, Newcastle University, Framlington Place, Newcastle NE1 4HH, United Kingdom

André J. Szameitat
Department of Psychology, Ludwig Maximilians University, Leopoldstrasse 13, 80802 Munich, Germany

Dirk Wildgruber
Department of Psychiatry and Psychotherapy, University of Tübingen, Osianderstrasse 24, 72076 Tübingen, Germany

Annette Sterr
Department of Psychology, University of Surrey, Guildford GU2 7XH United Kingdom

Chris J. Darwin
Department of Psychology, University of Sussex, Brighton BN1 9QG United Kingdom

(Received 16 July 2008; revised 22 April 2009; accepted 24 April 2009)

Although listeners are able to decode the underlying emotions embedded in acoustical laughter sounds, little is known about the acoustical cues that differentiate between the emotions. This study investigated the acoustical correlates of laughter expressing four different emotions: joy, tickling, taunting, and schadenfreude. Analysis of 43 acoustic parameters showed that the four emotions could be accurately discriminated on the basis of a small parameter set. Vowel quality contributed only minimally to emotional differentiation whereas prosodic parameters were more effective. Emotions are expressed by similar prosodic parameters in both laughter and speech.
 [DOI: 10.1121/1.3139899]

PACS number(s): 43.71.Bp, 43.70.Gr, 43.72.Ar [DOS] Pages: 354–366

I. INTRODUCTION

Laughter is a prominent part of human non-verbal communication; in social interaction it is uttered in a wide variety of different situations and emotional contexts.[1,2] Moreover, while its acoustical signal is easily identifiable,[3] it is also extremely variable.[4] Such variability is not random but, amongst other things, allows listeners reliably to perceive which of a number of different emotions is being expressed.[5] However, we do not know what acoustic properties of laughter cue the different emotions. The aims of the current study are to describe the acoustical properties of laughter sounds produced under different emotions and to test for differences between them.[6]

To our knowledge, previous studies on the acoustical structure of laughter investigated laughter emitted in single behavioral contexts.[4,8,9] However, studies directly comparing different laughter types are lacking. Thus, we derived hypotheses for acoustic cues conveying emotions in laughter from studies on emotions in speech. Numerous studies have shown that emotions are not predominantly communicated via lexical information but rather via emotional prosody (for reviews see Refs. 10–12). Different emotions in speech can be reliably identified via a small set of prosodic vocal parameters[11] such as fundamental frequency (F0), standard deviation of F0, intensity, duration of voiced elements, and energy below 1000 Hz.[12] These parameters are not unique to speech: emotional expression in musical performance is based on the same vocal indicators as has been reported for emotional speech prosody.[10] In addition, there is some evidence that similar effects are seen in non-verbal utterances[13,14] such as crying or screaming and in interjections (e.g., "yippee!" and "hurray!"). Thus, communication of emotions may rely on similar acoustic parameters in these different types of utterance.

In order to investigate emotional expressions in laughter, we analyzed four different portrayals of laughter sounds. First, we decided to test joyous and taunting laughter, as both arise from basic emotions[15] which have been regularly investigated in emotional facial and vocal expression and which differ strongly from each other.[5,13] Joyful laughter is based on joy, which resembles a positive emotion for both sender and listener, and promotes social bonding. In contrast, taunting laughter (which we consider to be synonymous to sneering laughter) is based on an aggressive, destructive emotion such as contempt or scorn, which humiliates the listener and segregates members from group context.[5] The third emotion we investigated was schadenfreude (pleasure in another's misfortune), which resembles an affect blend of taunt (Ger-

a)Author to whom correspondence should be addressed. Electronic mail: d.szameitat@gmx.de

man "schaden" = English harm) and joy (German "freude"). Although schadenfreude shares features with both, joyful and taunting laughter, it can be distinguished from the latter two emotions. Schadenfreude is similar to joy in that the sender enjoys the situation which is the misfortune of the other person. However, this joy does not (in contrast to joyful laughter) promote social bonding. Furthermore, and comparable with taunting laughter, schadenfreude aims at dominating the other person.[5] However, in schadenfreude (in contrast to taunt) the sender does not want to seriously harm the listener. Thus schadenfreude shares similarities with teasing, a behavior that is also found in other social contexts such as between friends and romantic couples.[16–18] The fourth laughter type we tested was laughter provoked by tickling (hereafter named tickling laughter), which is one of the first laughter expressions in children[19] and one of the very few laughter expressions also emitted by non-human primates.[20,21] It is still a matter of debate whether tickling laughter is based on an emotion[22] or if it is merely a reflex action[23] (however, for ease of reading we will subsume it under the category of emotional laughter). Tickling laughter is characterized by a high physical activation and, like joyful laughter, promotes social relationships.[22]

In order to allow for a good acoustical differentiation, we analyzed the laughs according to the three basic perceptual dimensions of vocal sounds, i.e., frequency, tempo, and intensity.[24,25] Scherer[12] suggested that differentiation between emotions may be hampered if too few acoustical parameters are investigated. Accordingly, we investigated a broad range of parameters for each perceptual dimension. This also allowed for a better comparison of our data with previously reported acoustical data on emotional vocal expressions, as previously investigated parameter sets were heterogeneous. Furthermore, we examined parameters characterizing voice quality, such as amount of voiced energy, as they are essential for characterizing emotions in the human voice[26] and for differentiating laughs.[27] In order to investigate a possible contribution of vowel quality to the encoding of emotions in laughter, further analyses dealt with potential phonological content in laughter.

If emotions in laughter are communicated via similar parameters to those expressing emotions in speech, we would expect that joyful laughter is characterized by a high laugh rate, high F0, and high intensity, similar to joyful speech,[9,28,29] while taunting laughter is characterized by a low laugh rate, low F0, and a low intensity, similar to taunting speech.[28,30–35] For schadenfreude and tickling laughter, no hypothesis could be derived as their emotional speech prosody has not yet been investigated.

II. METHOD

A. Data collection

For the portrayals of emotional laughter eight professional actors (three male) produced four types of laughter, i.e., joyous, tickling, schadenfreude, and taunting. The speakers were instructed to put themselves into the respective emotional state with the help of self-induction techniques and to laugh freely without thinking about the expression of the laughter. Instructions included an example scenario for each emotion; however, the interpretation and expression of the emotions was left to the speakers to decide for themselves (see Ref. 36 for a similar approach).

Sound recordings, using a DAT recorder (TASCAM DA-P) with the microphone (Sanyo MP-101) approximately 0.5 m in front of the talker, took place in a sound proof booth. Recordings were digitized at a sampling rate of 48 kHz (16 bits), normalized, and cut into individual laughter sequences.

TABLE I. Number of laughter sequences per speaker and emotion. ma-mc *male speakers*, fa-fe *female speakers*, J *Joy*, Ti *Tickling*, S *Schadenfreude*, Ta *Taunt*.

Speaker	J	Ti	S	Ta	Total
ma	6	1	3	1	11
mb	4	5	1	6	16
mc	6	5	6	6	23
fa	5	3	0	2	10
fb	4	6	2	6	18
fc	0	6	3	5	14
fd	5	4	4	6	19
fe	6	2	2	6	16
Total	36	32	21	38	127

B. Stimulus material

Sequences containing verbal material, interjections, and background noise were excluded from further analysis. Furthermore, only the laughter sequences that gave good expression of the emotions in a previous study[5] were used. This study divided 429 sequences into three subsets (120–153 sequences each). Each subset was then classified according to the underlying emotion in a four-choice classification paradigm by 24 (12 male) English native subjects (mean age 22 years, total $n=72$).[5] From all correctly classified sequences (i.e., classification above chance level, $p<0.05$, two-tailed), a stimulus set was chosen which was balanced with respect to emotion, speaker sex, and speaker identity. This set consisted of 127 laughter sequences (21–38 per emotion, 0–6 per emotion and speaker, Table I) and had an average correct classification rate of 63% (for details see Table II).

TABLE II. Classification results in percent as derived by listener's classification (Ref. 5). J *Joy*, Ti *Tickling*, S *Schadenfreude*, Ta *Taunt*. Bold type represents correct classification.

		Response			
		J	Ti	S	Ta
Stimulus	J	**61**	12	21	5
	Ti	13	**68**	15	4
	S	22	11	**54**	14
	Ta	6	4	20	**70**

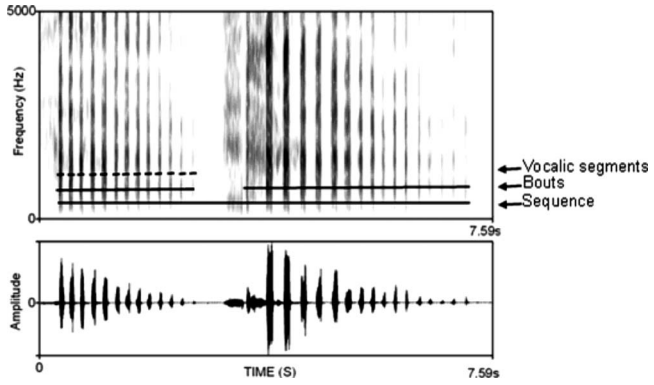


FIG. 1. Segmentation of a laughter sequence. Shown are the spectrogram (above) and oscillogram (below).

C. Acoustical analysis

The acoustic parameters were extracted using PRAAT 4.02.04.[37] Laughter sequences were segmented in the time domain according to vocalic segments (burst of energy of unvoiced and voiced exhaled breath having a single vocal peak) and bouts (either all segments from the first to the beginning of an inhaled breath or all segments between two inhaled breaths, Fig. 1). The boundaries of a segment were determined visually in the amplitude-time spectrum (distinct rise of energy from background noise into a single vocal peak) and transcribed into a script (Text-Grid function in PRAAT). On the basis of this segmentation, 43 acoustical parameters were calculated by PRAAT scripts for each individual sequence (Table III). To calculate the amplitude parameters, the values of the sounds were squared and convolved with a Gaussian window (Kaiser-20, side lobes below −190 dB, e.g., Intensity: Get mean function). Parameters of fundamental frequency were determined by an autocorrelation method [e.g., Sound: To Pitch (ac) function]. To avoid artifacts in F0 extraction, the F0 search range (pitch floor and pitch ceiling) was determined by visual inspection, i.e., by overlaying the automatically extracted pitch contours with a narrowband Fast Fourier Transform (FFT)-based spectrogram (30 ms, Gaussian window, pre-emphasis +6 dB/octave). For male speakers the F0 search range was always 75–600 Hz. For female speakers the F0 search range was highly variable; although it predominantly had an average range of 120–1000 Hz, the pitch ceiling could be as high as 2000 Hz. Formants were extracted by linear predictive coding [Gaussian-like window, Formant (burg) function],[38,39] a short-term spectral analysis approximating the spectrum of each analysis frame by five formants. The ceiling of the formant search range for the first five formants was 5000 Hz for male speakers and 5500 Hz for female speakers, respectively. For vocalic segments with ambiguous outcome in the automatic formant extraction, formant-peak locations were examined by visual inspection on a random basis. For this, the automatically detected formant bands were overlaid with a broadband FFT-based spectrogram (5 ms, Gaussian window, pre-emphasis +6 dB/octave). The harmonic-to-noise ratio (HNR) was calculated by a short-term HNR analysis performing an acoustic periodicity detection on the basis of a forward cross-correlation analysis [Harmonicity (cc) function] with a time resolution of 10 ms. The parameters center of gravity (CoG), kurtosis, and skewness were calculated on the basis of the averaged spectrum [Spectrum (fft) function].

For calculation of parameters based on vocalic segments (segment parameters, see Table III) acoustical measurements from laughter segments that were produced with a closed mouth, or where spectral measurement extraction was uncertain were excluded leaving 3947 (125) of the original 4238 (127) laughter segments (sequences) for analysis.

D. Statistical analysis

1. Parameter-wise analysis

To test if individual acoustical parameters differed between the emotions, individual analyses of variance (ANOVAs) were calculated for each of the 43 acoustical parameters.

In detail, for parameters based on laughter sequences (sequence parameters, see Table III) some parameters were averaged across bouts (averaged: N_Sg_Bt, BtDur, IntBtDur; not averaged: TotDur, N_Sg, N_Bt, LgRate). Next, individual two-factorial ANOVAs [*emotion* (4) × *speaker sex* (2), Bonferroni-corrected for 43 comparisons: overall $p<0.05$, i.e., individual alpha level=0.0012] were carried out. Additionally, pairwise comparisons between all four emotions were calculated for each acoustical parameter showing a significant effect of emotion using Tukey's HSD tests (corrected for six comparisons).

For the evaluation of the segment parameters (see Table III) careful consideration of the acoustical properties of the laughter signal is necessary in order to avoid artifacts in the statistical analysis. For instance, the average number of vocalic segments in the sequence differed significantly between emotions [one factorial ANOVA, $F(3,117)=3.731$; $p<0.05$]. In addition, for 20 of the segment parameters the factor *segment position* was significant (one factorial ANOVAs, all $p<0.05$, not corrected for multiple comparisons), indicating that many parameters change along the course of the laughter sequence. These two effects together might lead to artifacts in the statistical analysis. For example, two types of laughter may show a statistically significant difference with respect to the mean (averaged across segments) of a parameter that has a gradient of continually decreasing values along the laughter sequence (such as F0), although the true gradients of both laughter types are identical and the laughter types differ solely in the number of segments per bout.

In the same way, testing whether parameters change along the segments of bouts is complicated by the fact that the first segment was significantly longer than all following segments [mean duration first segment=129 ms, second segment=102 ms, Tukey-HSD contrasts for one factorial ANOVA, factor *segment position* (6), segments 1 vs 2, $p<0.001$; for all other combinations of segments 2–6, not significant] and 32 segment parameters correlated significantly with segment duration (Pearson's correlation coefficient, two-tailed, $n=1058-3932$, all $p<0.05$). Changes in a parameter with segment number may arise simply because the first segment is longer, and the parameter changes with

TABLE III. Investigated acoustical parameters. Parameters marked with (+) were subjected to the discriminant analysis.

Parameter	Abbreviation	Unit	Description
Sequence level			
Number of vocalic segments	N_Sg		Number of segments
Number of bouts	N_Bt		Number of bouts (separated by inbreath)
Segments per bout	N_Sg_Bt		Average number of segments in bout
Total duration	TotDur	ms	Duration from onset to end of sequence
Bout duration(+)	BtDur	ms	Average duration of laughter bouts
Inter bout duration	IntBtDur	ms	Average duration between bouts
Laugh rate(+)	LgRate	1/s	Average number of segments per second
Segment level			
Duration			
Segment duration(+)	SgDur	ms	Average duration of a segment
Inter segment duration	IntSgDur	ms	Average duration between the end of a segment to the start of the following segment within a bout
Event duration	EvntDur	ms	Average duration between the start of two consecutive segments within a bout, (SgDur+IntSgDur)
Amplitude (Amp)			
Amplitude ratio	AmpMN_Max		Ratio of mean intensity to maximal intensity; (mean Amp./maximal Amp.)
Amplitude bandwidth(+)	AmpBW	dB	Difference between maximal intensity and minimal intensity, (maximal Amp.−minimal Amp.)
Amplitude SD ratio	AmpSD_MN		Ratio of intensity standard deviation to mean intensity, (Amp. SD/mean Amp.)
Time of max. amplitude	tiAmpMax	ms	Relative position of max. Amp. measured from voice onset of segment
Fundamental frequency (F0)			
Mean F0(+)	F0MN	Hz	Average fundamental frequency measured across time segments (i).
Minimal F0	F0Min	Hz	F0Min=Minimum ($F0_i: 1 \leq i \leq N$)
Maximal F0	F0Max	Hz	F0Max=Maximum ($F0_i: 1 \leq i \leq N$)
F0 bandwidth	F0 BW	Hz	F0BW=F0Max−F0Min
F0 start	F0Start	Hz	$F0_i=1$
F0 end	F0End	Hz	$F0_i=N$
F0 change	F0Chg	Hz	F0Chg=F0End−F0Start
Time of max F0	tiF0Max	ms	Relative position of max. F0 measured from voice onset of segment
Formants			
F1(+), F2(+), F3, F4, F5	F1–F5	Hz	First to fifth formant
F1 bandwidth	BwF1	Hz	Bandwidth of first formant
Peak frequency (PF)			
Mean PF	PFMN	Hz	Average peak frequency measured across time segments (i).
Maximal PF(+)	PFMax	Hz	PFMax=Maximum ($PF_i: 1 \leq i \leq N$)
Ratio mean PF/mean F0	PFMN_F0MN		Ratio mean PF to mean F0
Ratio max PF/mean F0(+)	PFMax_F0MN		Ratio maximal PF to mean F0
Time of max. PF	tiPFMax	ms	Relative position of max. PF measured from voice onset of segment
Voice parameters			
Ratio of voiced elements(+)	% voic	%	Percent of time segments which had a clear harmonic structure
Mean harmonic-to-noise ratio (HNR)(+)	HNRMN		Average HNR
HNR SD	HNRSD		Standard deviation of HNR

TABLE III. (*Continued.*)

Parameter	Abbreviation	Unit	Description
Maximal HNR	HNRMax		Peak HNR
Time of max HNR	tiHNRMax	ms	Relative position of max. HNR measured from voice onset of segment
Jitter	Jitt	%	Measure for micro irregularities in F0
Shimmer	Shim	%	Measure for micro irregularities in amplitude of F0
Center of gravity$^{(+)}$	CoG	Hz	Frequency at which the energy of the signal is divided into half. Measure for the average height of the frequencies in the segment.
Skewness	Skew		Normalized skewness is the third central moment divided by the 1.5 power of the second central moment. Measure for how much the shape of the spectrum below the CoG is different from the shape above the CoG.
Kurtosis	Kurt		Normalized kurtosis is the fourth central moment divided by the square of the second central moment. Measure for how much the shape of the spectrum around the CoG is different from a Gaussian curve.

segment duration rather than with segment number. This problem also prevents us from saying whether such changes differ across emotions.

Different segment positions also had different sample sizes, whereby the sample size decreased with increasing segment position, with the exception of the first segment which had a smaller sample size than the second segment. A smaller sample size, however, might result in a less accurate estimate of the mean. For the examination of the segment parameters only segments with a sample size of at least 50% of the second segment were examined, which was true for all segments up to the eighth segment. Furthermore, due to the above mentioned particularities, the first segment was excluded from the analysis.

To test whether the average value of segment parameters differed between the emotions, the parameter values for segments 2–8 were first each averaged across bouts. These seven averaged values were then themselves averaged across segments resulting in one data point per sequence for each acoustical parameter. Individual two-factorial ANOVAs were carried out on these values [*emotion* (4) × *speaker sex* (2), Bonferroni-corrected for 43 comparisons] for each parameter. Furthermore, for each parameter pairwise comparisons of the emotions were conducted using Tukey's HSD test (corrected for six comparisons).

2. Variation of parameters along bouts

To test for parameter changes during the segments of a bout, the values for each of segments 2–8 were separately averaged across bouts, so that for each laughter sequence there was one data point for each of segments 2–8. Individual three-factorial ANOVAs [*emotion* (4) × *speaker sex* (2) × *segment position* (7)] were then carried out and the factor *segment position* was examined for significance (Bonferroni-corrected for 36 comparisons). To test if emotions differ in the change of parameters along the bouts, we examined, in a second step, the interaction *segment position* × *emotion* (Bonferroni-corrected for 36 comparisons). To understand potential interactions more thoroughly, we calculated, separately for each parameter, all pairwise combinations of emotions in separate ANOVAs [*emotion* (2) × *segment position* (7)]. Finally, to test for the direction of potential parameter changes along the bouts, we calculated a linear regression for each parameter and emotion.

3. Analysis of the first segment

The above statistical analysis used only the second to eighth segments. To test whether the first segment contains further information for differentiating between emotions beyond the one provided by segments 2–8 further analysis was made to test differences between the first and second segments. Parameter values for segments 1 and 2 were separately averaged across bouts and individual three factorial ANOVAs performed [*emotion* (4) × *speaker sex* (2) × *segment position* (2), Bonferroni-corrected for 36 comparisons]. A significant interaction between the factors *emotion* and *segment position* would indicate that differentiation of emotions depends on the segment. Further analysis will be conducted for such parameters to test whether the first segment provides information beyond the one carried by the second segment.

4. Identification of emotions

To test how well different emotions can be identified, a subset of acoustical parameters was subjected to a discriminant analysis (Table III). Parameters were chosen according to the following criteria: First, at least one parameter was chosen from each parameter domain [domains: (1) sequence parameter in general, on the segment level: (2) duration, (3) amplitude, (4) fundamental frequency, (5) formants, (6) peak frequency, (7) voice parameters, see Table III]. Second, only parameters showing significant differences between the emotions (individual two-factorial [*emotion* (4) × *speaker sex* (2)] ANOVAs, $p < 0.05$, Bonferroni-corrected for 43 comparisons) were selected, with the exception of the parameter bout duration, which was included since it missed the significance level only by a small margin ($p = 0.0013$ instead of the required $p < 0.0012$ for $p < 0.05$, Bonferroni-corrected

for 43 comparisons). Finally, we predominantly chose parameters which did not correlate with any other parameter. However, following Hammerschmidt and Jürgens,[26] we retained some correlated parameters which both theoretical considerations and empirical findings deemed important for characterizing prosodic structure. To assess the discriminative power of each individual parameter, we additionally calculated 12 separate discriminant analysis, one for each parameter.

5. Vowel quality

To identify the vowel quality of vocalic segments, F1-F2 plots were generated and compared with the standard vowel space representation according to Hillenbrand *et al.*[40] To examine if emotions are characterized by specific vowels, F1-F2 plots were compared with emotion recognition rates for each talker.

III. RESULTS

A. Differentiation of individual parameters

To examine the acoustical correlates of laughter sounds expressing different emotions, we first tested whether individual acoustical parameters differed between the emotions by conducting 43 individual two-factorial ANOVAs [*emotion* (4) × *speaker sex* (2)]. This analysis revealed that 26 out of 43 investigated parameters differed significantly between the four emotions (all $p<0.05$, Bonferroni-corrected, $F(42)=5.885-50.734$, Table IV). For sequences, the parameters number of bouts (N_Bt), temporal distance between bouts (IntBtDur), and laugh rate (LgRate) differed. For segments, two duration parameters (SgDur, EvntDur), many amplitude parameters (AmpBW, AmpSD_MN, tiAmpMax), most F0 parameters (F0MN, F0Min, F0Max, F0BW, F0Start, F0End), the first and second formants (F1, F2), all peak frequency parameters (PFMW, PFMax, PFMW_F0, PFMax_F0, tiPFMax), % of voiced elements, mean HNR, CoG, skewness, and kurtosis differed significantly between the emotions. Thus, the different laughter types clearly had different acoustical properties.

Additional analyses revealed that 21 acoustical parameters showed differences between male and female speakers (factor *speaker sex*, all $p<0.05$). The laughter of female speakers had higher frequencies (F1-F5, CoG, all F0 and PF parameters with the exception of F0Chg, tiPFMax), was more regular and more voiced (jitter, shimmer, HNR, % voiced elements), and the time of F0max measured from voice onset was longer (tiF0max). Moreover, six of the acoustical parameters showing differences between the emotions had a significant interaction between the factors *emotion* and *speaker sex* (EvntDur, F0MN, F0Min, F0Max, F0BW, F0Start, all $p<0.05$): male and female speakers thus modulated some parameters differently.

B. Differentiation of changing patterns of individual parameters

There was significant change along the course of the bout for 15 of the 36 segment parameters (three factorial ANOVAs [*emotion* (4) × *speaker sex* (2) × *segment position* (7)], factor *segment position*, all $p<0.05$, Bonferroni-corrected). The segment duration, many F0 parameters (F0MN, F0Min, F0Max, F0BW), some voice parameters (%voic, HNRMW, HNRSD), and one amplitude parameter (AmpMN_Max) decreased along bouts, while the ratio between PF and F0 (PFMW_F0, PFMax_F0), jitter and shimmer, and two amplitude parameters (AmpBW, AmpSD_MN) increased along bouts. However, only one parameter (PFMax_F0) showed a different pattern of change depending on the emotion (interaction *segment position* × *emotion*, $p<0.05$). This interaction was due to PFMax_F0 increasing more with increasing segment position in taunt than in joy or tickling laughter [individual three-factorial ANOVAs (*emotion* (2) × *speaker sex* (2) × *segment position* (7)], interaction *emotion* (taunt vs joy or taunt vs tickling, respectively) × *segment position*, $p<0.05$; linear regressions (all $p<0.05$): PFMax_F0: β taunt=0.32, β joy=0.10, β tickling =0.22). These results indicate that the pattern of parameter changes along the bout contributes only minimally to the differentiation of emotions.

C. The first segment

To test whether the first segment provides further information for acoustical differentiation beyond the one derived from the analysis of segments 2–8, we tested in individual three-factorial ANOVAs [*emotion* (4) × *speaker sex* (2) × *segment position* (2)] if the first and second segments (averaged across bouts) differed acoustically. A significant interaction between the factors segment and emotion was evident only for two acoustical parameters (both $p<0.05$, Bonferroni-corrected), i.e., % of voiced elements (%voic) and CoG. In detail, in joyous laughter the percentage of voiced elements was lower in the first than in the second segment, while there were no differences between the first and second segments for tickling, taunt, and schadenfreude. The CoG showed the opposite pattern for joy, since the first segment had higher values than the 2nd segment, while the 1st and 2nd segment did not differ for tickling, taunt, and schadenfreude. However, visual inspection of this pattern indicated that the differences between the emotions were larger in the second segment as compared to the first segment. Therefore, we suggest that the first segment adds only little additional information for the differentiation of emotions expressed in laughter.

D. Identification of emotions

To test how well different emotions can be identified, a discriminant analysis was conducted on the basis of a reduced parameter set. Acoustical parameters where chosen according to the following criteria: parameters which (1) described different acoustical cues, (2) differed significantly and strongly (high p-value) between the emotions, and (3) showed little correlation (for details see Sec. II D 4). The resulting parameter set consisted of the following 12 acoustical parameters: F0, F1, F2, SgDur, MaxPF_F0, MaxPF, AmpBW, %voic, HNRMN, CoG, BtDur, and LgRate (Table III). We found that the emotional category of the laughter

TABLE IV. Mean values for the four types of laughter and results of statistical tests. Pairwise t-tests were calculated for all combinations of laughter type [e.g., J-Ti *pairwise t-test joy vs tickling*, left arrows (<) joy significantly smaller than tickling, right arrows (>) joy significantly higher than tickling; all other comparisons equivalent]. (<, >) $p<0.05$, (<<, >>) $p<0.01$, (<<<, >>>) $p<0.001$. Gender effects : F *females*, M *males*. Abbreviations. Sex *speaker sex*, F *female speakers*, M *male speakers*, J *Joy*, Ti *Tickling*, S *Schadenfreude*, Ta *Taunt*. For further abbreviations and units of acoustical parameters see Table III.

		Means					t-tests					
Parameter	Sex	J	Ti	S	Ta	Total	J-Ti	J-S	J-Ta	Ti-S	Ta-Ti	Ta-S
Sequence level												
NrSg	F	32.5	30.7	33.9	30.3	31.5						
	M	31.7	42.2	33.9	38.8	36.2						
NrBt	F	3.0	4.3	3.5	3.3	3.5	<<<			>>	<<	
	M	2.8	4.6	2.8	3.4	3.4						
NrSg_Bt	F	13.1	7.6	10.5	9.5	10.1						
	M	12.5	11.1	12.3	11.3	11.8						
TotDur	F	7940	6749	7685	7376	7404						
	M	7540	8826	9029	8778	8436						
BtDur	F	2644	1390	2034	1945	1996						
	M	2481	1976	3018	2291	2431						
IntBtDur	F	698	329	439	515	498	>>>	>>>	>>>	<<	>	
	M	783	419	628	474	590						
LgRate	F	4.08	4.60	4.38	4.07	4.26	<<<			>>	<<	
	M	4.20	4.87	3.77	4.33	4.29						
Segment level												
Duration												
SgDur	F	88	82	90	109	94			<<	<<	>>>	
	M	90	85	116	101	97						
IntSgDur	F	114	105	112	107	109						
	M	123	100	144	113	120						
EvntDur	F	202	189	204	217	204					f>>>	
	M	214	185	259	214	217		m<		m<<<		m<
Intestity												
AmpMN_Max	F	0.928	0.913	0.912	0.898	0.912						
	M	0.918	0.922	0.907	0.914	0.916						
AmpBW	F	0.250	0.305	0.299	0.369	0.311		<	<<<		>>	
	M	0.266	0.251	0.310	0.291	0.278						
AmpSD_MN	F	0.081	0.099	0.100	0.120	0.101		<	<<<		>>	
	M	0.093	0.090	0.106	0.102	0.097						
tiAmpMax	F	44	42	49	60	49			<<	<	>>>	
	M	48	48	61	53	52						
Fundamental frequency												
F0MN	F	500	681	412	329	479	<<<		f>>>	f>>>	<<<	
	M	177	261	216	158	199		m<		m>		m<<
F0Min	F	431	599	366	296	421	<<<		f>>>	f>>>	<<<	
	M	154	237	189	148	178				m>		m<
F0Max	F	547	744	445	354	521	<<<		f>>>	f>>>	<<<	
	M	198	279	243	164	217		m<				m<<<
F0BW	F	117	146	79	58	100			>>>	f>>>	f<<<	
	M	44	41	55	16	39					m<<	m<<<
F0Start	F	481	713	430	331	485	<<<		f>>>	f>>>	<<<	
	M	198	268	252	157	215		m<<	m>			m<<<
F0End	F	447	604	394	294	432	<<<			>>>	<<<	
	M	144	252	178	116	177						
F0Chg	F	36	51	21	7	30						
	M	49	30	65	29	44						
tiF0Max	F	51	42	48	53	49						
	M	25	30	42	34	32						

TABLE IV. (*Continued.*)

		Means					t-tests					
Parameter	Sex	J	Ti	S	Ta	Total	J-Ti	J-S	J-Ta	Ti-S	Ta-Ti	Ta-S
Jitt	F	0.03	0.02	0.03	0.02	0.02						
	M	0.05	0.04	0.03	0.03	0.04						
Shim	F	0.12	0.13	0.15	0.14	0.13						
	M	0.24	0.21	0.19	0.18	0.21						
Formants												
F1	F	802	909	967	1052	936		<<	<<<		>>>	
	M	660	654	797	829	728						
F2	F	1654	1736	1666	1745	1707	<<<		<	>>		
	M	1462	1686	1485	1500	1526						
F3	F	2962	2907	3011	3027	2976						
	M	2666	2767	2685	2649	2688						
F4	F	3800	3757	3878	3913	3837						
	M	3523	3449	3603	3314	3471						
F5	F	4578	4661	4629	4604	4616						
	M	4205	4262	4240	4147	4211						
BwF1	F	153	172	241	155	171						
	M	192	157	164	122	161						
Peak frequency												
PFMW	F	870	1049	1077	1179	1049	<	<	<<<			
	M	540	672	822	890	713						
PFMax	F	856	1018	1195	1285	1089		<<<	<<<		>>>	
	M	649	715	917	943	791						
PFMW_F0	F	1.8	1.6	2.9	3.8	2.6		<<	<<<	<<<	>>>	>
	M	3.0	2.3	4.1	6.1	3.9						
PFMax_F0	F	1.7	1.5	3.2	4.2	2.7		f<<<	f<<<	f<<<	f>>>	f>>
	M	3.5	2.5	4.5	6.5	4.2						
tiPFMax	F	43	42	49	61	50			<<<		>>>	
	M	50	50	61	56	54						
Voice parameters												
%voic	F	87	82	74	66	77		>>	>>>	>	<<<	
	M	69	67	52	39	58						
HNRMW	F	11.2	11.4	8.7	8.3	9.9			>	>>	<<<	
	M	6.5	7.9	5.7	5.2	6.3						
HNRSD	F	4.8	5.0	4.5	5.2	4.9						
	M	4.1	4.1	4.3	4.7	4.3						
HNRMax	F	23.4	26.2	23.8	24.9	24.7						
	M	23.7	25.5	24.0	25.4	24.6						
tiHNRMax	F	44	40	48	52	46						
	M	46	47	60	45	49						
CoG	F	1163	1409	1440	1646	1427	<<	<<	<<<		>>	>
	M	804	1033	1139	1255	1034						
Skew	F	5.9	5.4	4.1	3.4	4.7			>>>		<<<	<
	M	6.0	4.9	5.7	3.2	5.0						
Kurt	F	92	79	52	30	62			>>>		<	
	M	95	53	85	33	68						

stimuli could be predicted with a high accuracy [discriminant analysis "enter-method" ("leave-one out cross validation"): mean 84% (76%), for details see Table V].

To test the discrimination power of each parameter individually, we calculated 12 separate discriminant analyses. These analyses revealed that emotions could be classified with an accuracy of 33.6%–48.0% (leave-one out cross validation) on the basis of a single parameter (Fig. 2).

E. Vowels

The vowel elements of the laughter sequences were predominantly based on central vowels characterized by middle F2 values, with vowel height varying from mid (ə) to open (a) (for details see Ref. 41).

To test whether vocalic elements contributed to emotional differentiation, first F1-F2 plots were analyzed for

TABLE V. Classification results in percent as derived by discriminant analysis. J *Joy*, Ti *Tickling*, S *Schadenfreude*, Ta *Taunt*. Bold type represents correct classification.

		Predicted			
		J	Ti	S	Ta
"enter-method"	J	**89**	3	6	3
	Ti	3	**94**	0	3
	S	24	10	**52**	14
	Ta	0	0	11	**89**
"leave-one out cross validation"	J	**81**	3	14	3
	Ti	6	**81**	3	10
	S	29	10	**43**	19
	Ta	0	0	14	**86**

each speaker individually and then compared with the speaker's individual recognition rates. F1-F2 plots for individual speakers revealed that the clusters of the vowel elements overlapped widely for most of the speakers and emotions. Furthermore, the variability in vocalic elements varied strongly with speaker identity, i.e., in four speakers the vowel elements differed between the emotions, and in three speakers the vowel elements showed virtually no difference. All speakers uttered almost exclusively central vowels (e.g., ɑ or ə), and in the rare cases where non-central vowels were expressed, recognition rates remained unchanged, which indicates that vowels were not used by the listeners to differentiate between emotions.

IV. DISCUSSION

Analysis of the expression of four different emotions in laughter revealed that they differ in a variety of acoustical parameters, and that they can be classified accurately (84%) on the basis of a small parameter set. Overall, prosodic parameters provided a good basis for classification, whereas vowel quality did not differ reliably between the emotions.

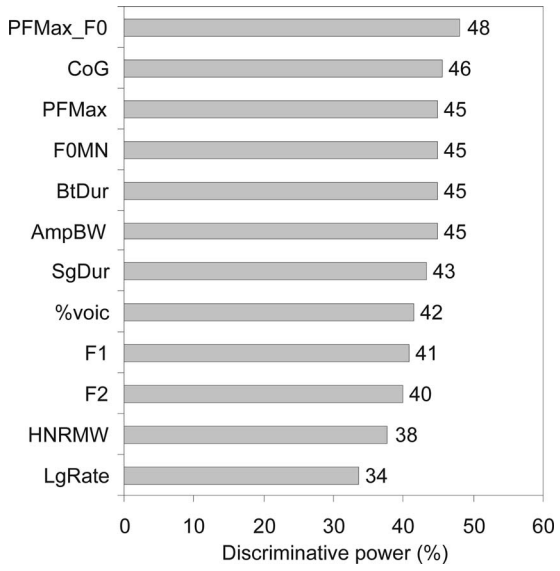


FIG. 2. Discriminative power of individual parameters. Calculated by separate discriminant analyses (leave-one out cross validation). For abbreviations of acoustical parameters, see Table III.

TABLE VI. Acoustical correlates. J *Joy*, Ti *Tickling*, S *Schadenfreude*, Ta *Taunt*; <</>> very *small/large values*; </> *small/large values*; = *middle values*; gender effect: f *females*; m *males*; bold type: *significantly different to all remaining laughter types*.

	J	Ti	S	Ta
Segment duration	=	<	>	>
Event duration	=	<	f = m **>>**	f > m =
Laugh rate	=	**>>**	=	=
Number of bouts	=	**>>**	=	=
Inter-bout duration	**>>**	**<<**	=	=
Intensity	<	<	>	>
F0	=	**>>**	=	<<
Peak frequency	<<	<	>	>>
PF/F0	<	<	=	**>>**
F1	<<	<	>	>>
F2	<	>	<	>
% voiced elements	>	>	<	<
HNR	>>	>	<	<<
Center of gravity	**<<**	=	=	**>>**
Skewness	=	=	=	**<<**
Kurtosis	=	=	=	<<

A. Prosodic characteristics of the four laughter types

Laughter sequences from the four emotions used here were associated with specific acoustical correlates (Table VI). Tickling laughter was rapid and high-pitched. Its F0 reached up to 1112 Hz for females (glottal whistles up to 1765 Hz) and up to 528 Hz for males and it had the shortest segment duration, inter-bout duration, and event duration, as well as the highest laugh rate and number of bouts. Furthermore, tickling laughter had more harmonic energy (HNR, %voic) than did schadenfreude and taunting laughter. The first formant and the peak frequency were rather low, leading in combination with the high F0 to low PF_F0 values. The second formant, on the other hand, was higher than in joyful and schadenfreude laughter, and comparable to taunting laughter. The intensity parameters were rather low.

Joyful laughter was rich in low-frequency energy and had the longest time between bouts. More specifically, it had the lowest peak frequency and first formant frequency, and its energy was the most concentrated in the lower frequency range (lowest CoG). In the time domain it stood out by having the longest temporal distance between bouts (IntBtDur). Its fundamental frequency was in the middle range, which, in combination with the low peak frequency, resulted in low PF_F0 values, which in turn were comparable to those of tickling laughter. Besides which, joyful laughter had a lot of harmonic energy (HNR, %voic), similar to tickling laughter. The second formant was rather low, i.e., lower than in tickling and taunting laughter. Also the intensity parameters were rather low, i.e., they were lower than in schadenfreude and taunting laughter.

Schadenfreude laughter did not show any outstanding characteristics, i.e., most of its parameters were in the middle

range. Specifically, schadenfreude laughter shared features with both joyful and taunting laughter (see Table V). In the time domain schadenfreude was comparable to joyful and taunting laughter. In the intensity domain, it was comparable to taunting laughter. Moreover, while the fundamental frequency and second formant were comparable to joyful laughter, the first formant and peak frequency were comparable to taunting laughter. This resulted in that the parameter PF_F0 was in the middle range, i.e., it was higher than in joyful and tickling laughter, but lower than in taunting laughter. Additionally, schadenfreude laughter had little harmonic energy (HNR, %voic), comparable to taunting laughter.

Taunting laughter had the lowest fundamental frequency, but the highest first formant and peak frequency giving the highest PF_F0 ratio. It also had the most energy concentrated in the higher frequency range (highest CoG) but the frequency distribution parameters skewness and kurtosis were lower in comparison to the remaining three laughter types. It had a small amount of harmonic energy (HNR, %voic) and a high segment duration whereby both parameters were comparable to schadenfreude laughter. Finally, its intensity parameters were higher than in joyful and tickling laugher.

B. Emotional expressions in laughter in comparison to speech

As shown in Sec. IV A, laughter sequences from the four emotions were associated with specific acoustical correlates. The question arises whether those acoustical correlates are unique for emotional expression in laughter, or whether commonalities exist to emotional expression in speech.

A number of findings support the latter hypothesis. First, the same parameters that showed reliable differences between the laughter types have also previously been reported to distinguish different emotions in speech, including F0 and PF, HNR, amplitude bandwidth, speech rate (compare laugh rate for laughter), and CoG.[26] Moreover, the acoustical correlates of joyful and taunting laughter were mainly in accordance with the theoretical predictions made for joyful and contemptuous emotional speech prosody by Scherer[11] (assuming that taunt and contempt refer to comparable emotions). Finally, the acoustic profiles for joyful and taunting laughter are very similar to the acoustic profiles of joyful and contemptuous speech prosody. (To our knowledge schadenfreude and tickling speech prosody have not been previously investigated). In detail, taunting laughter and contemptuous speech prosody were both characterized by a low mean F0[26,30–33,35] and low maximal F0, a low F0 bandwidth,[26,31] a long segment duration,[33,34,26] a long temporal distance of F0max measured from voice onset (tiF0Max),[26] a low amount of harmonic energy,[26] and both utterances were often produced with a "pressed" voice.[31] However, in contrast to contemptuous speech prosody, taunting laughter had an average instead of low laugh rate,[31,34] and the peak frequency was high instead of low.[26] Joyful laughter and joyful speech prosody were both characterized by a high F0 and F0 bandwidth.[10,11] Furthermore, both expressions showed decreased values for the first formant.[42] However, in contrast to joyful speech prosody, in joyful laughter the CoG was at low instead of middle[10] frequencies and the peak frequency was low instead of high.[26]

Taken together, most of the acoustical correlates for joy and taunt were in line with previous findings for the respective emotions when communicated via speech prosody. Differences in the findings may be caused by more fine-grained differences within the employed emotions.[12] Another possibility is that emotional communication in laughter and speech is not equivalent in all acoustical correlates.

C. Laughter portrayals in comparison to spontaneous laughter

Since the stimulus-material was based on laughter portrayals produced by professional actors the question arises whether such portrayals truly reflect spontaneously emitted laughs. With respect to speech literature, the majority of authors assumed such equivalence,[43,44] although some noted that emotional portrayals may overemphasize acoustical parameters so that they may be more intense and prototypical than spontaneous expressions.[45] However, a number of findings support the assumption of equivalence.

First, the majority of the acoustical parameters of our stimulus-material fell well within the range previously reported for spontaneously emitted laughs. For example, the reported fundamental frequency was in accordance with previous studies: the average F0 was 199 Hz for males [compared to a range of previously reported average F0 (Refs. 3, 4, 8, and 46–52) 126–424 Hz] and 476 for females [160–502 Hz (Refs. 3, 4, 8, 48, and 50–53)] respectively. Moreover, most of our temporal parameters were well within the range of previously reported data: mean segment duration was 95 ms in this study, (compared[3,48,49,51–53] to means of 60–370 ms), intersegment duration was 115 ms (compared[3,4,8,48,49,51,52] to means of 87–240 ms), mean bout duration was 2213 ms (compared[3,4,46,47,51–55] to means of 700–3970 ms), and mean laugh rate was 4.3 segments/s (compared[4,46–48,51,52,54] to means of 2.8–5.6). However, the mean number of segments per bout was 11 segments and therefore on the upper limit of previously reported data (compared[3,4,8,46,47,51,52,55] to means of 1.5–12.5). The relatively high number of segments per bout has probably been caused by the fact that speakers were asked to produce long laughter sequences (the stimulus-material was intended to be also used in another study requiring longer durations). Formant measurements were in accordance with previous findings,[4,50,51] with the exception of the first formant which was much higher than previously reported [this study: males (females) 728 (924) Hz; as compared to 535 (653) Hz,[4] 543 (559),[50] females 650 Hz, [51])]. Detailed analyses revealed that high F1 values were not due to an artifact in formant extraction, but most likely reflect extreme positions adopted by the vocal tract during laughter in combination with physiological constraints accompanying production of a "pressed" voice, as reported in Ref. 41. Finally, analysis of vowel quality of vocalic segments showed that most of the vowels were based on central vowels, with only occasional deviants, which is in accordance with previous findings.[4,48,51,52,56,57] Taken to-

gether, the majority of the acoustical parameters measured in this study were in accordance with previous findings.

Second, the specific acoustical correlates of the two laugh utterances joy and taunt showed many commonalities with the respective emotions in emotional speech prosody (see Sec. IV B). Finally, laugh portrayals and spontaneous laughs are very hard to tell apart, as assessed by listeners discrimination[58] as well as the laughter's acoustical structure.[59] However, to answer the question conclusively as to whether portrayals truly reflect spontaneously emitted laughter, an investigation of emotional expression in spontaneous laughter is needed.

D. Differentiations on the basis of vowel quality

Emotional laughter is sometimes, for example, in comic strips, illustrated with certain vowels, e.g., joyous laughter is depicted as /hahaha/, taunt as /hohoho/, tickling as /hihihi/, or schadenfreude as /hɛhɛhɛ/, which may indicate a contribution of vowel quality to the encoding of emotions in laughter. However, vowel quality contributed only minimally to the discrimination of emotions in laughter, since laughter sequences were almost exclusively based on central vowels and the rare use of non-central vowels had no significant influence on the recognition rate.

Another hypothesis relating vowel quality with emotion was suggested by Ruch and Ekman.[23] They suggested that during the production of "reflexlike" laughter the vocal tract remains in a neutral position so that such laughs are not articulated, while emotional laughter would involve supralaryngeal structures leading to a diversity in vowel elements. However, our data did not support this assumption, since tickling laughter, which could be interpreted as a reflexlike laughter type, showed the same vowel elements as schadenfreude and taunt, i.e., (ə), (a), and (ɑ) vowels. In contrast, joyful laughter tended to involve more (ə) vowels, which are characterized by a neutral vocal tract, than in the other laughter types. Therefore, it was not the reflexlike laughter type, i.e., tickling laughter, which was predominantly based on unarticulated vowels, but joyful laughter, an emotional laugh utterance.

E. Emotions in laughter in comparison to other non-verbal vocalizations

The question arises how laughter should be integrated in the framework of non-verbal vocalizations. Wundt[60] classified non-verbal emotional vocalizations into two categories. In the first category are primary affective vocalizations, which he described as relicts of a pre-language period, e.g., panic shrieks (German "naturlaute," primary interjections, raw affect bursts).[59–61] In the second category are secondary affective vocalizations, which were assimilated into language, and eventually conventionalized, e.g., "yucky!" or "hooray!" (secondary interjections, affect emblems).[60–62] Scherer[62] assumed that primary affective vocalizations are direct externalizations of motor behaviors reflecting push effects, while secondary affective vocalizations are primarily influenced by socio-cultural norms reflecting pull effects.

That non-verbal vocalizations can indeed be classified into these primary and secondary vocalizations is supported by a study of Schröder.[13] In his study some non-verbal vocalizations could be classified according to the emotions solely on the basis of their transcripts (e.g., German: "igitt," "yippie"), while others could not (e.g., yawning out of boredom). Furthermore, Dietrich *et al.*[14] showed that the transition between the two categories is continuous. Therefore, non-verbal affective vocalizations can communicate emotions via the same mechanism as that known for emotional communication via speech, i.e., lexical meaning (word content) and emotional prosody. Moreover, non-verbal vocalizations can be arranged on a continuous scale, whereby primary affective vocalizations differ merely on the basis of emotional prosody, while secondary affective vocalizations can differ in both emotional prosody and lexical meaning.[14]

The question arises where laughter should be placed on this (continuous) scale. In the present study we showed that laughter is predominantly based on central vowels and therefore is foremost not articulated. Furthermore, different emotional laughs did not differ according to a systematic variation in vowel quality, which might have been served as lexical information. Moreover, laughter is estimated to be 7 million years old,[63] and thus its existence predates the evolution of language.[23] Based on these findings, we suggest that laughter is a primary affective vocalization, whereby various emotional expressions differ foremost in emotional prosody.

F. Vocal expression of emotions

With regard to the origin of emotional speech prosody, an intriguing hypothesis has been suggested. With the development of human language intensive neuronal and physiological changes took place in order to enable the production and perception of speech.[64] As the production of language and non-verbal affect vocalizations is based on the same physiological structures, i.e., the vocal tract, it has been suggested that with the development of human speech neural structures subserving speech production have been superimposed upon already existing structures subserving the production of non-verbal affective vocalizations.[28] Accordingly, emotional prosody is assumed to predate language development and to derive from animal communication.[21,28] However, evidence supporting this theory is sparse, since only little is known about emotional prosody in animal communication.[65,66]

Interestingly, some marked features of laughter may provide tentative support for this theory. Laughter is inborn, evident by the fact that also deaf-blind born children laugh.[67] It emerges in babies at the age of 4 months, and thus long before language acquisition.[23,68] Also in phylogeny it predates language evolution,[63] and it is one of the few vocalizations not only uttered by humans but also by non-human primates.[21] Therefore, laughter seems to be a phylogenetically old communication signal dating back to our primate ancestors.

A comparison of emotional expression in laughter and speech reveals numerous striking commonalities. In both

laughter and speech emotions are expressed by similar acoustical parameters, in particular peak frequency, F0, temporal patterns, and resonance characteristics of the vocal tract (for emotional speech prosody see Ref. 26). Even more specifically, discrete emotions, such as joy and taunt, have highly comparable acoustical correlates when expressed in laughter and in speech. In line with the idea that the same emotional prosody underlies laughter and speech, behavioral studies revealed that the classification accuracy for emotional laughter[5] falls within the range reported for emotional speech prosody.[10] Additionally, the confusion matrices derived from the classification of emotions in laughter (see Tables II and V) and speech show similar patterns, and distinct emotions are characterized by similar values in arousal, valence, and dominance in laughter and speech.[5] This striking convergence strongly supports the hypothesis that emotions are communicated via the same mechanism in laughter and speech, i.e., emotional prosody.

Thus, the existence of emotional prosody in laughter, a phylogenetically old communication signal derived from animal communication, is one of the few indications based on empirical data which support the hypothesis[28] that emotional prosody is a communication system dating back prior to the evolution of language.

V. CONCLUSIONS

The present study showed that laughter sequences from the four emotions—joy, schadenfreude, taunt, and tickling—were associated with distinct acoustical correlates. Accordingly, the present study supports the hypotheses that acoustic distinction between different types of laughter exists, and that this acoustic variability is a potent tool for communicating the sender's emotional state to the listener. Crucially, we found that acoustical correlates of emotions in laughter had much in common with emotional expression in speech, supporting a common underlying mechanism for the vocal expression of emotions. The existence of emotional expression in laughter, a non-verbal signal existing long before development of human language, provides suggestive evidence that vocal emotional expression also existed long before evolution of language. That emotional modulation in laughter is primarily based on respiration and phonation rather than on articulation (i.e., vowel quality) suggests that only little supralaryngeal modeling is involved in vocal emotional expression, and this is a finding consistent with the notion that supralaryngeal structures become only centrally involved with the production of language.

ACKNOWLEDGMENTS

This work was supported by grants from the Marie Curie Foundation and the German Research Foundation (DFG AL357/1 and WI2101/2). This work was also supported by The Lena Teague Bequest. We thank Maria Bale for her helpful comments on this paper.

[1]F. Poyatos, "The many voices of laughter—A new audible-visual paralinguistic approach," Semiotica **93**, 61–81 (1993).
[2]H. Giles and G. S. Oxford, "Towards a multidimensional theory of laughter causation and its social implications," Bull. Br. Psychol. Soc. **23**, 97–105 (1970).
[3]H. Rothgänger, G. Hauser, A. C. Cappellini, and A. Guidotti, "Analysis of laughter and speech sounds in Italian and German students," Naturwiss. **85**, 394–402 (1998).
[4]J.-A. Bachorowski, M. J. Smoski, and M. J. Owren, "The acoustic features of human laughter," J. Acoust. Soc. Am. **110**, 1581–1597 (2001).
[5]D. P. Szameitat, K. Alter, A. J. Szameitat, C. J. Darwin, D. Wildgruber, S. Dietrich, and A. Sterr, "Differentiation of emotions in laughter at the behavioral level," Emotion **9**, in press (2009).
[6]When referring to the concept of emotion, we follow the definition of Scherer (Ref. 7, p. 140) who stated that an emotion is a "relatively brief episode of synchronized responses by all or most organismic subsystems to the evaluation of an external or internal event as being major significance."
[7]K. R. Scherer, "Psychological models of emotion," in *The Neuropsychology of Emotion*, edited by J. C. Borod (Oxford University Press, Oxford, 2000), pp. 137–162.
[8]J. Vettin and D. Todt, "Laughter in conversation: Features of occurrence and acoustic structure," J. Nonverb. Behav. **28**, 93–115 (2004).
[9]J. Vettin and D. Todt, "Human laughter, social play, and play vocalizations of non-human primates: An evolutionary approach," Behaviour **142**, 217–240 (2005).
[10]P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance—Different channels, same code?," Psychol. Bull. **129**, 770–814 (2003).
[11]K. R. Scherer, "Vocal affect expression: A review and a model for future research," Psychol. Bull. **99**, 143–165 (1986).
[12]K. R. Scherer, "Vocal communication of emotions: A review of research paradigms," Speech Commun. **40**, 227–256 (2003).
[13]M. Schröder, "Experimental study of affect bursts," Speech Commun. **40**, 99–116 (2003).
[14]S. Dietrich, H. Ackermann, D. P. Szameitat, and K. Alter, "Psychoacoustic studies on the processing of vocal interjections: How to disentangle lexical and prosodic information?," Prog. Brain Res. **156**, 295–302 (2006).
[15]P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique," Psychol. Bull. **115**, 268–287 (1994).
[16]D. Keltner, R. C. Young, E. A. Heerey, C. Oemig, and N. D. Monarch, "Teasing in hierarchical and intimate relations," J. Pers. Soc. Psychol. **75**, 1231–1247 (1998).
[17]A. Mooney, R. Creeser, and P. Blatchford, "Children's views on teasing and fighting in junior school," Educ. Res. **33**, 103–112 (1991).
[18]L. A. Baxter, "Forms and functions of intimate play in personal relationship," Human Communication Research **18**, 336–363 (1992).
[19]R. W. Washburn, "A study of the smiling and laughing of infants in the first year of life," Genet. Psychol. Monogr. **6**, 397–537 (1929).
[20]J. A. R. A. M. van Hooff, "A comparative approach to the phylogeny of laughter and smiling," in *Nonverbal Communication*, edited by R. A. Hinde (University Press, Cambridge, 1972), pp. 209–241.
[21]C. Darwin, *The Expression of the Emotions in Man and Animals*, 3rd ed., edited and republished by P. Ekman (Harper Collins, London, 1872).
[22]J. Panksepp and J. Burgdorf, "'Laughing' rats and the evolutionary antecedents of human joy?," Physiol. Behav. **79**, 533–547 (2003).
[23]W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotion, Qualia, and Consciousness*, edited by A. Kaszniak (Word Scientific, Tokyo, 2001), pp. 426–443.
[24]J. Pittam and K. R. Scherer, "Vocal expression and communication of emotion," in *Handbook of Emotions*, edited by M. Lewis and J. M. Haviland (Guilford, New York, 1993), pp. 185–197.
[25]K. R. Scherer, "Methods of research on vocal communication: Paradigms and parameters," in *Handbook of Methods in Nonverbal Behavior Research*, edited by K. R. Scherer and P. Ekman (Cambridge University Press, Cambridge, 1982), pp. 136–198.
[26]K. Hammerschmidt and U. Jürgens, "Acoustical correlates of affective prosody," J. Voice **21**, 531–540 (2007).
[27]J.-A. Bachorowski and M. J. Owren, "Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect," Soochow J. Math. **12**(3), 252–257 (2001).
[28]K. R. Scherer, "Speech and emotional states," in *Speech Evaluation in Psychiatry*, edited by J. K. Darby (Grune and Stratton, New York, 1981), pp. 189–220.
[29]T. Johnstone and K. R. Scherer, "Vocal communication of emotion," in *Handbook of Emotions*, edited by M. Lewis and J. M. Haviland-Jones (Guildford, New York, 2000), pp. 220–235.
[30]G. Fairbanks and W. Pronovost, "Vocal pitch during simulated emotion,"

Science **88**, 382–383 (1938).
[31]L. Anolli and R. Ciceri, "The voice of emotions: Steps to a semiosis of the vocal non-vocal communication of emotion," in *Oralité et gestualité—Interactions et comportments multimodaux dans la communication*, edited by C. Cavé, I. Guaitella, and S. Santi (L'Harmattan, Paris, 2001), pp. 175–178.
[32]R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," J. Pers. Soc. Psychol. **70**, 614–636 (1996).
[33]L. Leinonen, T. Hiltunen, I. Linnankoski, and M.-L. Laakso, "Expression of emotional-motivational connotations with a one-word utterance," J. Acoust. Soc. Am. **102**, 1853–1863 (1997).
[34]G. Fairbanks and L. Hoaglin, "An experimental study of the durational characteristics of the voice during the expression of emotion," Speech Monogr. **8**, 85–90 (1941).
[35]I. Fónagy and K. Magdics, "Emotional patterns in intonation and music," Z. Phon. Sprachwissenschaft Kommunikationsforsch. **16**, 293–326 (1963).
[36]M. Bulut and S. Narayanan, "On the robustness of overall F0-only modifications to the perception of emotions in speech," J. Acoust. Soc. Am. **123**, 4547–4558 (2008).
[37]P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," Ver. 4.02.04, http://www.praat.org (Last viewed July 16, 2008).
[38]D. G. Childers, *Modern Spectrum Analysis*, (IEEE, New York, 1978).
[39]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, (Cambridge University Press, New York, 1992).
[40]J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111 (1995).
[41]D. P. Szameitat, C. J. Darwin, A. J. Szameitat, D. Wildgruber, A. Sterr, S. Dietrich, and K. Alter, "Formant characteristics of human laughter," in Interdisciplinary Workshop on the Phonetics of Laughter (Satellite Workshop of the 16th ICPHS 2007), Saarbrücken, Germany, 2007, pp. 9–13.
[42]M. Kienast and W. F. Sendlmeier, "Acoustical analysis of spectral and temporal changes in emotional speech," in Proceedings of the ISCA Workshop on Emotion and Speech, Belfast, Northern Ireland, 2000, pp. 92–94.
[43]K. R. Scherer, "Vocal affect signalling: a comparative approach," in *Advances in the Study of Behavior*, edited by J. Rosenblatt, C. Beer, M.-C. Busnel, and P. J. B.Slater, (Academic, New York, 1985), pp. 189–244.
[44]J. R. Davitz, "A review of research concerned with facial and vocal expression of emotion," in *The Communication of Emotional Meaning*, edited by J. R. Davitz (McGraw-Hill, New York, 1964), pp. 13–29.
[45]J. A. Russell, "Is there a universal recognition of emotion from facial expression? A review of cross-cultural studies," Psychol. Bull. **115**, 102–141 (1994).
[46]D. E. Mowrer, L. L. LaPointe, and J. Case, "Analysis of five acoustic correlates of laughter," J. Nonverb. Behav. **11**, 191–199 (1987).
[47]L. L. La Pointe, D. M. Mowrer, and J. L. Case, "A comparative acoustic analysis of the laugh responses of 20- and 70-year-old males," Int. J. Aging Hum. Dev. **31**, 1–9 (1990).
[48]R. R. Provine and Y. L. Yong, "Laughter: A stereotyped human vocalization," Ethology **89**, 115–124 (1991).
[49]J. D. Boeke, "Mikroskopische phonogrammstudien (Microscopic phonogramm studies)," Pfluegers Arch. Gesamte Physiol. Menschen Tiere **50**, 297–318 (1891).
[50]P. A. Milford, "Perception of laughter and its acoustical properties," Ph.D. thesis, Pennsylvania State University, Pennsylvania, 1980.
[51]C. Bickley and S. Hunnicutt, "Acoustic analysis of laughter," in International Conference of Spoken Language Processing, 1992, pp. 927–930.
[52]M. M. Makagon, E. S. Funayama, and M. J. Owren, "An acoustic analysis of laughter produced by congenitally deaf and normally hearing college students," J. Acoust. Soc. Am. **124**, 472–483 (2008).
[53]E. E. Nwokah, H.-C. Hsu, P. Davies, and A. Fogel, "The integration of laughter and speech in vocal communication: A dynamic systems perspective," J. Speech Lang. Hear. Res. **42**, 880–894 (1999).
[54]J. D. Boeke, "Mikroskopische phonogrammstudien (Microscopic phonogramm studies)," Pfluegers Arch. Eur. J. Physiol. **76**, 497–516 (1899).
[55]S. Kipper and D. Todt, "Dynamic-acoustic variation causes differences in evaluations of laughter," Percept. Mot. Skills **96**, 799–809 (2003).
[56]M. S. Edmonson, "Notes on laughter," Anthropol. Linguist. **29**, 23–34 (1987).
[57]E. E. Nwokah, P. Davies, A. Islam, H.-C. Hsu, and A. Fogel, "Vocal affect in three-year-olds: A quantitative acoustic analysis of child laughter," J. Acoust. Soc. Am. **94**, 3076–3090 (1993).
[58]P. Ekman, "What we have learned by measuring facial behavior," in *What the Face Reveals*, edited by P. Ekman and E. L. Rosenberg (Oxford University Press, New York, 1997), pp. 469–485.
[59]J. A. Bea and P. C. Marijuán, "The informal patterns of laughter," Entropy **5**, 205–213 (2003).
[60]W. Wundt, *Völkerkunde. Eine Untersuchung der Entwickungsgesetze von Sprache, Mythos und Sitte*, Die Sprache Band 1 (*Ethnology. An Investigation of the Development of Language, Myth, and Custom*, Language, Vol. **1**) (Kröner, Leipzig, 1900).
[61]F. Kainz, *Psychologie der Sprache (Psychology of Speech)* (Ferdinand Enke, Stuttgart, 1940).
[62]K. R. Scherer, "Affect bursts," in *Emotions: Essays on Emotion Theory*, edited by S. H. M. van Goozen, N. E. van de Poll, and J. A. Sergeant (Lawrence Erlbaum, Hillsdale, NJ, 1994), pp. 161–193.
[63]C. Niemitz, "Visuelle Zeichen, Sprache und Gehirn in der Evolution des Menschen—Eine Entgegnung auf McFarland (Visual Signs, Language and the Brain in the Evolution of Humans—A Reply to McFarland)," Z. Sem. **12**, 323–336 (1990).
[64]P. Lieberman, *On the Origins of Language* (Collier Macmillan, London, 1975).
[65]U. Jürgens, "Das Trojansche Klassifikationsschema emotionaler Intonation und seine Anwendung auf Totenkopfaffenlaute (The Trojan classification scheme for emotional intonation and its application to the squirrel monkey)," Bonn. Zool. Beitr. **44**, 141–146 (1993).
[66]U. Jürgens, "Vocalization as an emotional indicator—A neuroethological study in the squirrel monkey," Behaviour **69**, 88–117 (1979).
[67]I. Eibl-Eibesfeldt, *Ethology: The Biology of Behavior*, (Holt, Rinehart and Winston, New York, 1970).
[68]E. E. Nwokah, H.-C. Hsu, O. Dobrowolska, and A. Fogel, "The development of laughter in mother-infant communication: Timing parameters and temporal sequences," Infant Behav. Dev. **17**, 23–35 (1994).

Cross-language differences in cue use for speech segmentation

Michael D. Tyler[a)]
MARCS Auditory Laboratories and School of Psychology, University of Western Sydney, Locked Bag 1797, Penrith South DC, New South Wales 1797, Australia

Anne Cutler
Max Planck Institute for Psycholinguistics, Nijmegen 6500 AH, The Netherlands and MARCS Auditory Laboratories, University of Western Sydney, Locked Bag 1797, Penrith South DC, New South Wales 1797, Australia

(Received 17 September 2008; revised 24 March 2009; accepted 13 April 2009)

Two artificial-language learning experiments directly compared English, French, and Dutch listeners' use of suprasegmental cues for continuous-speech segmentation. In both experiments, listeners heard unbroken sequences of consonant-vowel syllables, composed of recurring three- and four-syllable "words." These words were demarcated by (a) no cue other than transitional probabilities induced by their recurrence, (b) a consistent left-edge cue, or (c) a consistent right-edge cue. Experiment 1 examined a vowel lengthening cue. All three listener groups benefited from this cue in right-edge position; none benefited from it in left-edge position. Experiment 2 examined a pitch-movement cue. English listeners used this cue in left-edge position, French listeners used it in right-edge position, and Dutch listeners used it in both positions. These findings are interpreted as evidence of both language-universal and language-specific effects. Final lengthening is a language-universal effect expressing a more general (non-linguistic) mechanism. Pitch movement expresses prominence which has characteristically different placements across languages: typically at right edges in French, but at left edges in English and Dutch. Finally, stress realization in English versus Dutch encourages greater attention to suprasegmental variation by Dutch than by English listeners, allowing Dutch listeners to benefit from an informative pitch-movement cue even in an uncharacteristic position. [DOI: 10.1121/1.3129127]

PACS number(s): 43.71.Hw, 43.71.Sy, 43.71.Es [AJ] Pages: 367–376

I. INTRODUCTION

Listening to continuous speech is easy in the native language and often difficult in a foreign language, and one of the reasons for this is that segmenting a continuous-speech stream into its component words encourages language-specific solutions. Among the sources of information which can help locate word boundaries are phonotactic sequencing constraints (e.g., the sequence /mg/ cannot be syllable-internal, but must contain a boundary, as in *some good*). Listeners make use of such constraints to segment speech (McQueen, 1998). In Finnish, which has word-level vowel harmony, two successive syllables containing vowels from different harmony classes must belong to different words; listeners make use of this knowledge in segmentation too (Suomi *et al.*, 1997; Vroomen *et al.*, 1998). English and Dutch are languages with variable lexical stress, but in both languages there is a strong statistical tendency for stress to fall word-initially; this too is effectively exploited by listeners in these languages (Cutler and Norris, 1988; Vroomen *et al.*, 1998).

Each of these factors is clearly language-specific. Stress placement is not a relevant factor for the many languages without stress, vowel harmony match is irrelevant in languages without vowel harmony, and though some phoneme sequence constraints (e.g., /mg/) hold across languages, many are language-specific. Sequences that cannot be syllable-internal in English (and hence must contain a boundary) may occur syllable-internally in other languages (e.g., /kv/ in German, /mr/ in Czech), and acceptable syllable-internal sequences in English may force a boundary in other languages (e.g., /ld/ as in *cold, build* would contain a boundary in German or Dutch or other languages with obligatory syllable-final obstruent devoicing). Thus the ease with which listeners segment continuous speech in their native language is in part based on efficient exploitation of the probabilities specific to that language.

Conversely, the difficulty of segmenting speech in a non-native language is in part based on unfamiliarity with that language's probabilities, and, worse, application of segmentation procedures encouraged by the native language to input for which they are inappropriate. Highly proficient German listeners to English draw on German phonotactic constraints in segmenting English (Weber and Cutler, 2006). French listeners, who can use a syllable-based segmentation procedure effectively with their native language, apply the same procedure to input in English (Cutler *et al.*, 1986) and in Japanese (Otake *et al.*, 1993), although syllabic segmentation is not used by native speakers of either of these languages. Likewise, Japanese listeners, whose native language encourages a mora-based segmentation procedure, also apply

[a)]Author to whom correspondence should be addressed. Electronic-mail: m.tyler@uws.edu.au

that procedure to input in English (Cutler and Otake, 1994) and in French (Otake *et al.*, 1996), although, again, native speakers of neither language do this.

The studies showing that listeners make use of language-specific probabilities in speech segmentation have mostly used the word-spotting task (McQueen, 1998; van der Lugt, 2001; Vroomen *et al.*, 1998; Weber and Cutler, 2006). But with word-spotting, which exploits knowledge of a vocabulary, the same input cannot be presented to different language groups. This can however be achieved with artificial-language learning (ALL) techniques. In ALL studies, listeners are typically exposed for minutes on end to a continuous stream of speech made up of novel (but phonotactically acceptable) "words," and tested post-exposure on their recognition of the recurring constituent components. For instance, they might hear *pabikutibudogolatudaropitibudopabikudaropigolatu*, containing the recurring trisyllables *daropi, pabiku, golatu*, and *tibudo*; successful segmentation would enable listeners to accept these items as the words, and reject *kutibu, dogola*, and any other sequence which only occurred through juxtaposition of recurring items.

Listeners can perform this task using only the information in the transitional probability between syllables of the input (Saffran *et al.*, 1996b), and the resulting learning has been shown to generalize beyond the exposure materials (Mirman *et al.*, 2008). If the exposure materials contain useful phonetic cues, such as tiny pauses between the constituent items (Toro *et al.*, 2008), prosodic contours grouping the syllables into words (Vroomen *et al.*, 1998), or vowel harmony within words (Vroomen *et al.*, 1998, for Finnish listeners), then listeners can use these cues too. The ALL task has the further advantage that it can be used with populations without a lexicon, such as prelinguistic infants, where it has yielded valuable insights into the statistical learning capacities of young language learners (Saffran *et al.*, 1996a; Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003).

This independence of lexical knowledge makes ALL studies also suitable for direct comparisons across languages. If the component items are made up of phonemes with a high cross-language frequency of occurrence, the input can be virtually language-neutral. Although ALL research has been a real growth area in recent years, and, in particular, many infant/adult comparisons have been undertaken, the primary focus has been the use of syllable-to-syllable transitional probability (TP) rather than of the language-specific information (phonotactic constraints, rhythmic structure), which, as described above, has been shown to characterize speech segmentation by adult listeners. Presumably in consequence of this, there have been remarkably few direct cross-language comparisons with ALL techniques.

This is somewhat surprising given that ALL techniques allow manipulation of segmentation cues. Vroomen *et al.* (1998) accompanied their word-spotting study on Finnish with an ALL study involving Finnish, French, and Dutch listeners, in which they manipulated two cues: vowel harmony and a cue they called stress that was realized as a fundamental frequency (f0) contour rising across the first syllable of a trisyllabic item, and then gradually decreasing to a baseline across the second and third syllables. They found differences in the use of these two cues across the three groups: Finnish listeners made use of the vowel harmony cue while the other two groups did not, and Finnish and Dutch listeners used the "stress" cue while French listeners did not. Both patterns were in agreement with the phonological facts: Finnish has vowel harmony while the other languages do not, and Finnish is a fixed stress language, Dutch is a free stress language, while French is not a stress language.

The currently available results suggest that listeners can use any segmentation cue of which they have had experience in their native language, from the universally computable cue of TP through any aspect of language-particular structure, but that they may not make use of cues with which they are unfamiliar. However, many more dimensions of this account remain to be explored. In infancy, for instance, there is evidence that TP allows discovery of the language-specific cues that will be most effective for acquiring the native vocabulary, after which the language-specific cues are used in preference to TP (Johnson and Jusczyk, 2001; Johnson and Seidl, 2009; Thiessen and Saffran, 2003). For adults, we have as yet little information about the relative strength of alternative cues; we do not know whether TP is the only type of information that is universally computable while all other cues are language-specific, or whether there are also cues which will prove to be universal; finally, we do not know whether a given cue is used in same way by all listeners who make use of it.

A prediction may easily be made concerning a candidate for a universal non-TP cue: final lengthening. It has been known for at least a century that regular sounds varying in duration tend to be heard as forming iambic sequences, that is, with the longer elements in final position (Woodrow, 1909). For linguistics, Hayes (1995) formulated the iambic/trochaic law, whereby intensity contrast produces trochaic grouping while durational contrast produces iambic grouping. Bolinger (1978) claimed that there are just two prosodic universals: the signaling of prominence and the signaling of juncture, and for the latter, Vaissière (1983) in a cross-language survey proposed that pre-boundary final lengthening is linguistically universal. Testing the iambic/trochaic law in an experiment with speech (synthetic CV syllables separated by 200 ms silent intervals) and nonspeech (square waves, similarly spaced) stimuli, Hay and Diehl (2007) found that both English- and French-speaking listeners preferred to group durationally varying sequences so as to produce an iambic rhythm. Saffran *et al.* (1996b), in the only cross-position preference ALL study with adult listeners, found that English speakers benefited from final lengthening over and above TP information, but not from initial lengthening.

A prediction may also be made concerning relative sensitivity to cues. Recent evidence from English and Dutch has revealed subtle differences in the use of the acoustic correlates of lexical stress. These two closely related languages both have variable lexical stress, and the phonological determinants of Dutch and English stress placement are virtually identical (van der Hulst, 1999). However, unstressed syllables show vowel reduction far more often in English than in Dutch; as a result of this, lexico-statistical analyses reveal

that vowel quality suffices to effect distinctions between words to a greater extent in English than in Dutch, and taking suprasegmental cues to stress into account in speech recognition yields a more substantial payoff in Dutch than in English (Cutler and Pasveer, 2006). For example, *cigar* and *cigarette* have different vowels in the first syllable in most dialects of English, while the cognate words in Dutch have the same vowel; and the words *octopus* and *October* exist in both languages, but begin to differ segmentally on the fourth phoneme in English (the second vowel in English *octopus* is reduced), but only on the fifth phoneme in Dutch. Although suprasegmental cues distinguish different levels of stress in both languages, making use of these cues (in *ci-* or in *octo-*, in these examples) thus pays off more for distinguishing words in Dutch. Listeners act in accord with this, showing stronger effects of suprasegmental mismatch in word recognition in Dutch (Donselaar *et al.*, 2005) than in English (Cooper *et al.*, 2002). Indeed, in judging the source of English syllables differing only in stress level (e.g., *mus-* from *music* vs *museum*), Dutch listeners outperformed native English listeners (Cooper *et al.*, 2002), and although there were significant acoustic differences between members of these syllable pairs on all suprasegmental dimensions affected by stress, the responses of the Dutch listeners were more closely correlated with the acoustic variation than were those of the native listeners (Cutler *et al.*, 2007). English listeners' judgments of English stress are principally determined by vowel quality rather than by any suprasegmental cue (Fear *et al.*, 1995), whereas Dutch listeners' judgments of stress in the same English stimuli are more fine-grained and make better use of the suprasegmental cues (Cutler, 2009), as indeed do their stress judgments in their native language (Sluijter *et al.*, 1997). It is thus reasonable to predict that suprasegmental cues may be better exploited by Dutch than by English listeners in ALL tasks too.

In the present ALL study we use the cues most often manipulated in the one-language studies: lengthening and pitch movement. Both are suprasegmental cues familiar to all our listeners. As noted above, lengthening is associated universally with iambic structures; it is a cue to a right-edge boundary. Pitch movement is principally associated with the expression of prominence, but it exhibits no positional restrictions. We use the two cues orthogonally in right- and left-edge positions, and contrast them with no separate cue; the TP structure in all the materials is otherwise identical.

We present these stimuli to listeners from three languages: English, French, and Dutch. This comparison allows us first to examine the effects on segmentation performance of cross-language differences in preferred prosodic structure. French has more right-edge (iambic) and English and Dutch more left-edge (trochaic) boundary phenomena. In general, we therefore predict that French listeners will show greater sensitivity to cues in item-final position while English and Dutch listeners will show greater sensitivity to cues in item-initial position. Further, French has no stress while both English and Dutch have stress, and in both the latter languages stress differences have acoustic reflections in pitch movement and in duration, whereby in both, stress affects f0 more strongly than it affects duration. If the universal status of the durational cue prevails over its language-specific realizations, then a final lengthening cue would prove useful to all listeners, and to a greater extent than an initial lengthening cue (or no cue other than TP information). Initial lengthening, if it is useful as a cue, may, however, prove useful to English and Dutch listeners to a greater extent than to French listeners.

The cross-language comparison also allows us to examine the relative sensitivity of English and Dutch listeners to the cues we are manipulating, both of which are suprasegmental in nature. As described above, Dutch listeners have been shown to display greater sensitivity to suprasegmental cues to stress in their own language than English listeners do in theirs and also to be more sensitive than English listeners to the suprasegmental cues to stress which English offers. We predict that if differences appear in how the cues are used by these two listener groups with prosodically highly similar native phonologies, then the differences will be in the direction of greater exploitation of the cues we provide by the Dutch listeners than by the English.

II. EXPERIMENT 1: VOWEL LENGTHENING CUES

A. Method

1. Participants

In each of the three language groups, 24 participants were randomly assigned to each cue condition (TP-only, left-edge cue, and right-edge cue; $n=24\times9=216$). French participants were psychology students at the Université de Bourgogne, Dijon, France. All had acquired French from birth, with the exception of one participant in the TP-only condition who acquired French at the age of 3 (Arabic first language [L1]), and all but eight participants had learned some English at school. Around half of the participants had also learned Spanish. The three conditions (TP-only, left-edge cue, right-edge cue) were matched as closely as possible: each condition contained 21 female participants, and the mean ages were, respectively, 19.25 (s.e.m. 0.34), 18.92 (s.e.m. 0.21), and 19.33 (s.e.m. 0.34) years. Dutch participants were recruited from the participant panel at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and were university students from a variety of academic disciplines. All were native speakers of Dutch, all had learned English at school; the majority had also learned German, and two-thirds had learned French. In the three conditions, there were, respectively, 18, 18, and 19 females, and the mean ages were 21.83 (s.e.m. 0.60), 22.04 (s.e.m. 0.56), and 21.00 (s.e.m. 0.45) years. The English speakers were first-year psychology students at the University of Western Sydney, Australia. All were native speakers of English, and there was no systematic pattern of exposure to other languages in the sample. In the three conditions there were, respectively, 19, 21, and 22 females, and the mean ages were 23.17 (s.e.m. 1.13), 24.42 (s.e.m. 1.77), and 21.13 (s.e.m. 0.88) years.

2. Stimulus materials

The artificial language consisted of the concatenation of nine words. In ALL studies the words usually have an equal

number of syllables (e.g., three: Saffran *et al.*, 1996b; Vroomen *et al.*, 1998). However, if that were the case here, the isochronous rhythm created by the addition of a vowel lengthening cue could improve participants' performance over and above any influence of the cue itself. The words of the language used here therefore varied in length—six trisyllabic words and three words of four syllables. The 30 consonant-vowel syllables were constructed by exhaustively combining six consonants and five vowels that occur in the phoneme inventories of French, Dutch, and English: /p, b, m, f, s, k/, and /a, i, ɛ, ɔ, u/.

There is always the possibility that listeners may come to an ALL task with biases that influence word-boundary detection during the exposure phase (Reber and Perruchet, 2003). For example, certain syllables, consonants, or vowels may occur with a higher probability at word boundaries in the listener's native language, and combinations of syllables in the artificial language may resemble real words. To counteract such effects, we randomly allocated the 30 syllables to words, without repetition, such that the nine words of the artificial language were composed of a different unique combination of syllables for each of the 24 participants in any given condition. For each combination, TP-only, left-edge-cue, and right-edge-cue versions were created; thus each combination was presented to one participant per condition, and the conditions were balanced for variation across syllable combinations.

The 24×3 languages were generated using the MBROLA diphone synthesizer (Dutoit *et al.*, 1996). Each consonant and vowel was assigned a base length of 116 ms, resulting in a syllable length of 232 ms (following Peña *et al.*, 2002). As a partial control for effects of phonetic differences between realizations of the vowels and consonants across the three languages, and effects of native-language experience, half of the participants in each language group heard a language synthesized with a male French voice (MBROLA's fr1 diphone database) and the other half a male Dutch voice (the n12 diphone database). There is currently no Australian English diphone database for MBROLA. Note that each phoneme was coarticulated with the following phoneme, regardless of its position in the word.

The f0 was set to a monotone 120 Hz in all three cue conditions, and in the cued conditions the vowel in either the first syllable (left-edge cue) or last syllable (right-edge cue) of each statistical word was lengthened by 60 ms. The initial exposure lasted for a total of 11–12 min., depending on the condition, and was divided into five blocks of equal length to help maintain participants' attention. A 5-s fade-in and fade-out was applied to each block so that participants would not have access to word-boundary cues from the beginning and end of the sequence. The words were presented 19 times per block and in random order, with the sole constraint that a given word could not follow itself.

The test phase included 27 pairs of items. One member of each pair was a word from the language and the other was a part-word, that is, a sequence of syllables that occurred in the stream but crossed a word boundary. For half of the participants, the part-words were formed from the last two (or three) syllables of a word and the first syllable of another, and for the other half they were formed from the last syllable of a word and the first two (or three) syllables of another. All nine words of the language were used in the test phase, along with nine part-words. Each word was paired with three part-words and each part-word was paired with three words to counteract learning during the test phase. All items in the test phase were presented with the vowel lengthening cue corresponding to the stimulus condition (either TP only, left- or right-edge), following Vroomen *et al.* (1998).[1] The words and part-words were separated by an interstimulus interval of 500 ms. The order of words and part-words in the item pairs was counterbalanced, and the item pairs were presented in random order. The five exposure blocks and 27 pairs of test items were presented over headphones using a computer.

3. Procedure

Participants were tested in groups of one, two, or three, each seated in front of a different computer. Verbal instructions were given by the first author, who interacted with the French participants in French and with the Dutch and Australian participants in English. To ensure that all of the directions were understood, written instructions were also provided in the participant's native language.

Participants were instructed that they would hear an artificial language, consisting of a sequence of syllables with no pauses; they were asked to pay attention to the exposure stream without reflecting too much on what they were hearing, or trying to guess the purpose of the experiment. If they felt their attention start to wander they were to try to focus again on the task. Participants were made aware that there would be a test phase after the exposure phase.

Before the test phase, the participants were told that the artificial language consisted of nonsense strings that the experimenter had designated to be the words of the language. The purpose of the test phase was to find out if they had learned anything about those words during exposure. It was stressed that they would not be real words in the participant's native language, and that any resemblance to known words would be coincidental. Participants listened to each pair of a word and a part-word and indicated, by pressing the keys "1" or "2," whether the word of the language was the first or second member of the pair. They were told to guess if unsure.

B. Results and discussion

Percent correct scores were calculated for each participant, and then each group's mean score was derived from these values. The mean percent correct responses for each language group in each of the three cue conditions are presented in Fig. 1, and the exact values and standard errors of the mean are presented in Table I.[2] An alpha level of 0.05 was used for all statistical tests unless otherwise specified.

Before assessing whether word segmentation improved in the edge cue conditions, relative to TP-only, it is necessary first to test whether participants segmented the artificial language in the TP-only condition. If participants did not segment the artificial speech stream, then their performance on the two-alternative forced-choice test would not be greater

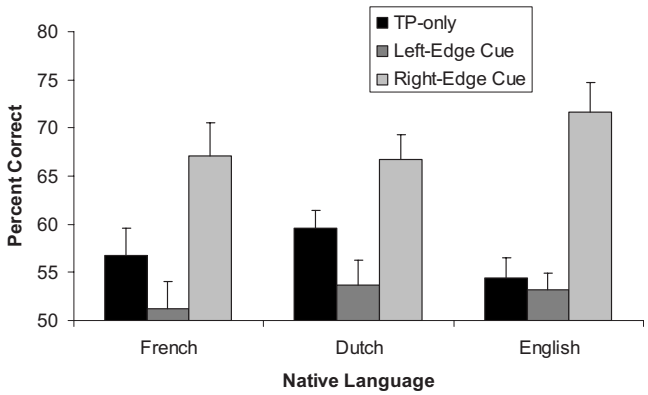


FIG. 1. Mean percent correct scores in the test phase for each language group in each cue condition in Experiment 1 (vowel-lengthening cues). Error bars represent s.e.m.

than chance (50%). For reference, the left-edge and right-edge cue conditions were also analyzed. Results of one-sample *t*-tests against a chance score of 50% are shown in Table I. Participants from each language group performed above chance in the TP-only and right-edge-cue conditions, but none of the language groups performed above chance in the left-edge-cue condition.

Having established that the artificial language can be segmented on the basis of TP cues only, our second analysis used planned contrasts to test whether segmentation was affected by cue (i.e., left-edge or right-edge vowel lengthening) and whether that varied according to the listener's native-language background. The *left-edge difference* and the *right-edge difference* contrasts compared the baseline TP-only condition with the left-edge and right-edge cue conditions, respectively. As these contrasts are not orthogonal, a Bonferroni correction was applied to the analysis (see Betz and Levin, 1982). Two additional planned contrasts assessed the influence of the listener's native language on learning. To test for effects of native-language prosodic preferences, the *language type* contrast compared the scores of French listeners (iambic, no-stress) with the combined scores of Dutch and English listeners (trochaic, with stress). The *stress language prosodic sensitivity* contrast compared scores of Dutch and English listeners only, to test performance of listeners who are more sensitive to suprasegmental information (i.e., Dutch) versus those who are less sensitive (i.e., English). Four interaction contrasts were also calculated to test for differential effects of cue location as a function of language background.

The results of the planned contrast analysis are shown in Table II. The only significant contrast was right-edge difference, with participants' scores being significantly higher in the right-edge than TP-only condition. Although performance in the left-edge condition dropped to chance level, the score was not significantly lower than the TP-only condition. None of the interaction contrasts was significant.

Experiment 1 showed, therefore, that all participants, regardless of language background, benefited from vowel lengthening if and only if it was a right-edge cue. Cross-language differences in preferred prosodic structure had no effect. This is consistent with a universal status for final lengthening as a boundary cue.

III. EXPERIMENT 2: PITCH-MOVEMENT CUES

A. Method

1. Participants

Another 216 participants, 72 from each of the same populations as for Experiment 1, took part in Experiment 2; again 24 were randomly assigned to each cue condition. All French participants had acquired French from birth, with the exception of one participant in the TP-only condition (L1: Mauritian creole), and two in the right-edge-cue condition (L1: Mandinka, Portuguese) who all acquired French in early childhood. All had also learned some English at school, around half had learned Spanish, and around one-third had learned German. In the TP-only, left-edge-cue and right-edge-cue conditions, respectively, there were 20, 18, and 18 females, and the mean ages were 20.82 (s.e.m. 0.76), 21.17 (s.e.m. 1.71), and 20.33 (s.e.m. 0.50) years. Dutch participants were native speakers of Dutch, had learned English at school, and had also been exposed to some French at school. In the three conditions, there were, respectively, 19, 19, and 20 females, and the mean ages were 20.63 (s.e.m. 0.49), 20.13 (s.e.m. 0.39), and 20.59 (s.e.m. 0.65) years. The English participants were again all native speakers of that language; 12 had taken some French at school. In the three conditions there were, respectively, 16, 18, and 20 females, and the mean ages were 20.29 (s.e.m. 1.11), 20.04 (s.e.m. 0.95), and 20.96 (s.e.m. 1.26) years.

2. Stimuli and procedure

The artificial languages used in this experiment had the same structure, and were constructed in the same manner as those used in Experiment 1, with the exception of the fundamental frequency characteristics and the vowel length. Syllable duration was that of the TP-only condition in Experiment 1 (232 ms). The f0 was set to a monotone 120 Hz for all syllables in the TP-only condition, while in the cued con-

TABLE I. Mean percent correct scores, standard error of the mean, and $t(23)$ values from a one-sample *t*-test against chance (50%) for Experiment 1 (vowel-lengthening cues). Values marked with an asterisk were significant at the 0.05 level.

	TP-only			Left-edge cue			Right-edge cue		
Native Language	*M*	s.e.m	$t(23)$	*M*	s.e.m	$t(23)$	*M*	s.e.m	$t(23)$
French	56.79	2.76	2.46*	51.23	2.78	0.44	67.13	3.39	5.05*
Dutch	59.57	1.84	5.21*	53.70	2.60	1.42	66.67	2.63	6.34*
English	54.48	2.06	2.17*	53.24	1.64	1.98	71.60	3.11	6.95*

TABLE II. Planned contrast analysis for Experiment 1 (vowel-lengthening cues). Contrast values are percent mean difference scores, and asterisks indicate significance with Bonferroni adjustment.

Contrast	$F(1,207)$	Contrast value	s.e.m.	Bonferroni 95% confidence interval	
				Lower	Upper
Language type	0.66	1.49	1.83	−2.65	5.63
Prosodic sensitivity	0.01	−0.21	2.12	−4.99	4.57
Left-edge difference	3.97	−4.22	2.12	−9.00	0.56
Right-edge difference	29.62*	11.52	2.12	6.74	16.30
Language type × left-edge difference	0.20	−2.01	4.49	−13.32	9.31
Language type × right-edge difference	0.16	−1.78	4.49	−13.09	9.54
Prosodic sensitivity × left-edge difference	0.80	−4.63	5.19	−17.70	8.44
Prosodic sensitivity × right-edge difference	3.74	−10.03	5.19	−23.10	3.04

ditions a parabolic f0 contour with its peak at 170 Hz was imposed on the cued syllable (following Thiessen and Saffran, 2003). Procedure was as in Experiment 1.

B. Results and discussion

Mean percent correct responses for each language group in each of the three conditions of Experiment 2 are displayed in Fig. 2, and the exact values and standard errors of the mean are listed in Table III.[3] As can be seen, all participants performed above chance in the TP-only condition. The pattern of learning across conditions was similar to Experiment 1 for French listeners, but a different pattern emerged here for Dutch and English listeners. Dutch listeners performed above chance in all conditions, whereas English listeners performed above chance in all but the right-edge-cue condition.

Planned contrasts were applied to the data as in Experiment 1, and the results are shown in Table IV. As in Experiment 1, the significant right-edge difference contrast shows that, across the three listener groups, participants scored higher in the right-edge condition than the TP-only condition. (That is, the magnitude of the difference for French and Dutch listeners combined was sufficient to compensate for the English listeners' chance performance in the right-edge condition.) The significant Language type × left-edge difference interaction contrast shows that stress-language listeners (Dutch and English) benefited from the left-edge cue more than the French listeners. The prosodic sensitivity × right-edge difference contrast qualifies the significant overall right-edge difference by showing that Dutch listeners benefited more from the right-edge cue than English listeners. These analyses confirm the pattern of results seen in Fig. 2—French listeners benefit from right-edge cues only, Dutch listeners benefit from both left- and right-edge cues, whereas English listeners benefit from left-edge cues only.

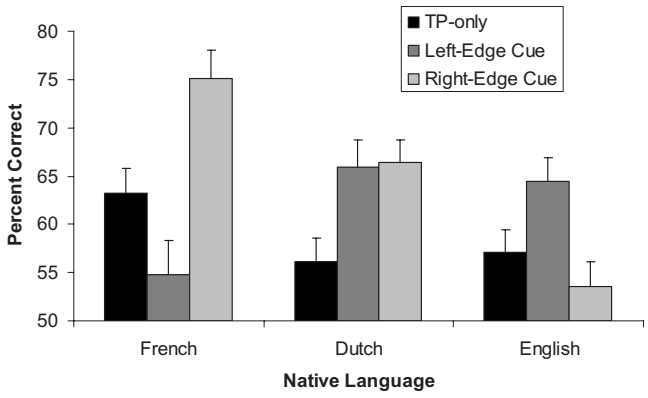


FIG. 2. Mean percent correct scores in the test phase for each language group in each cue condition in Experiment 2 (pitch-movement cues). Error bars represent s.e.m.

Experiment 2 has thus shown that pitch-movement cues to segmentation produce different results with listeners from different languages. In contrast to the universally consistent pattern revealed for vowel lengthening cues in Experiment 1, sensitivity to pitch-movement cues to segmentation is dependent on language-specific factors: preferred prosodic structure and patterns of stress realization.

IV. CONCLUSIONS

ALL as a tool for investigating speech segmentation reveals both universal similarities and cross-linguistic differences. Consistent with our predictions, based on the linguistic literature and one prior result with English-speaking listeners, lengthening realized in final position proved a powerful cue, significantly improving the segmentation performance of listeners from all language backgrounds we tested. Also consistent with our predictions, based on the cross-language differences in prosodic structure, listeners whose languages exhibit a preference for trochaic stress benefited more from a left-edge pitch cue, while listeners whose language displays preferred iambic prominence benefited more from the same cue realized in right-edge position. Finally, our prediction concerning the two highly similar stress languages, Dutch and English, was also supported: Dutch listeners exploited the presence of a pitch cue to a greater extent than English listeners did.

Thus exactly the same artificial-language input can be parsed differently if experience with the native language encourages attention to different cues in the input. Even when languages encourage use of the same cue in the same position, this does not entail that the way in which the cue is used is also the same; it too can differ. However, there are also cues that appear to be used in the same way across prosodically different languages.

Vowel lengthening is a powerful segmentation cue if it is associated with the right edge of structural components. The

TABLE III. Mean percent correct scores, standard error of the mean, and $t(23)$ values from a one-sample t-test against chance (50%) for Experiment 2 (pitch-movement cues). Value marked with an asterisk were significant at the 0.05 level.

	TP-only			Left-edge cue			Right-edge cue		
Native language	*M*	s.e.m	$t(23)$	*M*	s.e.m	$t(23)$	*M*	s.e.m	$t(23)$
French	63.27	2.50	5.31*	54.78	3.55	1.35	75.15	2.93	8.59*
Dutch	56.17	2.43	2.54*	65.90	2.90	5.48*	66.36	2.39	6.84*
English	57.10	2.33	3.05*	64.51	2.36	6.15*	53.55	2.52	1.41

widespread use of final-element lengthening suggests that it does not derive solely from linguistic structure but is more general in application (consider that lengthening of units in final position is also observed in music: Lindblom, 1978; Palmer, 1997). Hay and Diehl (2007) interpreted their finding of similar grouping preferences in speakers of French and English as supporting an explanation of the iambic/trochaic law in terms of general auditory mechanisms rather than linguistically based regularities. Our present findings are fully consistent with a general explanation of this nature.

Both the studies of Hay and Diehl (2007) and of Thiessen and Saffran (2003) were apparently conceived in the expectation that English listeners might prefer, or be more sensitive to, left-edge lengthening rather than right-edge lengthening. In both cases the authors cited the results showing that English listeners use stress in segmentation, combined with the fact that lengthening is a correlate of English stress. However, stress in English is multiply determined (see Cutler, 2005, for a review); there are suprasegmental cues in f0, duration, and amplitude, but stress placement is most highly correlated with segmental structure: syllable weight and vowel quality. Furthermore, the literature clearly shows that for lexical-level segmentation, English listeners mainly rely on the segmental information. Both in natural listening and in the laboratory, their preferred segmentation strategy is to postulate a word boundary at the onset of any strong syllable, that is, any syllable containing a full vowel (Cutler and Butterfield, 1992; Cutler and Norris, 1988). This heuristic pays off very effectively in segmenting typical English speech (Cutler and Carter, 1987), and has correctly accounted for the findings on segmentation of English when it is incorporated into computational models of word recognition in continuous speech (Norris *et al.*, 1997; Norris and McQueen, 2008). The literature also shows that English listeners make less use of suprasegmental cues for lexical processing than the acoustic structure of speech supports (Fear *et al.*, 1995; Cooper *et al.*, 2002), although the use of durational cues for syntactic processing is robust (Scott, 1982). Thus the previous findings and our current findings are fully consistent.

Lexical segmentation in normal speech recognition operates as efficiently and as rapidly as listeners can manage. Current models of spoken-word recognition (e.g., Norris and McQueen, 2008) assume that listeners evaluate multiple lexical hypotheses concurrently, and, in fact, this process itself parses continuous input into a sequence of words; nevertheless, listeners exploit the further cues to segmentation which speech signals provide, at many different levels, as ample empirical evidence attests (Mattys *et al.*, 2005). Explicit segmentation using cues in the signal, and segmentation arising from concurrent evaluation of word hypotheses, can moreover be shown to be distinct processes. Cutler and Butterfield (1992) analyzed a corpus of errors of segmentation in English, in which the predicted effect of vowel quality was found: a syllable was more likely to be erroneously interpreted as word-initial if it contained a full vowel. Effects of lexical structure on segmentation, however, were not observed; thus syllables were not more often taken as word-initial when they began more words, and erroneously reported words were more frequent than the actual words in the input only when boundaries were correct, not when boundaries had been inferred from the vowel quality cue. This pattern is consistent with use of cues in the input to deliver an initial segmentation which then constrains the set of lexical hypotheses for evaluation.

Silence abutting a word would putatively be the strongest of all such cues, and even tiny silences inserted between words produce significant improvement in ALL performance

TABLE IV. Planned contrast analysis for Experiment 2 (pitch-movement cues). Contrast values are percent mean difference scores, and asterisks indicate significance with Bonferroni adjustment.

				Bonferroni 95% confidence interval	
Contrast	$F(1,207)$	Contrast value	s.e.m.	Lower	Upper
Language type	4.03	−3.81	1.90	−8.09	0.48
Prosodic sensitivity	4.08	−4.42	2.19	−9.37	0.52
Left-edge difference	1.73	2.88	2.19	−2.07	7.83
Right-edge difference	7.94*	6.17	2.19	1.23	11.12
Language type × left-edge difference	13.46*	−17.05	4.65	−28.76	−5.34
Language type × right-edge difference	3.40	8.57	4.65	−3.15	20.28
Prosodic sensitivity × left-edge difference	0.19	2.32	5.37	−11.21	15.84
Prosodic sensitivity × right-edge difference	6.55*	13.73	5.37	0.21	27.26

(Toro *et al.*, 2008). Although vowel lengthening as a right-edge cue most strongly signals syntactic boundaries (Klatt, 1975), even in doing so it would still function as a cue to the end of a word as well as of the phrase of which the word was a part. However, in fact, there is also consistent lengthening associated with the right edge of words (Beckman and Edwards, 1990). All this makes vowel lengthening a right-edge cue of considerable power, and listeners make appropriate use of it, as the results of Experiment 1 showed. It is even noteworthy that the listeners in all groups in Experiment 1 performed slightly (albeit not significantly) worse than with TP alone when the vowel lengthening cue was associated with the left edge of words. This may also be due to the power of the cue as a signal of right edges, but here placed, under such an interpretation, in conflict with the TP cues.

The appearance of a consistent pattern across languages in one experiment, however, does not mean that cross-language differences in prosodic structure exert no influence on ALL segmentation. The relative strength of the cues in different positions clearly varied across the three listener groups we tested, most clearly in Experiment 2 in which we manipulated pitch movement. In that experiment, French listeners gained no benefit at all from a left-edge cue and English listeners gained no benefit at all from a right-edge cue. This is exactly in accord with the preferred expression of syllabic prominence (left edge of lexical units in English, right edge of lexical units—strictly speaking, of clitic groups—in French).

Note that the pitch-movement cue as we (and others) have manipulated it is something of a caricature of what pitch does in stress variation. However, undeniably it is the case that whatever it does, it does it in characteristically different locations in English and in French, and precisely that predicted difference turned up in our results.

The pitch cue that we used was realized locally on a left- or right-edge syllable. Note that this has not always been the case when pitch cues have been manipulated in ALL studies of segmentation. For example, the pitch cue manipulated by Vroomen *et al.* (1998) (referred to by those authors as "stress") was actually a word-level prosodic contour grouping all three syllables of the words they used into a consistent prosodic shape, with prominence on the initial syllable. It seems that this should have been a relatively easy cue to use in an ALL experiment, so that their finding that French listeners did not benefit from such a cue is somewhat surprising. Vroomen *et al.* (1998) interpreted the finding in terms of language-specific prosodic structure; the characteristic prosodic shape of French words does not correspond to the prosodic shape they used as a cue. However, in a replication of their experiment, Tyler (2006) demonstrated that French listeners indeed showed significant benefit from such a prosodic shape cue, both in learning a more complex artificial language (analogous to the ones used in the present study) and in learning from the very materials used in the Vroomen *et al.*, 1998 study.[4]

Interestingly, in one ALL study in which a pitch cue was manipulated at the syllabic level, as we did, no benefit accrued for listeners. This was a study with Spanish listeners conducted by Toro-Soto *et al.* (2007). The Spanish lexicon has a strong preponderance of words with penultimate stress, and Toro-Soto *et al.* (2007) tested the value of such a stress cue (realized as an increase in syllable pitch on the penultimate syllable of trisyllabic words). Listeners performed no better than chance with this cue (and significantly worse than their performance with TP information only, or with an initial or final stress cue of the same kind). Thus for a cue to be useful for segmenting an artificial language, it seems that it should preferably be aligned with word edges.

Finally, we also observed in Experiment 2 a significant difference, again as we had predicted, in the learning performance of our English and our Dutch participants. The Dutch listeners profited from the pitch-movement cue in both initial position and final position, while the English listeners showed a benefit only in initial position. In final position, the Dutch listeners' performance was significantly better than that of the English listeners; the latter actually performed somewhat worse in this condition than with TP-only, again consistent with an effect which was powerfully unidirectional and in conflict, under the preferred interpretation, with TP.

The Dutch listeners, however, were able to override such a preference and make use of pitch movement when it uncharacteristically provided a consistent right-edge cue. Thus in this experiment, as in preceding studies (Cooper *et al.*, 2002; Donselaar *et al.*, 2005; Cutler *et al.*, 2007; Cutler, 2009), Dutch listeners displayed greater sensitivity than English listeners to the suprasegmental cues to stress in speech, especially, in this case, the pitch-movement cue; they could learn to use it as a marker of lexical identity irrespective of its position in words, while the English listeners used it only in the position in which it most commonly occurs in English.

We interpret this not as a reflection of cross-language differences in positional marking of stress; this is overwhelmingly initial in both English [for which Cutler and Carter (1987), reported 90% of the lexical vocabulary to be strong-initial] and Dutch (for which Schreuder and Baayen (1994), reported 87.5%). Rather, it provides yet further evidence for the greater sensitivity of Dutch listeners to suprasegmental structure. We note here that a finding by Thiessen and Saffran (2004) in an ALL experiment can also be interpreted in this light. Thiessen and Saffran (2004) observed that infants could base speech segmentation on the exploitation of spectral tilt (the relative distribution of amplitude across the spectrum, a strong cue to stress in that stressed syllables show significantly greater amplitude in the higher spectral regions than unstressed syllables). Adult English-speaking listeners, however, ignored variation in spectral tilt unaccompanied by other stress cues. Spectral tilt in isolation has been shown to be an effective cue to stress for Dutch listeners (Sluijter *et al.*, 1997) but not for English listeners (Campbell and Beckman, 1997; see Cutler, 2005, for further discussion).

This consistent pattern of findings underlines how powerful are the distributional statistics of language-specific phonology in determining listeners' attention to speech cues, and how efficiently adult listening exploits native-language probabilities. The widespread occurrence of vowel reduction in unstressed syllables in English has encouraged English lis-

teners to skip attention to suprasegmental variation for lexical identification, since vowel quality will virtually always provide all the information that the suprasegmental variation encodes, and vowel quality must be attended to anyway. The frequent occurrence in Dutch of unstressed syllables with full vowels (such as the first syllables of *sigaar* "cigar" or *Oktober* "October") has made Dutch listeners realize that they can profit from attending to suprasegmental variation, because they can distinguish more rapidly between potential words than would be possible on the basis of segmental information alone.

ALL techniques have enabled us to observe the effect of these different language-specific patterns, and at the same time to appreciate the strength of language-universal effects, because they have allowed a direct comparison of the use of speech cues in exactly the same input across listeners with different native languages. We have seen that listeners are very good at segmentation on the basis of distributional information alone (the TP-only conditions, in which listeners could only identify recurring items on the basis of their sequential probabilities, consistently produced above-chance performance from all listener groups). Beyond that, they can also make use of other cues where these are available. But not all cues are equal. A durational cue is a significant help, but only when it is in the universally preferred right-edge position. A pitch-movement cue is also a significant help, but for most listeners such a cue is only helpful in the characteristic native-language position—right-edge for French listeners and left-edge for English listeners. Only Dutch listeners, whose language encourages careful attention to suprasegmental information, displayed sufficient flexibility to exploit a consistent pitch-movement cue both in its expected position and in an uncharacteristic mapping.

Speech segmentation is one of the most useful language processing skills. It develops early in life, and it directly assists language learning: facility with speech segmentation in the first year of life is associated with enhanced vocabulary development in the following years (Newman *et al.*, 2006). From the earliest stages it is adapted to the native language and exploits the distributional probabilities of the input (McQueen, 1998; van der Lugt, 2001). This of course has the inevitable downside that listening to a non-native language is rendered more difficult where the structure of the native language encourages segmentation procedures inappropriate for the other language (Weber and Cutler, 2006). However, this is apparently a small price to pay for the streamlined efficiency with which native input can be divided into its component words. Even with the impoverished nonword input on offer in an ALL experiment, this efficiency can be seen in action. In ALL, listeners know the input is not real language, and they know that their only task is to use the information in the input to the best of their ability. Nonetheless, they succeed in using cues only to the extent that the cue position, and/or the cue type, coincides with their native-language experience. The artificial nature of ALL input allows these effects to be observed, in that it provides a direct window onto cross-language differences.

ACKNOWLEDGMENTS

The first author was supported by a postdoctoral fellowship from the Conseil Régional de Bourgogne, France, the NWO SPINOZA project "Native and Non-native Listening" (A.C.), and NIH Grant No. DC00403 (PI: C. Best). The authors thank Pierre Perruchet for his time and advice, especially with design and generation of the ALL stimuli; Kylie Tyler, Marloes van der Goot, Eelke Spaak, Jess Hartcher-O'Brien, and Susan Wijngaarden for research assistance; and Nigel Nettheim, Kate Stevens, and the reviewers for helpful comments on the manuscript.

[1]Note that Bagou *et al.* (2002) found no difference between cued and uncued test items in a similar task.

[2]There was no effect of synthesis voice (French or Dutch), and this factor did not interact with any other variable. All analyses reported here therefore collapse across that counterbalancing factor.

[3]As in Experiment 1, there was no effect of synthesis voice (French or Dutch), and this factor did not interact with any other variable, so analyses are again collapsed across that counterbalancing factor.

[4]The difference in results was attributed to audio presentation differences (headphones in Tyler 2006, and loudspeakers in Vroomen *et al.* 1998).

Bagou, O., Fougeron, C., and Frauenfelder, U. H. (**2002**). "Contribution of prosody to the segmentation and storage of 'words' in the acquisition of a new minilanguage," in *Proceedings of Speech Prosody 2002*, edited by B. Bel, and I. Marlien (Association pour la promotion de la phonétique et de la linguistique, Aix-en-Provence, France), pp. 159–162.

Beckman, M. E., and Edwards, J. (**1990**). "Lengthenings and shortenings and the nature of prosodic constituency," in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston, and M. E. Beckman (Cambridge University Press, Cambridge), pp. 152–178.

Betz, M. A., and Levin, J. R. (**1982**). "Coherent analysis-of-variance hypothesis testing strategies: A general approach," J. Educ. Stat. **7**, 193–206.

Bolinger, D. L. (**1978**). "Intonation across languages," in *Universals of Human Language*, Phonology, Vol. **2**, edited by J. H. Greenberg (Stanford University Press, Stanford), pp. 471–524.

Campbell, N., and Beckman, M. E. (**1997**). "Stress, prominence, and spectral tilt," in *Intonation: Theory, Models, and Applications (Proceedings of a European Speech Communication Assoc. Workshop, September 18–20, 1997)*, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis (ESCA and University of Athens Department of Informatics, Athens), pp. 67–70.

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.

Cutler, A. (**2005**). "Lexical stress," in *The Handbook of Speech Perception*, edited by D. B. Pisoni, and R. E. Remez (Blackwell, Oxford), pp. 264–289.

Cutler, A. (**2009**). "Greater sensitivity to prosodic goodness in non-native than in native listeners (L)," J. Acoust. Soc. Am. **125**, 3522–3525.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Mem. Lang. **33**, 824–844.

Cutler, A., and Pasveer, D. (**2006**). "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," in *Proceedings of the Third International Conference on Speech Prosody*, edited by R. Hoffman, and H. Mixdorff (TUD, Dresden), pp. 250–254.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Mem. Lang. **25**, 385–400.

Cutler, A., Wales, R., Cooper, N., and Janssen, J. (**2007**). "Dutch listeners' use of suprasegmental cues to English stress," in *Proceedings of 16th*

International Congress of Phonetic Sciences, edited by J. Trouvain and W. J. Barry (Saarbrücken, Germany), pp. 1913–1916.
van Donselaar, W., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. A **58A**, 251–273.
Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and van der Vrecken, O. (**1996**). "The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes," in *Proceedings of the Fourth International Conference on Spoken Lang. Processing*, edited by H. T. Bunnell, and W. Idsardi, pp. 1393–1396.
Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.
Hay, J. S. F., and Diehl, R. L. (**2007**). "Perception of rhythmic grouping: Testing the iambic/trochaic law," Percept. Psychophys. **69**, 113–122.
Hayes, B. (**1995**). *Metrical Stress Theory: Principles and Case Studies* (University of Chicago Press, Chicago).
Johnson, E. K., and Jusczyk, P. W. (**2001**). "Word segmentation by 8-month-olds: When speech cues count more than statistics," J. Mem. Lang. **44**, 548–567.
Johnson, E. K., and Seidl, A. H. (**2009**). "At 11 months, prosody still outranks statistics," Dev. Sci. **12**, 131–141.
Klatt, D. H. (**1975**). "Vowel lengthening is syntactically determined in connected discourse," J. Phonetics **3**, 129–140.
Lindblom, B. (**1978**). "Final lengthening in speech and music," in *Nordic Prosody*, edited by E. Gårding, G. Bruce, and R. Bannert (Department of Linguistics, Lund University, Lund, Sweden), pp. 85–100.
Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.
McQueen, J. M. (**1998**). "Segmentation of continuous speech using phonotactics," J. Mem. Lang. **39**, 21–46.
Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (**2008**). "The link between statistical segmentation and word learning in adults," Cognition **108**, 271–280.
Newman, R., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P. W., and Dow, K. A. (**2006**). "Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis," Dev. Psychol. **42**, 643–655.
Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: A Bayesian model of continuous speech recognition," Psychol. Rev. **115**, 357–395.
Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). "The possible-word constraint in the segmentation of continuous speech," Cognit Psychol. **34**, 191–243.
Otake, T., Hatano, G., Cutler, A., and Mehler, J. (**1993**). "Mora or syllable? Speech segmentation in Japanese," J. Mem. Lang. **32**, 258–278.
Otake, T., Hatano, G., and Yoneyama, K. (**1996**). "Speech segmentation by Japanese listeners," in *Phonological Structure and Language Processing: Cross-Linguistic Studies*, edited by T. Otake, and A. Cutler (Mouton de Gruyter, Berlin), pp. 183–201.
Palmer, C. (**1997**). "Music performance," Annu. Rev. Psychol. **48**, 115–138.
Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (**2002**). "Signal-driven computations in speech processing," Science **298**, 604–607.
Reber, R., and Perruchet, P. (**2003**). "The use of control groups in artificial grammar learning," Q. J. Exp. Psychol. A **56A**, 97–115.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996a**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.
Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996b**). "Word segmentation: The role of distributional cues," J. Mem. Lang. **35**, 606–621.
Schreuder, R., and Baayen, R. H. (**1994**). "Prefix stripping re-revisited," J. Mem. Lang. **33**, 357–375.
Scott, D. (**1982**). "Duration as a cue to the perception of a phrase boundary," J. Acoust. Soc. Am. **71**, 996–1007.
Sluijter, A. M. C., van Heuven, V. J., and Pacilly, J. J. A. (**1997**). "Spectral balance as a cue in the perception of linguistic stress," J. Acoust. Soc. Am. **101**, 503–513.
Suomi, K., McQueen, J. M., and Cutler, A. (**1997**). "Vowel harmony and speech segmentation in Finnish," J. Mem. Lang. **36**, 422–444.
Thiessen, E. D., and Saffran, J. R. (**2003**). "When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants," Dev. Psychol. **39**, 706–716.
Thiessen, E. D., and Saffran, J. R. (**2004**). "Spectral tilt as a cue to word segmentation in infancy and adulthood," Percept. Psychophys. **66**, 779–791.
Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (**2008**). "Finding words and rules in a speech stream: Functional differences between vowels and consonants," Psychol. Sci. **19**, 137–144.
Toro-Soto, J. M., Rodríguez-Fornells, A., and Sebastián-Gallés, N. (**2007**). "Stress placement and word segmentation by Spanish speakers," Psicologica **28**, 167–176.
Tyler, M. D. (**2006**). "French listeners can use stress to segment words in an artificial language," in *Proceedings of the 11th Australasian International Conference on Speech Sci. & Tech.*, edited by P. Warren, and C. I. Watson (Australasian Speech Sci. and Technol. Assoc. Inc., Auckland, New Zealand), pp. 222–227.
Vaissière, J. (**1983**). "Language-independent prosodic features," in *Prosody: Models and Measurements*, edited by A. Cutler, and D. R. Ladd (Springer-Verlag, Hamburg), pp. 53–66.
van der Hulst, H. (**1999**). *Word Prosodic Systems in the Languages of Europe* (Mouton de Gruyter, Berlin).
van der Lugt, A. H. (**2001**). "The use of sequential probabilities in the segmentation of speech," Percept. Psychophys. **63**, 811–823.
Vroomen, J., Tuomainen, J., and de Gelder, B. (**1998**). "The roles of word stress and vowel harmony in speech segmentation," J. Mem. Lang. **38**, 133–149.
Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**, 597–607.
Woodrow, H. (**1909**). "A quantitative study of rhythm: The effect of variations in intensity, rate, and duration," Archives of Psychol. **14**, 1–66.

Audio-visual identification of place of articulation and voicing in white and babble noise[a)]

Magnus Alm[b)] and Dawn M. Behne
Department of Psychology, Norwegian University of Science and Technology, N-7491 Trondheim Norway

Yue Wang
Department of Linguistics, Simon Fraser University, Burnaby, British Columbia V5A 156 Canada

Ragnhild Eg
Department of Psychology, Norwegian University of Science and Technology, N-7491 Trondheim Norway

(Received 25 June 2008; revised 15 April 2009; accepted 15 April 2009)

Research shows that noise and phonetic attributes influence the degree to which auditory and visual modalities are used in audio-visual speech perception (AVSP). Research has, however, mainly focused on white noise and single phonetic attributes, thus neglecting the more common babble noise and possible interactions between phonetic attributes. This study explores whether white and babble noise differentially influence AVSP and whether these differences depend on phonetic attributes. White and babble noise of 0 and −12 dB signal-to-noise ratio were added to congruent and incongruent audio-visual stop consonant-vowel stimuli. The audio (A) and video (V) of incongruent stimuli differed either in place of articulation (POA) or voicing. Responses from 15 young adults show that, compared to white noise, babble resulted in more audio responses for POA stimuli, and fewer for voicing stimuli. Voiced syllables received more audio responses than voiceless syllables. Results can be attributed to discrepancies in the acoustic spectra of both the noise and speech target. Voiced consonants may be more auditorily salient than voiceless consonants which are more spectrally similar to white noise. Visual cues contribute to identification of voicing, but only if the POA is visually salient and auditorily susceptible to the noise type.
 [DOI: 10.1121/1.3129508]

PACS number(s): 43.71.Sy [AJ] Pages: 377–387

I. INTRODUCTION

Audio-visual speech perception (AVSP) denotes the human ability to benefit from both the auditory and visual modality in perceiving and interpreting speech. Research has shown that the visual modality plays an important role in perception of speech (e.g., Green and Kuhl, 1989; Massaro, 1987; McGurk and MacDonald, 1976) and its contribution is emphasized in conditions where the speech signal is degraded by noise (e.g., Erber, 1969; MacLeod and Summerfield, 1987; Sumby and Pollack, 1954). Previous research on AVSP has focused on the effect of white noise (e.g., Dodd, 1977; Fixmer and Hawkins, 1998), an auditory distractor infrequently present in everyday human speech communication, while the effect of babble noise is much less explored (e.g., Behne *et al.*, 2006; Markides, 1989; Wang *et al.*, 2008). The phonetic attributes place of articulation (POA) and voicing differ in susceptibility to noise (e.g., Miller and Nicely, 1955). However, whereas the effect of noise level on these attributes is well established (e.g., Jiang *et al.*, 2006; Miller and Nicely, 1955), the effect of different types of noise is not thoroughly explored. Research shows that speech perception proficiency by hearing-impaired and second language learners improves with training by directing visual attention to the specific cues relevant to the different attributes of speech (e.g., Lansing and McConkie, 1999; Massaro and Light, 2004; Wang *et al.*, 2008). The current study contributes to identifying where visual attention should be directed for POA and voicing in babble noise. Consequently, the present study on speech perception investigates the influence of different types and levels of noise, using audio-visual (AV) stimuli that differ in AV saliency (i.e., POA and voicing).

By using AV stimuli in which auditory syllables were dubbed onto visual syllables with different POAs, McGurk and MacDonald (1976) elegantly demonstrated vision's contribution to speech perception. Not only did responses frequently correspond to the visual component but also the AV cues were often fused and hence perceived as an intermediate percept to that of the auditory and visual components (i.e., AV-fusion). These findings show the bimodal nature of speech perception and are supported by later studies (e.g., Green and Kuhl, 1989; MacDonald and McGurk, 1978; Massaro, 1987; Summerfield and McGrath, 1984).

The availability of bimodal information renders speech perception less vulnerable to white noise (e.g., Erber, 1969; MacLeod and Summerfield, 1987; Sumby and Pollack, 1954). Macleod and Summerfield (1990) demonstrated that in white noise AV cues increase signal-to-noise ratio (SNR) thresholds by 6.1 dB compared to auditory speech alone.

[a)] Portions of this work were presented at Acoustics'08 and Interspeech 2008. This work is based on Magnus Alm's MA thesis, Norwegian University of Science and Technology (NTNU), Trondheim Norway, 2007.

[b)] Author to whom correspondence should be addressed. Electronic mail: magnus.alm@svt.ntnu.no

Since optimal hearing conditions greatly favor the auditory cues in AVSP, visual and AV-fusion responses are more likely to occur in noisy environments (e.g., Dodd, 1977; Easton and Basala, 1982; Ross *et al.*, 2007). However, whereas an increase in auditory noise results in a proportional shift toward visual responses, the relationship between noise level and AV-fusions is not linear (e.g., Ross *et al.*, 2007; Sommers *et al.*, 2005). Previous research (e.g., Dodd, 1977; Easton and Basala, 1982; Fixmer and Hawkins, 1998) demonstrates that moderate white noise facilitates AV-fusion responses, whereas extremely positive or negative SNRs favor the auditory or visual modality, respectively (Ross *et al.*, 2007).

Most studies on AVSP in noise referred to above used white noise. White noise has a flat power spectral density, meaning equal intensity levels across frequencies in a given band, while babble noise is a fluctuating signal with a low-frequency dominated spectral shape. Thus, given the same energy, white noise covers a wider range of frequencies than babble noise at any given time. The masking effect of noise is partly determined by the frequency overlap between a target stimulus and noise, and the size and frequency of uninterrupted speech intervals influence the intelligibility of the target signal in noise (i.e., glimpsing effect) (Cooke, 2006; Miller and Lickider, 1950).

To what extent noise influences the different modalities' contribution to AVSP depends on the characteristics of the speech signal (e.g., Binnie *et al.*, 1974). When perceiving speech, phonetic attributes such as POA and voicing have different susceptibilities to noise. Miller and Nicely (1955) showed that for auditory signals, POA identification is far more susceptible to noise than voicing identification. Whereas the identification of POA suffers at 6 dB SNR, voicing identification is robust at −12 dB SNR. These findings are supported by Jiang *et al.* (2006), who found that voicing identification was above chance level to a SNR of −15 dB. For stop consonants the acoustical cues important for POA and voicing identification differ. For POA identification the formant transition is deemed the most important cue (Blumstein *et al.*, 1982; Delattre *et al.*, 1955; Stevens and Blumstein, 1978), whereas for identification of voicing, the voice onset time (VOT) is considered the most important cue (Eimas and Corbit, 1973). In distinguishing consonants, the formant transition involves subtle acoustic variations within a wide range of frequencies, whereas VOT is associated with temporal variations in the time interval between consonant release and voice onset. An acoustically distinct event defines the end of VOT; that is, the speech signal shifts from a relatively flat intensity distribution across frequencies (aspiration) to a more fluctuating intensity distribution across frequencies (voice onset). This acoustic event is more temporally distinct than the subtle acoustic variations distinguishing different formant transitions and may contribute to voicing identification being less susceptible to noise than POA. POA's lack of auditory robustness is greatly compensated for by salient visual cues; that is, seeing the face of the speaker greatly aids perceivers in identifying POA in noise (e.g., Binnie *et al.*, 1974). This visual benefit is not found for voicing identification (e.g., Behne *et al.*, 2006; Binnie *et al.*, 1974). Behne *et al.* (2006) found that responses to incongruent AV syllables that varied in terms of voicing always matched the auditory component, independent of which component of the stimulus was voiced and independent of the presence of noise. The lack of visual access to the activity in the vocal folds may explain the poor visual contribution to voicing identification.

The current study employs the McGurk paradigm (McGurk and MacDonald, 1976) to explore whether white and babble noise influence the use of the auditory and visual modality differently, and whether these noise type differences depend on which phonetic attribute is being assessed. White noise is widely used in AV research, while babble noise, arguably a more common speech interference, is seldom used. Babble noise differs from white noise acoustically and possibly semantically. While phonetic attributes' susceptibility to different noise levels is well established (e.g., Jiang *et al.*, 2006; Miller and Nicely, 1955), the susceptibility to different noise types is less explored. Two hypotheses are tested: first, white noise is expected to result in fewer auditory and more visual and AV-fusion responses than babble noise, because white noise occludes more of the target signal than babble noise, that is, covers more frequencies. Second, white noise is expected to result in fewer auditory and more visual responses than babble noise for the noise susceptible POA identification, while no such noise type difference will be observed for the noise robust voicing identification.

II. METHODS

A. Design

Responses were collected for stimulus presentations consisting of congruent and incongruent AV stimuli containing stop-vowel syllables, varying in terms of POA, voicing, noise type, and noise level, using a repeated measures design.

B. Participants

Fifteen native Norwegian speakers within the range of 20–31 years of age ($M=24$, $SD=3$) participated in the experiment, seven of which were males and eight of which were females. The participants were students recruited from The Norwegian University of Science and Technology (NTNU) and Sør-Trøndelag University College (HIST). All participants reported normal hearing and normal or corrected-to-normal vision.

C. Stimuli

As shown in Table I, six congruent and ten incongruent AV stimuli were used in the experiment. The congruent stimuli were used as a control to measure the general AV intelligibility of the syllables across noise conditions (i.e., different noise types and noise levels). The incongruent syllables were used to test the hypotheses.

The AV stimuli were made from audio and visual recordings of six different monosyllables that differed in POA and voicing: Labial /ba/ and /pa/, alveolar /da/ and /ta/, and velar /ga/ and /ka/.

TABLE I. The six congruent (left) and the ten incongruent (right) AV syllables used as test materials. The first four incongruent syllables differ in POA. The remaining six differ in voicing.

Congruent stimuli		Incongruent stimuli	
Auditory	Visual	Auditory	Visual
		POA stimuli	
ba	ba	ba	ga
pa	pa	pa	ka
da	da	ga	ba
ta	ta	ka	pa
ga	ga	Voicing stimuli	
ka	ka	ba	pa
		pa	ba
		da	ta
		ta	da
		ga	ka
		ka	ga

As shown in Table I, congruent stimuli refer to stimuli in which the audio and visual components correspond and incongruent stimuli refer to stimuli in which the audio and visual components differ. For incongruent POA stimuli, the audio and visual components differ in POA, whereas for incongruent voicing stimuli the audio and visual components differ in voicing. Both POA and voicing stimuli included alternative stimulus structures: the POA stimulus structure was either $A_{\text{labial}}V_{\text{velar}}$ or $A_{\text{velar}}V_{\text{labial}}$, and the voicing stimulus structure was either $A_{\text{voiced}}V_{\text{voiceless}}$ or $A_{\text{voiceless}}V_{\text{voiced}}$.

A POA stimulus was either voiced or voiceless, whereas a voicing stimulus was either labial, alveolar, or velar. Therefore, the POA stimuli revealed participants' AV perception of POA, as well as the effect of consonant voicing on this POA perception. The voicing stimuli revealed participants' AV perception of voicing, as well as the effect of consonant POA on this voicing perception.

Four noise backgrounds were added to the congruent and incongruent AV stimuli: two intensity levels of white and babble noise. This resulted in four noise conditions in addition to the quiet condition.

1. AV recordings

The current study used AV recordings that were made for a previous study (see Behne *et al.*, 2006; Behne *et al.*, 2007). The AV recordings of a male speaker were made in the Speech Laboratory at the Department of Psychology, NTNU. The speaker had an urban Eastern-Norwegian dialect that is familiar to most Norwegians. The speaker was clean shaven and any artificial distractors, such as glasses and jewelry, were removed prior to video recording.

The speaker was instructed to keep a relatively flat intonation, avoiding a decline or incline at the end of a syllable. He was also told to keep facial gestures, such as eye blinks, to a minimum.

The speaker was seated inside a sound-insulated room with a Sony DCR-TRV50E camera positioned 90 cm in front of him and a Røde NT3 microphone positioned 50 cm to the left and 10 cm above his head. Two parallel audio recordings were made: one from the video camera's internal microphone and one from the external microphone (i.e., Røde). The sound from the external microphone was fed through an M-Audio Firewire 1814 box to a Apple Macintosh G5 computer, where two audio channels were recorded at a sampling rate of 44.1 kHz using PRAAT version 4.3.22 (Boersma and Weenink, 2006).

To avoid any visual distractions in the stimuli, the speaker was seated with his back toward a gray monotonous wall. His face was centered in the camera frame, leaving 4 in. of empty space on either side of the cheeks and 1 in. of empty space above the head when the video was displayed on a 17 in. computer monitor. The scene was thus believed to represent the visual field common to participants engaged in natural face-to-face communication.

The syllables used in the experiment consisted of six consonants /b, d, g, p, t/ and /k/, followed by the vowel /a/. The speaker repeated each of the syllables eight times to create a set of alternatives for finding one in which both auditory and visual qualities were good. The resulting QuickTime video file had a visual quality of 30 frames/s at a resolution of 640×480 pixels. The MPEG4/H264 video compression algorithm was used at a bit-rate of 57 600 kbps. The video file was segmented into separate syllables, using the software IMOVIE HD 5.0.2. The audio files derived from the external microphone were segmented using PRAAT (Boersma and Weenink, 2006).

The segmented video and audio files were rated independently by two different persons. Highly rated video segments were those in which syllable articulations were explicit and eye blinks or other unwanted facial gestures few. A highly rated audio segment implied a natural syllable pronunciation and a relatively even intonation, accompanied by no unwanted noise, such as that from movement in the recording environment. The six required AV syllables were selected based on the highest additive audio and visual ratings. The audio segments were edited in PRAAT to ensure the same unweighted intensity for all syllables and the corresponding video clips were cut in iMovie HD to meet the length of the longest video syllable (1960 ms). The auditory speech signals had an average length of 423 ms (range 377–468 ms), measured from consonant release for voiceless syllables and from onset of the voice bar for voiced syllables. Voiced syllables were measured from the onset of the voice bar in order to take into account an instance of prevoicing. The auditory speech signals were initiated 573 ms after the onset of the video clips, hereby enabling initiation of noise signals 573 ms prior to the auditory speech signals and thus avoiding artifacts caused by sudden onset of noise.

2. Noise signal

Two types of noise were used in the experiment: babble noise and white noise. The babble noise was recorded during lunchtime in a cafeteria at NTNU, using an Okay II DM-801 microphone connected to an SHG Note 40750 laptop via its built-in soundcard, and using a sampling frequency of 44.1 kHz. A segment of the recording was extracted in which babble was prominent and other sounds such as coughs and the rattling of cutlery were minimal. No individual voices

could be discerned in the babble segment. The white Gaussian noise used in the experiment was generated using the "Create sound" function in PRAAT (Boersma and Weenink, 2006). The babble and the white noise segments were cut to a length of 1960 ms, equaling the length of the video clips.

Two SNRs were used in the experiment. SNR was calculated by subtracting the mean noise intensity level from the mean speech signal intensity level. Different phonetic attributes, in this respect POA and voicing, are associated with different vulnerabilities to noise. A prior study by Behne *et al.* (2006) showed that for AV stimuli differing in POA, babble noise at a SNR of 0 dB led to an increase in the use of the visual modality, whereas no such shift was evident for AV stimuli that vary in voicing at this noise level. These results are supported by a firm body of research (e.g., Jiang *et al.*, 2006; Miller and Nicely, 1955). To further assess the AV benefit for voicing identification, a noise level at which visual responses to voicing stimuli occur had to be established. Miller and Nicely (1955) demonstrated that auditory voicing perception is robust at SNRs up to −12 dB, whereas Jiang *et al.* (2006) revealed that a SNR of −15 dB resulted in auditory responses that were rather arbitrary. The interval between −12 and −15 dB SNR was therefore considered. A pilot study with seven participants indicated a shift in use of modality near SNR −12 dB. Furthermore, this level was associated with predictable responding; that is, most participant responses corresponded either to the visual or the auditory part of the signal and were hence not subjected to merely guessing. The use of noise at 0 and −12 dB SNRs therefore seemed reasonable.

The two noise intensity levels were adjusted to 0 and −12 dB (relative levels) using PRAAT (Boersma and Weenink, 2006), resulting in a noise subset of five: white and babble noise of 0 dB SNR, white and babble noise of −12 dB SNR, and quiet.

3. Assembling stimuli

As shown in Table I, six congruent and ten incongruent AV stimuli were used in the experiment. The congruent stimuli were created by importing the video clips into iMovie HD and substituting the auditory speech signals recorded by the video camera's internal microphone with the corresponding auditory speech signals from the same syllable recorded by the external microphone. The incongruent stimuli were produced in the same manner, except that the original auditory speech signals were substituted with auditory speech signals that differed in POA or voicing. The auditory and visual signals were temporally aligned by synchronizing the consonant burst in the acoustic signal with mouth opening in the video clip. The resulting congruent and incongruent AV syllables constituted the quiet condition of the experiment.

The four noise segments (i.e., babble and white noise at 0 and −12 dB SNRs) had the same length as the film clips, ensuring that the noise covered the entire syllable. The noise segments were imported into iMovie HD and added to the already existing congruent and incongruent AV syllables. To combat the problem of differing total amount of energy imposed by the different noise levels, the new audio files were adjusted in PRAAT so that each stimulus had the same average intensity.

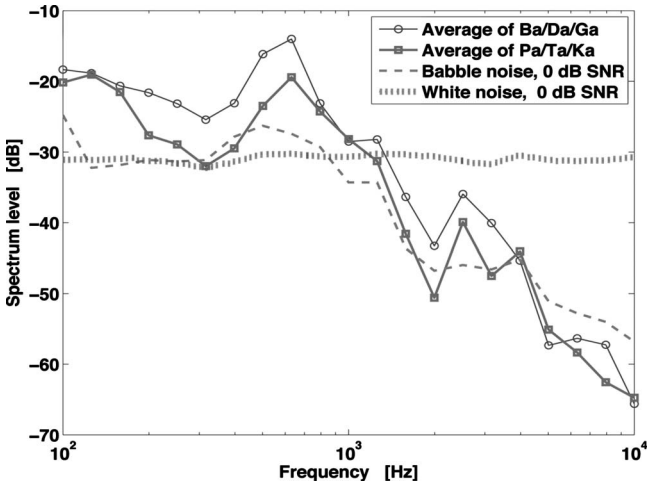


FIG. 1. Spectrum levels for the two categories of speech stimuli and the two noise types. For the speech stimuli, levels represent the average of the three voiced syllables and the three voiceless syllables, respectively, with a time window of 100 ms, starting at the burst. Levels are averaged across 1/3 octave bands. The level reference is arbitrary but the same for all curves.

Figure 1 illustrates spectrum levels for syllables and noise types for a 100 ms time window, starting at the consonant burst. For Norwegian stop consonants in initial position, aspiration is the most prominent voicing distinction, with unaspirated stops (/ba, da/ and /ga/) generally having a short voice onset lag, and aspirated stops (/pa, ta/ and /ka/) having a long voice onset lag (Halvorsen, 1998). These differences in VOT contribute to explaining why the average of the voiced syllables (/ba, da/ and /ga/) have a higher overall intensity than the average of the voiceless syllables (/pa, ta/ and /ka/) in this particular time segment. Figure 1 also shows that white noise has a flat spectrum, whereas for babble noise the spectrum decreases as frequency increases.

The 16 different AV syllables in five different auditory backgrounds resulted in a total of 80 different stimuli used in the experiment.

D. Procedure

The experiment was carried out in the Speech Laboratory at the Department of Psychology, NTNU. To mimic circumstances typical for face-to-face communication, participants were seated facing 17 in. monitors (1440 × 900 pixels) at approximately 50 cm distance. Audio signals were conveyed to AKG K271 stereo closed dynamic circumaural studio headphones, and the presentation level was fixed at 68 dBA (corresponding to a frontally incident free-field sound pressure level around 68 dBA) for all participants.

Participants were presented four stimulus blocks, each block containing the 80 different stimuli. Thus every distinct stimulus appeared four times during the experiment, but only once in a block. The stimuli in each block were randomized.

Participants were told to indicate which among the six alternative syllables (ba, da, ga, ka, pa, and ta) best corresponded to the syllable perceived. Because of possible per-

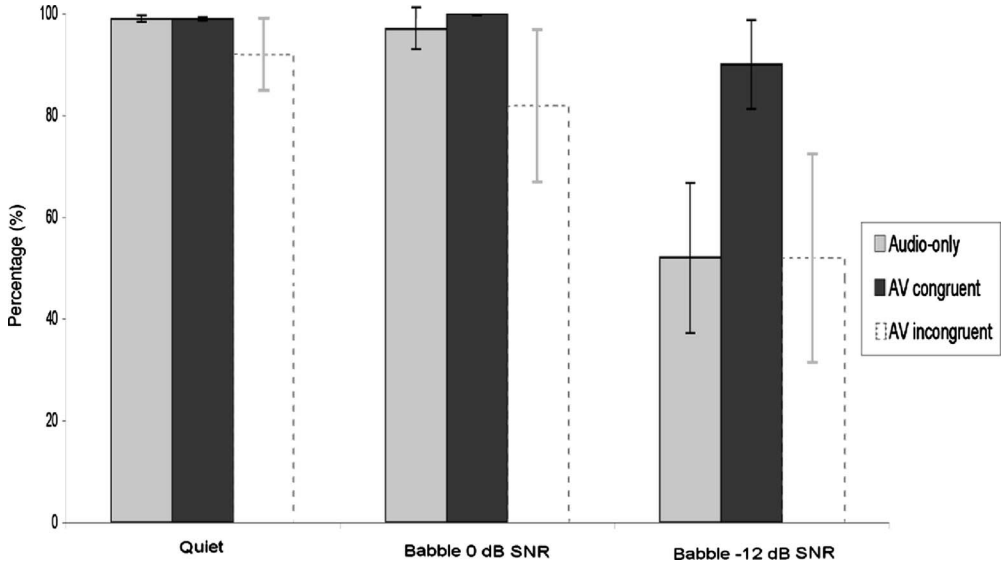


FIG. 2. Mean percent correct responses in the AO and AV congruent conditions in quiet and in babble noise at 0 and −12 dB SNRs. Mean percent auditory responses for all AV-incongruent stimuli in quiet and in babble noise at 0 and −12 dB SNR are also included. The AV- congruent and AV- incongruent responses were collected in the current study, whereas the AO responses were collected in a parallel study.

ceived ambiguity due to incongruent AV signals or noise, the participants were told that no wrong answers existed. The participants were frequently reminded to look at the talker's face throughout the entire duration of every clip to ensure that the participants received both auditory and visual input.

The experiment took approximately 1 h with a 10 min break included halfway.

III. RESULTS

A. Comparisons with audio-only and AV congruent stimuli

Audio-only (AO) and AV-congruent results were included in order to (1) establish whether the syllables used were good tokens of their respective categories, (2) address whether the babble noise unique in this study affects syllable perception in the absence of visual influence, and (3) assess whether responses to the AV-incongruent stimuli were likely to be based on chance.

Results from the AO condition were taken from a parallel study using the same stimuli as those in the current study. The two participant groups were recruited from the same university population, were comparable in age (current study: $M=24$, $SD=3$; parallel study: $M=23$, $SD=4$), and were equally balanced for gender. The two participant groups also had almost identical response patterns in the AV-incongruent conditions (e.g., mean percents A-match for POA stimuli were 44.4% and 45% [$F(1,43)=0.036$, $p>0.85$]).

Figure 2 shows percent correct responses from the AO and AV-congruent conditions collapsed for all syllables. AO results show that the auditory stimuli are good tokens of their respective categories, with participants getting nearly all responses correct in quiet and in babble noise of 0 dB SNR. The sharp drop off in correct responses at −12 dB SNR in the AO condition is greatly compensated for by relevant visual information in the AV-congruent condition. This indicates that the visual components are also good tokens of their respective categories.

Results from the AO condition further indicate that the auditory perception of syllables is robust in babble noise. Even at SNR of −12 dB participants responded well above chance, with 52% correct.

The AV-incongruent stimuli did not allow for correct or incorrect responses; the response options either matched the visual component, the audio component, or were intermediate to the two. In the AO and AV-congruent conditions, participants had a 17% (100% divided by six response alternatives) possibility of giving a correct response by chance. The high correct response percentage obtained for the AO and AV-congruent stimuli renders it unlikely that observed differences for the incongruent stimuli are due to chance responses.

Figure 2 also includes the overall percentage of auditory responses for all AV-incongruent stimuli in the current study. The pattern of responses shows that reliance on auditory cues is negatively influenced by incongruent visual cues. Compared to the AO condition, the introduction of incongruent visual information resulted in a decrease in the overall reliance on the auditory input by 5.5% in quiet and 10.7% in babble noise in the 0 dB SNR. In babble noise in the −12 dB SNR, the AV- incongruent and AO stimuli both received 52% audio responses, although the AV-incongruent mean had a larger variance ($SD=41$) than AO ($SD=29$). The strength and variation of this visual influence are considered in detail in the analyses of the AV-incongruent stimuli, where it is used as a means of assessing the relative influence of white and babble noise on AV perception of POA and voicing.

B. AV incongruent stimuli

The incongruent stimuli consisted of POA and voicing stimuli and are used to assess shifts in the contribution of

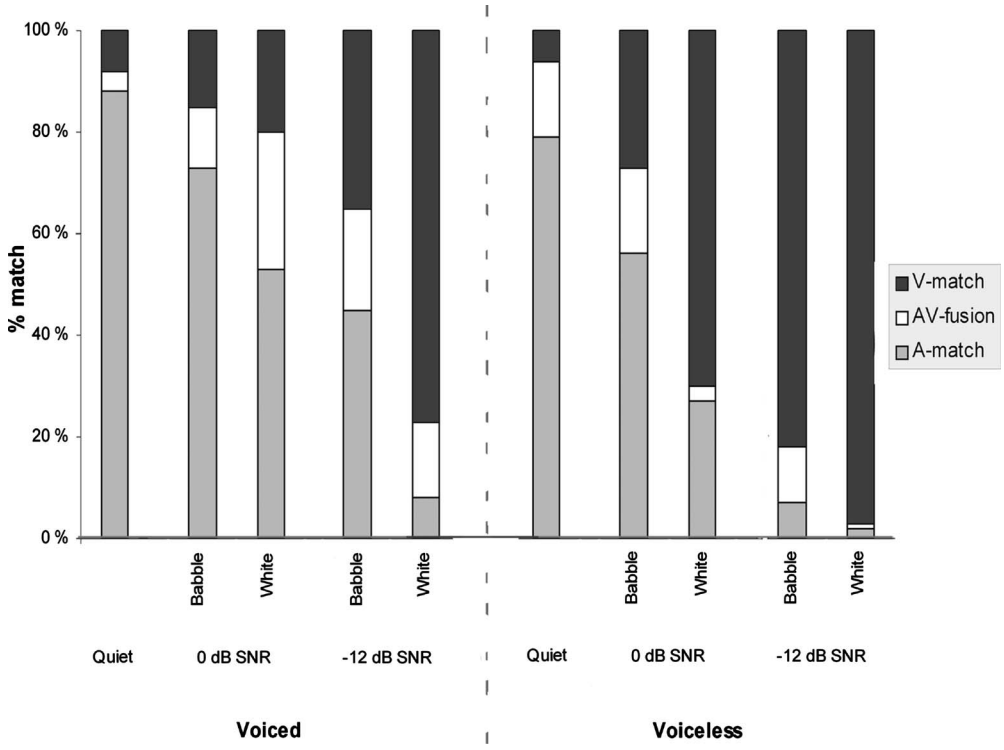


FIG. 3. Mean percent responses for A-match, V-match, and AV-fusion for incongruent POA stimuli for voiced and voiceless consonants presented in quiet and in babble and white noise in the 0 and −12 dB SNR conditions.

modalities in noise. For POA stimuli, responses were coded based on whether they matched the stimuli's audio component (A-match), visual component (V-match), or were intermediate to the audio and visual component (AV-fusion). For voicing stimuli, responses were coded based on whether they matched the stimuli's audio component (A-match) or video component (V-match). Match for a POA stimulus implied correspondence between response and stimulus in terms of POA, whereas match for a voicing stimulus implied correspondence between response and stimulus in terms of voicing; thus a "perfect" match (i.e., response matching stimulus for both POA and voicing) was not required for either POA or voicing stimuli.

1. Quiet condition

For the incongruent POA and voicing stimuli, the response patterns in the quiet conditions constitute baseline AV perception, and are considered points of reference for the effects of noise.

The results of the quiet condition for the incongruent POA stimuli are illustrated in Fig. 3. In quiet 84% of the responses matched the auditory component compared to an overall of 45% in babble, and 23% in white noise. Paired t-tests with Bonferroni corrected p-values revealed that participants gave significantly more A-match responses in quiet than in babble noise of 0 dB SNR ($p<0.004$), babble noise of −12 dB SNR ($p<0.004$), white noise of 0 dB SNR ($p<0.004$), and in white noise of −12 dB SNR ($p<0.004$). No significant differences between voiced and voiceless stimuli were obtained in the quiet condition.

For incongruent voicing stimuli presented in quiet, 100% of the responses match the audio component, a result identical to that obtained for white and babble noise at 0 dB SNR.

2. POA stimuli

Data for POA stimuli (Table I) were analyzed with three repeated measures analyses of variance (ANOVAs) where noise type (white and babble), noise level (0 and −12 dB SNRs), voicing (voiced and voiceless), and POA stimulus structures ($A_{\text{labial}}V_{\text{vclar}}$,$A_{\text{vclar}}V_{\text{labial}}$) were independent variables and A-match, V-match, and AV-fusion were dependent variables. Results from these analyses are reported in Table II. Results for POA stimulus structure are not discussed in the current article. *Post hoc* analyses (paired-samples t-tests) were performed using the Bonferroni correction for multiple comparisons and a 5% rejection level.

Figure 3 shows the percent A-match, V-match, and AV-fusion for voiced and voiceless stimuli in white and babble noise at 0 and −12 dB SNRs. The analyses show significant main effects of noise type, noise level, and voicing for A-match and V-match. A significant main effect for AV-fusion was only obtained for voicing. In addition, the interaction between these factors was significant for A-match, V-match, and AV-fusion (see Table II).

(a) Noise type. The quality of the background noise clearly influenced the participants' reliance on the audio and visual component. Figure 3 shows that participants generally relied less on the audio component in white noise ($M=23\%$, $SE=2.74$) than in babble noise ($M=45\%$, $SE=2.04$). *Post hoc* analyses revealed that white noise resulted in sig-

TABLE II. Statistical results for the three sets of repeated measures ANOVAs for the POA stimuli, showing main effects and interactions of noise type, noise level and voicing for A-match, V-match, and AV-fusion.

Place of articulation stimuli							
		Auditory match		Visual match		AV-fusion	
Factor(s)	*df*	*F*	*P*	*F*	*P*	*F*	*P*
Noise type	1,14	196.98	<0.001	80.71	<0.001	2.63	n.s
Noise level	1,14	117.09	<0.001	206.92	<0.001	2.01	n.s
Voicing	1,14	53.80	<0.001	135.52	<0.001	22.15	<0.001
Noise type × noise level	1,14	0.44	n.s	1.07	n.s	4.55	n.s
Noise type × voicing	1,14	13.13	<0.003	0.54	n.s	5.52	<0.034
Noise level × voicing	1,14	0.00	n.s	0.05	n.s	0.11	n.s
Noise type × noise level × voicing	1,14	14.30	<0.002	31.17	<0.001	13.13	<0.003

nificantly fewer A-match responses than babble noise for voiced ($p<0.024$) and voiceless stimuli ($p<0.012$) in the 0 dB SNR, and voiced stimuli in the −12 dB SNR ($p<0.012$), but not for voiceless stimuli in the −12 dB SNR.

The responses matched the visual component to a significantly greater extent in white noise ($M=66\%$, $SE=4.27$) than babble noise ($M=40\%$, $SE=3.55$). *Post hoc* analyses revealed no significant noise type effect for voiced stimuli in the 0 dB SNR, but white noise resulted in significantly more V-match responses than babble noise for voiceless stimuli in the 0 dB SNR ($p<0.012$) and for voiced stimuli in the −12 dB SNR ($p<0.012$).

(b) Noise level. Consistent with previous findings (e.g., Erber, 1969; MacLeod and Summerfield, 1987; Ross *et al.*, 2007), noise level greatly influenced the way participants utilized audio and visual cues. The increase in noise level from 0 to −12 dB SNR resulted in participant responses corresponding more with the visual component and less with the audio component, and this pattern was evident for all types of noise and voicing.

As illustrated in Fig. 3, responses match the audio component to a significantly greater extent in the 0 dB SNR ($M=52\%$, $SE=3.66$), than in the −12 dB SNR ($M=15\%$, $SE=1.64$). *Post hoc* analyses revealed that the 0 dB SNR resulted in more A-match responses than the −12 dB SNR for voiced ($p<0.012$) and voiceless stimuli in white noise ($p<0.012$), and for voiced ($p<0.012$) and voiceless stimuli in babble noise ($p<0.012$). The noise level effect on V-match was opposite that of A-match.

(c) Consonant voicing. As can be seen from Fig. 3, consonant voicing clearly influenced perception of the AV syllables. Surprisingly, voiced consonants resulted in more A-match responses ($M=45\%$, $SE=2.8$) than did voiceless consonants ($M=23\%$, $SE=2.65$). This result contrasts with previous findings (e.g., Behne *et al.*, 2006; McGurk and MacDonald, 1976). The most remarkable observation in this respect is the size of the difference in mean A-match responses between voiced and voiceless consonants, and the voicing effect's consistency across conditions. Voiced stimuli resulted in significantly more A-match responses than voiceless stimuli in white noise in the 0 dB SNR ($p<0.012$), in babble noise in the 0 dB SNR ($p<0.024$), and in babble noise in the −12 dB SNR ($p<0.012$), but not in white noise in the −12 dB SNR.

Responses match the visual component to a lesser extent for voiced consonants ($M=36\%$, $SE=3.76$) than for voiceless consonants ($M=69\%$, $SE=4.05$). Voiced stimuli resulted in fewer V-match responses than voiceless stimuli in white noise in the 0 dB SNR ($p<0.012$), in white noise in the −12 dB SNR ($p<0.036$), and in babble noise in the −12 dB SNR ($p<0.012$).

AV-fusion data replicate previous findings (e.g., Green and Kuhl, 1991; McGurk and MacDonald, 1976) demonstrating fewer AV-fusion responses for voiceless consonants ($M=9\%$, $SE=2.47$) than for voiced consonants ($M=19\%$, $SE=3.42$). However, the effect of consonant voicing shows that the degree of AV-fusion is dependent on noise type; that is, only white noise led to a significant voicing effect. *Post hoc* analyses showed reliably more AV-fusion responses for voiced consonants than voiceless consonants in white noise in the 0 dB SNR ($p<0.012$).

(d) Summary of POA stimuli results. In general participants gave fewer A-match and more V-match responses in white noise than in babble noise. The 0 dB SNR resulted in more A-match responses and fewer V-match responses compared to the −12 dB SNR. Considering consonant voicing, more A-match and AV-fusion responses and fewer V-match responses were given for voiced than for voiceless consonants.

3. Voicing stimuli

Data for voicing stimuli (Table I) were analyzed with two repeated measures ANOVAs where noise type (white and babble), noise level (0 and −12 dB SNRs), POA (labial, alveolar, and velar), and voicing stimulus structure ($A_{\text{voiced}}V_{\text{voiceless}}$,$A_{\text{voiceless}}V_{\text{voiced}}$) were independent variables and A-match and V-match were dependent variables. The results of these analyses are reported in Table III. Results for voicing stimulus structure are not discussed in the current article. Note that only two response categories (i.e., A-match and V-match) were available for voicing stimuli and any response had to fall into one of the two. Thus, when the percentage *A*-match is found, the percentage V-match is already known. A significant effect for A-match therefore implies a

TABLE III. Statistical results for the two sets of repeated measures ANOVAs for the voicing stimuli, showing main effects and interactions of noise type, noise level, and place of articulation for A-match and V-match,.

Voicing stimuli			
		Auditory or visual match	
Factor(s)	*df*	*F*	*P*
Noise type	1,14	33.88	<.001
Noise level	1,14	100.46	<.001
POA	2,28	17.01	<.001
Noise type × noise level	1,14	41.88	<.001
Noise type × POA	2,28	20.79	<.001
Noise level × POA	2,28	19.09	<.001
Noise type × noise level × POA	2,28	18.97	<.001

significant effect for V-match. Results are therefore only described in detail for A-match, although hereby implicating the opposite effect for V-match. *Post hoc* analyses (paired-samples t-tests) were performed using Bonferroni correction for multiple comparisons and a 5% rejection level.

Analyses show a main effect of noise type, noise level and POA, in addition to an interaction between noise type, noise level, and POA (see Table III). Figure 4 depicts these results.

(a) Noise level. In the 0 dB SNR condition, participants gave 100% A-match responses whereas, as Fig. 4 illustrates, significantly fewer responses matched the auditory component for the −12 dB SNR ($M=87\%$, $SE=1.27$). Figure 4 clearly reveals the robustness of voicing in noise and hence replicates previous research (e.g., Jiang *et al.*, 2006; Miller and Nicely, 1955). *Post hoc* analyses revealed that the 0 dB SNR resulted in significantly more A-match responses than the −12 dB SNR for alveolar stimuli in white noise ($p<0.024$) and for labial ($p<0.024$) and alveolar stimuli in babble noise ($p<0.024$).

(b) Noise type. As depicted in Fig. 4, responses matched the auditory component to a significantly greater extent in white noise ($M=97\%$, $SE=0.56$) than in babble noise ($M=90\%$, $SE=0.76$). No significant noise type differences were found for the 0 dB SNR where 100% A-match responses was observed across conditions. At the −12 dB SNR, however, *post hoc* analyses revealed that white noise resulted in significantly more A-match responses than babble noise for labial ($p<0.024$) stimuli.

(c) POA. As shown in Fig. 4, most A-match responses were observed for velar stimuli ($M=98\%$, $SE=0.6$), followed by alveolar ($M=93\%$, $SE=1.03$) and labial stimuli ($M=89\%$, $SE=1.48$). No significant POA differences were found in the 0 dB SNR. Whereas POA differences in white noise in the −12 dB SNR did not reach significance, *post hoc* analyses revealed significant POA differences for babble noise in the −12 dB SNR. Here, labial consonants resulted in fewer A-match responses than velar consonants ($p<0.024$), whereas alveolar consonants resulted in fewer A-match responses than velar consonants ($p<0.024$).

(d) Summary of voicing results. In general significantly more A-match responses were obtained in the 0 dB SNR than in the −12 dB SNR. White noise generally resulted in more *A*-match responses than did babble noise. For POA, fewer A-match responses were found for labial stimuli than for alveolar and velar stimuli.

IV. DISCUSSION

The present study used white and babble noise to test AV perception of POA and voicing of stop consonants. POA and voicing stimuli were used because these phonetic attributes are known to be differentially susceptible to noise level (e.g., Jiang *et al.*, 2006; Miller and Nicely, 1955) and thus possibly unequally susceptible to noise type.

Results from the POA stimuli revealed that POA identification was more susceptible to white noise than babble noise, whereas voicing stimuli showed that voicing identifi-

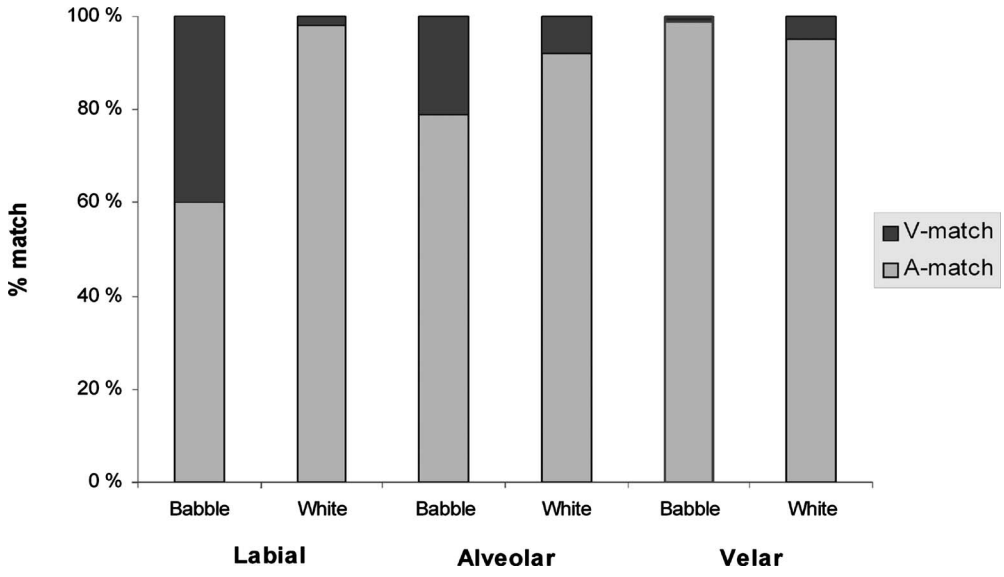


FIG. 4. Percent A-match and V-match responses for labial, alveolar, and velar stimuli, presented in babble and white noise in the −12 dB SNR.

cation was more susceptible to babble noise than white noise. When the POA of the visual and audio components differed, white noise resulted in significantly fewer auditory responses (23%) than did babble noise (45%). POA's susceptibility to white noise is consistent with previous research (Jiang *et al.*, 2006; Miller and Nicely, 1955) and supports the notion of a "glimpsing effect;" that is, white noise occludes the target signal more than babble noise because it allows fewer acoustical windows open to the perceiver (e.g., Carhart *et al.*, 1969; Cooke, 2006; Miller and Lickider, 1950). White noise has a flat intensity across frequencies and covers a wider frequency range than babble noise, and the speech signal is consequently less perceptually accessible in white noise than in babble noise.

The analysis of noise type effect on voicing revealed an interesting new finding. For voicing stimuli, the responses matched the auditory component to a greater extent for white noise than for babble noise. Previous research with AO stimuli has shown that perception of voicing is very robust in white noise and is nearly unaffected up to the threshold −15 dB SNR (Jiang *et al.*, 2006). The results for white noise stimuli in the current study replicate these findings, with participants almost solely responding consistent with the auditory component (97%). However, for babble noise, only 90% of the responses matched the auditory component. This difference is substantial considering that voicing identification is known to be very robust in noise (Jiang *et al.*, 2006; Miller and Nicely, 1955). The greater negative effect of babble noise on auditory responses challenges whether the glimpsing effect (i.e., the proportion of the frequency range which is masked) can fully account for the effect of noise.

The surprising difference between the effects of white and babble noise on POA and voicing identification can be considered from the perspective of peripheral masking. Peripheral masking takes place when noise and a target signal overlap in time and frequency, making parts of the target signal inaudible (Watson and Kelly, 1981). "Peripheral" denotes that this kind of interference distorts the clarity of the target signal on the auditory nerve and refers to the acoustical characteristics of the target signal and the noise.

Both phonetic attributes (i.e., POA and voicing) and noise types (i.e., white and babble noise) differ in acoustical characteristics. The formant transitions between the consonant and the vowel in a consonant-vowel syllable yield important cues for identifying POA (Blumstein *et al.*, 1982; Stevens and Blumstein, 1978) and especially the F_2 and F_3 transition (Delattre *et al.*, 1955; Liberman *et al.*, 1954). An essential cue for identification of voicing is VOT (e.g., Eimas and Corbit, 1973; Lisker and Abramson, 1964; Massaro and Oden, 1980), defined as the time that elapses between consonant release and when the vocal folds begin to vibrate (Lisker and Abramson, 1964). The aspiration associated with VOT is characterized by a flatter power spectral density than the formant transition, which in turn is characterized by high intensity levels at the frequencies associated with formants. Consequently, the transition between the aspiration and the voice onset is characterized by the signal shifting from a relatively flat intensity distribution across frequencies (aspiration) to a more fluctuating intensity distribution across frequencies (formants in connection with voice onset).

Babble noise has a fluctuating intensity with a low-frequency dominated spectral shape, while white noise is stationary and has a flat spectral shape, as illustrated in Fig. 1. The fluctuating nature of the babble implies that its momentary level might vary by several decibels above or below the average level, whereas the intensity of the white noise varies little over time.

In identification of voicing, babble noise may interfere more with the target signal than white noise because high energy densities at low frequencies obscure the transition between the aspiration and the voice onset in the target signal; that is, babble noise makes it harder to discern when the target signal shifts from relatively flat intensity (aspiration) to a more fluctuating intensity (voice onset) by adding fluctuation to the aspiration. The onset of F_1 (0–1000 Hz) may, for example, be hard to distinguish in babble noise, because the babble noise has high energy density around the same frequency band as F_1. Babble noise may thus obscure temporal differences in VOT between voiced and voiceless signals by making it harder to perceive the point in time when the aspiration stops and the voicing starts.

Considering the effect of POA on perception of voicing stimuli, an interesting pattern emerges. First, voicing is robust in noise and hence only the −12 dB SNR revealed a significant POA effect. Whereas babble noise revealed significant differences in use of modality between labial, alveolar, and velar stimuli, no such differences were obtained for white noise. To the authors' knowledge, an effect of POA on use of modality in perception of voicing has not previously been reported, although some research (e.g., Lisker and Abramson, 1970) indicates an interaction between POA and voicing. In the current study labial consonants resulted in more visual responses than alveolar and velar consonants, respectively, in babble noise of −12 dB SNR. Research (e.g., Bengherel and Pichora-Fuller, 1982; Walden *et al.*, 1977) shows that labial consonants are more visually salient than velar consonants, since production of labial consonants provokes more prominent facial movements than velar consonants. Furthermore, research (Cooper *et al.*, 1952; Liberman, 1952) has demonstrated that the spectrum of the stop burst influences the way POA is perceived. The burst spectrum of labial stops has low frequency dominance, whereas the bursts of alveolar and velar stops have high and midfrequency dominance, respectively. Babble noise is characterized by high spectral densities at low frequencies and thus occlude the burst of labial stops more than alveolar and velar stops.

In the current study overwhelmingly more auditory responses were given when the POA stimuli were voiced compared to voiceless. For voiced stimuli 45% of the responses were auditory compared to only 23% for voiceless stimuli. This effect was consistent across noise conditions; that is, voiced POA stimuli resulted in more auditory responses than voiceless POA stimuli in both noise types and in both noise levels.

How may voicing change the perception of POA? Jackson (2001) showed that at voice onset the auditory distinc-

tiveness between voiced stop consonants is more prominent than it is for voiceless stop consonants. It may be argued that voiced stops are more auditorily prominent than voiceless stops because the production of voiced stops is characterized by vocal activity, absent in word-initial aspirated voiceless stops. The spectral distribution of aspiration in voiceless stops is in many respect reminiscent of white noise. White noise conditions may therefore pose a more serious threat to accurate auditory perception of voiceless syllables than it does to voiced syllables, since the spectral similarity between the white noise and the target stimuli enhances the masking effect (i.e., peripheral masking) (e.g., Watson and Kelly, 1981). Indeed, fewer responses matched the voiceless auditory stimuli in conditions with white noise than babble noise. This was especially true for the 0 dB SNR, where only 27% of the responses matched the auditory component in white noise, while 53% of the responses matched the auditory component in babble noise.

V. CONCLUDING REMARKS

The contribution of the auditory and visual modality to speech perception is highly dependent on noise type and phonetic attributes. The two modalities are used to a different degree in white and babble noise, for voiced and voiceless consonants and for labial, alveolar, and velar consonant. The current study took AVSP research further by assessing the effect of babble noise on speech perception. Previous research has mainly focused on white noise, although babble noise arguably is a more common distractor in natural speech communication. The results for noise type revealed that voicing identification is more susceptible than predicted from research based solely on white noise. Surprisingly, and contrary to the findings for POA identification, the identification of voicing is susceptible to babble noise, but not to white noise. The results also indicate interdependency between voicing and POA in babble noise; that is, the use of modality in identification of voicing depends on the POA of the target signal. The difference between white and babble noise found in this study is most likely attributed to peripheral masking.

This study supports previous research in demonstrating the robustness and flexibility of human speech perception and that these qualities are much related to its natural bimodal character. An integrated, acute, and highly adaptable system of AVSP is demonstrated by the way changes in the speech signal and acoustical context affect the contribution of auditory and visual information to speech perception.

ACKNOWLEDGMENTS

The authors wish to thank Peter Svensson for his assistance with the acoustic analyses and for his judicious comments. They would also like to thank the Associated Editor Allard Jongman and the anonymous reviewers for their insightful comments and suggestions.

Behne, D., Wang, Y., Alm, M., Arntsen, I., Eg, R., and Valsø, A. (**2006**). "Audio-visual speech perception with age in quiet and café noise," poster presented at the Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, HI.

Behne, D., Wang, Y., Alm, M., Arntsen, I., Eg, R., and Valsø, A. (**2007**). "Changes in audio-visual speech perception during adulthood," in Proceedings of the International Conference of Audio-Visual Speech Processing (AVSP), Hilvarenbeek, The Netherlands.

Bengherel, A. P., and Pichora-Fuller, M. K. (**1982**). "Coarticulation effects in lipreading," J. Speech Hear. Res. **25**, 600–607.

Binnie, C. A., Montgomery, A. A., and Jackson, P. L. (**1974**). "Auditory and visual contributions to the perception of consonants," J. Speech Hear. Res. **17**, 619–630.

Blumstein, S., Isaacs, E., and Mertus, J. (**1982**). "The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **72**, 43–50.

Boersma, P., and Weenink, D. (**2006**). "Praat: Doing phonetics by computer (Version 4.3.22)," from http://www.praat.org/ (Last viewed February 6, 2006).

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Cooke, M. P. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cooper, F., Delattre, P., Liberman, A., Borst, J., and Gerstman, L. (**1952**). "Some experiments on perception of synthetic speech sounds," J. Acoust. Soc. Am. **24**, 597–606.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1955**). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. **27**, 769–773.

Dodd, B. (**1977**). "The role of vision in the perception of speech," Perception **6**, 31–40.

Easton, R. D., and Basala, M. (**1982**). "Perceptual dominance during lip reading," Percept. Psychophys. **32**, 562–570.

Eimas, P. D., and Corbit, J. D. (**1973**). "Selective adaptation of linguistic feature detectors," Cognit Psychol. **4**, 99–109.

Erber, N. P. (**1969**). "Interaction of audition and vision in the recognition of oral speech stimuli," J. Speech Hear. Res. **12**, 423–425.

Fixmer, E., and Hawkins, S. (**1998**). "The influence of quality of information on the McGurk effect," in Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP), edited by D. Burnham, J. Robert-Ribes, and E. Vatikiotis-Bateson, Terrigal, Australia.

Green, K. P., and Kuhl, P. K. (**1989**). "The role of visual information in the processing of place and manner features in speech perception," Percept. Psychophys. **45**, 34–42.

Halvorsen, B. (**1998**). "Timing relations in Norwegian stops," thesis, University of Bergen.

Jackson, P. J. B. (**2001**). "Acoustic cues of voiced and voiceless plosives for determining place of articulation," in Proceedings Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis, CRAC 2001, Aalborg, Denmark, pp. 19–22.

Jiang, J., Chen, M., and Alwan, A. (**2006**). "On the perception of voicing in syllable-initial plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105.

Lansing, C. R., and McConkie, G. W. (**1999**). "Attention to facial regions in segmental and prosodic visual speech perception tasks," J. Speech Lang. Hear. Res. **42**, 526–539.

Liberman, A., Delattre, P., and Cooper, F. (**1952**). "The role of selected stimulus variables in the perception of unvoiced stop consonants," Am. J. Psychol. **65**, 497–516.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (**1954**). "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," Psychol. Monogr. **68**, 1–13.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lisker, L., and Abramson, A. S. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," in Proceedings of the Sixth International Congress of Phonetic Sciences, Prague (Academia, Prague), pp. 563–567.

MacDonald, J., and McGurk, H. (**1978**). "Visual influences on speech perception process," Percept. Psychophys. **24**, 253–257.

MacLeod, A., and Summerfield, Q. (**1987**). "Quantifying the contribution of vision to speech perception in noise," Br. J. Audiol. **21**, 131–141.

MacLeod, A., and Summerfield, A. Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.

Markides, A. (**1989**). "Background noise and lip-reading ability," Br. J. Audiol. **23**, 251–253.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Erlbaum, Hillsdale, NJ).

Massaro, D. W., and Light, J. (**2004**). "Using visible speech to train percep-

tion and production of speech for individuals with hearing loss," J. Speech Lang. Hear. Res. **47**, 304–320.
Massaro, D. W., and Oden, G. (**1980**). "Evaluation and integration of acoustic features in speech perception," J. Acoust. Soc. Am. **67**, 996–1013.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Miller, G. A., and Lickider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (**2007**). "Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments," Cereb. Cortex **17**, 1147–1153.
Sommers, M. S., Spehar, B., and Tye-Murray, N. (**2005**). "The effect of signal-to-noise ratio on auditory-visual integration: Integration and encoding are not independent," J. Acoust. Soc. Am. **117**, 2574.
Stevens, K., and Blumstein, S. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, Q., and McGrath, M. (**1984**). "Detection and resolution of audio-visual incompatibility in the perception of vowels," Q. J. Exp. Psychol. **36A**, 51–74.
Walden, B. E., Prosek, R. A., and Montgomery, A. A. (**1977**). "Effects of training on the visual recognition of consonants," J. Speech Hear. Res. **20**, 130–145.
Wang, Y., Behne, D., and Jiang, H. (**2008**). "Linguistic experience and audio-visual perception of non-native fricatives," J. Acoust. Soc. Am. **124**, 1716–1726.
Watson, C. S., and Kelly, W. J. (**1981**). "The role of stimulus uncertainty in the discrimination of auditory patterns," in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howards (Erlbaum, Hillsdale, NJ), pp. 37–59.

Left hand finger force in violin playing: Tempo, loudness, and finger differences

Hiroshi Kinoshita[a] and Satoshi Obata
Biomechanics and Motor Control Laboratory, Graduate School of Medicine, Osaka University, 1-17, Machikaneyama, Toyonaka, Osaka 560-0043, Japan

(Received 3 April 2008; revised 28 April 2009; accepted 29 April 2009)

A three-dimensional force transducer was installed in the neck of a violin under the A string at the D5 position in order to study the force with which the violinist clamps the string against the fingerboard under normal playing conditions. Violinists performed repetitive sequences of open A- and fingered D-tones using the ring finger at tempi of 1, 2, 4, 8, and 16 notes/s at *mezzo-forte*. At selected tempi, the effects of dynamic level and the use of different fingers were investigated as well. The force profiles were clearly dependent on tempo and dynamic level. At slow tempi, the force profiles were characterized by an initial pulse followed by a level force to the end of the finger contact period. At tempi higher than 2 Hz, only pulsed profiles were observed. The peak force exceeded 4.5 N at 1 and 2 Hz and decreased to 1.7 N at 16 Hz. All force and impulse values were lower at softer dynamic levels, and when using the ring or little finger compared to the index finger. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3139908]

PACS number(s): 43.75.De, 43.75.St [NHF] Pages: 388–395

I. INTRODUCTION

When playing bowed string instruments (violin, viola, cello, and double bass), the left hand fingers press the string against the fingerboard to control pitch by temporarily shortening the speaking length of the string. During this string-pressing motion, the force exerted by the finger can be divided in two components. First the finger force has to overcome the transversal reaction force due to string tension when the string is pressed down. The *string reaction force* is here defined as the force needed to press the string down so it barely touches the fingerboard. It is a property of the instrument, slowly varying from note to note. The *string clamping force* is the additional force applied by the player to press the string against the fingerboard in order to establish a stable string termination. Askenfelt and Jansson (1992) measured string reaction forces on the violin between 1.5 and 2.0 N at the middle of the string, increasing with about a factor of 2 toward the nut and the upper end of the fingerboard. The string reaction force will vary somewhat with bow force since it lowers the string to some extent. Typical values of bow force of 0.5–2.0 N has been reported in violin playing (Askenfelt, 1989). The effect on the reduction in string height was not reported.

Askenfelt and Jansson (1992) noted that the string clamping force had not been measured previously. However, they hypothesized that professional players would use little excessive force for clamping the string once the string has begun vibrating. The idea of a light pressure, sufficient to hold down the string, is also recommended by violin pedagogues (e.g., Flesch, 1930; Galamian, 1962).

Baader *et al.* (2005) recently examined the kinematics of left finger movement while violinists of varying skill levels performed an exercise on the D string. In all subjects they identified a close temporal coupling between the movements of the left hand fingers and the right hand, controlling bow motion. Interestingly, the timing of the moment of fingerboard contact relative to the start of a new bow stroke at bow changes varied among violinists, possibly reflecting a difference in the fingering strategy among performers. They also found that left hand fingers started to move almost one tone in advance of the production of the target tone, indicating that fingering motions are made in an anticipatory fashion. A search of literature revealed no other scientific studies that addressed the issues of left hand fingering motion and associated forces.

There are at least three reasons why an understanding of the forces exerted by the left fingers in string playing is important. First, the control of pitch requires not only a precise positioning of the fingertip on the string. In order to quickly achieve Helmholtz motion, a force must rapidly be applied by the finger, which is sufficiently high to swiftly bring the string down to the fingerboard, and establish a firm string-fingerboard contact The string clamping force will in principle need to be increased when string amplitude and bow force is increased in loud playing, in order to secure a well-defined string termination. Among other things, the static transversal force on the string in the bowing direction scales with bow force. The timing and magnitude of the string clamping force therefore can be an important source of information for understanding players' control of sound and timbre.

Second, the impulse delivered to the string and fingerboard represents the mechanical work of the left hand required for sound generation, and relates to the efficiency of performance. Third, values of peak force as well as impulse can be linked to the level of mechanical stress to which the players' left hand is constantly exposed. Neuromuscular and skeletal disorders in violinists and violists are found twice as

[a]Author to whom correspondence should be addressed. Electronic mail: hkinoshita@moted.hss.osaka-u.ac.jp

 0001-4966/2009/126(1)/388/8/$25.00

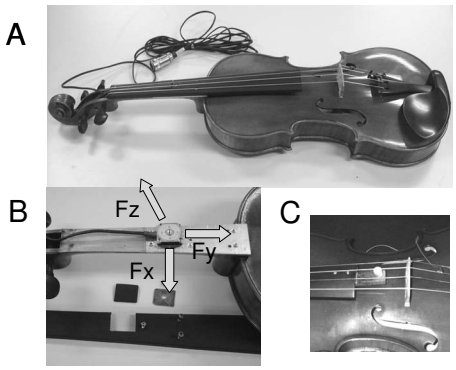


FIG. 1. Photos showing (a) the experimental violin, (b) force transducer after the fingerboard is removed, and (c) a small magnet fixed at the end of the fingerboard for string velocity measurement. The force transducer provided the mediolateral (Fx), longitudinal or fore-aft (Fy), and vertical (Fz) components of the string clamping force. A small piece ($18\times18\times2$ mm^3) was cut from the fingerboard and attached to the force sensor to measure the string clamping force.

often in the left than as in the right hand (Brandfonbrener, 1990; Lederman, 2006). Information regarding the string clamping force could thus help in understanding the etiology of playing-related problems.

In order to record and evaluate the string clamping force when playing the violin, we installed a three-dimensional (3D) miniature force transducer in the neck of an ordinary violin. Using this violin, a series of studies was conducted to investigate how string clamping force was influenced by variables set by the score and by players' individual performance techniques. In the present study, we examined the effects of tempo, sound pressure level, and the fingers used during simple exercises.

II. METHODS

A. Subjects

Eleven young Japanese violin players [9 females and 2 males, mean age ± standard deviation (SD)=23.5± 4.3 years] with more than 14 years of violin training under the guidance of professional violin teachers served as subjects in the experiments. All subjects had started their violin training before the age of 5 years. Five of them were professional solo performers and/or members of professional symphony orchestras. The others were either undergraduate or graduate students from professional performer training courses in the violin at the Department of Music at Japanese universities. Recruitment of the student players was based on their background (at least several participations in domestic and/or international violin competitions) and recommendations by their university violin teachers. All subjects reported that they were right-handed, and had not experienced serious playing-related physical problems in the past. Informed consent was obtained from all subjects, and the study was approved by the University Ethics Committee.

B. Experimental violin

A classical 4/4-size Yamaha violin (model: V10G) with a set of Thomastik-Infeld violin strings (type: blue)[1] was used [see Fig. 1(a)]. The string reaction force for note D5 played on the A string was assessed to be 2.2 N.[2] The reduction in string reaction force by the application of bow force was estimated to be around 0.07 N at *piano* and 0.15 N at *forte* levels.[3]

The violin was equipped with a 3D miniature force transducer (model: USL06-H5-50N-C, Tec-gihan Co., Kyoto, Japan) screwed onto a 2-mm thick duralumin plate, which in turn was fixed to the flat surface of the neck (fingerboard removed); see Fig. 1(b). The dimensions of the transducer were $20\times20\times5$ mm^3 (mass 3 g without wire). Maximum measurement ranges were 50 N in the vertical and 25 N in the lateral and longitudinal directions [see the definition of the force directions in Fig. 1(b)]. The force resolution was 0.03 N in the vertical direction and 0.015 N in the lateral and longitudinal directions. The non-linearity and hysteresis were less than 1% in all directions. The wire from the transducer was located under the fingerboard and extended from the peg box (see Fig. 1).

A thin duralumin plate ($18\times18\times1.5$ mm^3) was screwed to the force-detecting facing of the transducer. A piece cut from the fingerboard ($18\times18\times2$ mm^3) was then glued to the plate. The thickness of this small piece was fine-adjusted so that its slightly curved surface was flush with the fingerboard surface. The center of the force transducer was located below the A string at a distance of 78 mm from the nut, which corresponded to the position of D5. The force signals were amplified using a three-channel strain gauge amplifier (model: DSA-03Am Tec-gihan Co., Kyoto, Japan) and stored on a computer via a 12-bit analog/digital (A/D) converter sampling at a frequency of 2 kHz.

In order to monitor transverse string velocity, a small cobalt magnet (7 mm in diameter and 9 mm in thickness) was attached to the violin, located 2 mm under the A string and 30 mm from the bridge (approximately at the bowing point). The magnet was glued on an edge of a small wooden stick ($8\times10\times60$ mm^3), attached to the fingerboard [see Fig. 1(c)]. The two ends of the A string were connected to the A/D converter, and the induced signal was amplified (×20) and recorded on the computer along with the force signals at a sampling rate of 10 kHz.

C. Experimental room and sound recording

A Nihon-Koden soundproof and shielded-chamber (2.4 m high, 2.4 m wide, and 4.2 m long) with an ambient noise level of around 38 dB (sound pressure level) was used for the entire experiment. The reverberation time of the chamber was around 0.17 s at 0.5 kHz, and 0.13 s at 1 kHz. Radiated violin sound was sampled using a sound-level meter (model: NA-27, RION Co., Japan) placed approximately 1 m from the center of the top plate of the violin. The height of the sound-level meter was adjusted approximately to the level of the violin strings for each subject. The sound-level signal was stored in the computer at 2 kHz, synchronized with force data. In addition, the audio signal from the sound-level meter was monitored on line to provide the subject with feedback on their sound level.

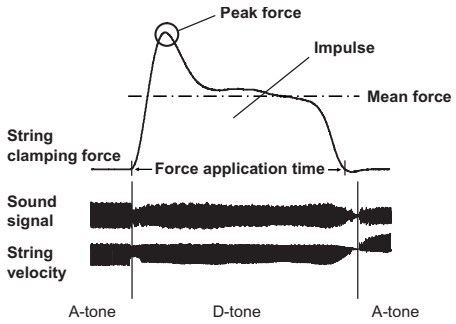


FIG. 2. A representative example of time-history curve for the string clamping force, sound signal, and string velocity during the repetitive A-D tone production task performed at a 2-Hz tempo at *mf*. Parameters evaluated are also shown, including force-application time, peak and mean values of string clamping force, impulse, and sound and string velocity signals.

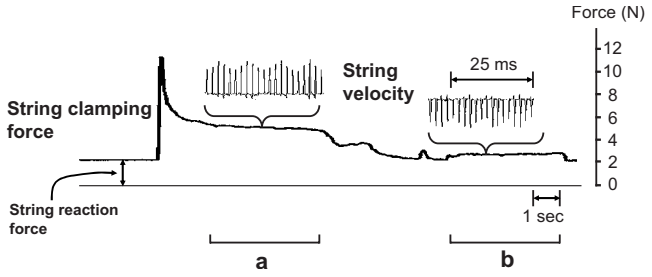


FIG. 3. String clamping force and string velocity during a trial to detect the minimum clamping force. The quality of the tone during the minimum force production part (phase b) was evaluated from the string velocity signal (periodic slip/stick triggering condition) and compared to that during the initial period (phase a) with steady-state Helmholtz motion before accepting the data for subsequent analysis

D. Experimental tasks

All subjects performed four sets of experimental tasks. These were (1) repetitive A-D tone production (alternating open string—fingered string with force recording), (2) melody performance, (3) minimum string clamping force measurement, and (4) maximum finger force measurement. All tasks were performed using the experimental violin, with the subjects seated in a chair facing the sound-level meter. To perform the first three tasks, each subject used a bow brought by him or herself. The tasks were performed with alternating up-strokes and down-strokes at a 1-Hz pace, using three-fourth of the bow hair length to keep the bow speed as constant as possible across all tasks (about 50 cm/s).The bow-bridge distance was maintained constant at around 30 mm from the bridge during all tasks.

1. Repetitive A- and D-tone production task

In the first set of experimental tasks, each subject performed 20 repetitions of alternating A- and D-tones at five predetermined tempi (1, 2, 4, 8, and 16 Hz or notes/s; nominal durations 1000, 500, 250, 125, and 63 ms) at a medium sound pressure level (mezzo-forte=75–77 dB). An example of registrations from this task, including string clamping force and sound and string velocity signals, is shown in Fig. 2. The tempi corresponded to quarter, 8th, 16th, 32nd, and 64th notes with quarter notes played at 60 beats/min (see the scores in Fig. 4). At 1, 4, and 16 Hz, the subjects also performed the same task at lower (p=70–72 dB) and higher (f=80–82 dB) sound pressure levels. These levels were determined based on the results of preliminary tests with the experimental violin performed by three violinists.

The left hand was basically placed in the first position, using the ring finger for playing the D-tone. To examine finger differences, three other cases using the index finger in the third position, the middle finger in the second position, and the little finger in the first position were also tested for selected combinations of tempi (2 and 8 Hz) at *mf* level. In total, 17 combinations of tempi, sound level, and finger were included in this first task, which all were repeated 20 times.

Instructions given to subjects regarded the tempo, dynamic level, finger used, and to play all tones without vibrato. No specific instructions concerning how fast the string should be pressed down or the angle of finger relative to string or fingerboard were given.

2. Melody tasks

In the second set of tasks, each subject played a short etude at a 1-Hz tempo and a musical excerpt from a Japanese nursery rhyme at a 2-Hz tempo. Both were played four times at *mf* sound level without vibrato. In each piece, the D following the open A was played using the ring finger eight times. These tones were written as quarter and eight notes in the 1-Hz etude and 2-Hz melody, respectively.

3. Minimum string clamping force measurement

In order to assess the excess force used in the first two tasks for clamping the string against the fingerboard above the minimum force required for producing the D-tone with normal quality, the minimum string clamping force needed to be determined in a separate experiment. This was done for the ring finger at p, *mf*, and f levels for all subjects. While bowing continuously at 1 Hz, the subjects started to generate the D-tone at the target sound pressure level (Fig. 3). They then searched for the minimum string clamping force by increasing or decreasing the force while continuously judging the tone quality. After reaching the minimum force, which still produced a normal tone quality, the subjects maintained the tone for 3–4 s while keeping this minimum force.

The quality of the string motion at the minimum force was examined by inspection of the string velocity signal. Guettler and Askenfelt (1997) reported that string players are quite sensitive to the noise produced when the slipping intervals deviate substantially from the nominal periodicity in the onset transient in bowed string attacks. Applying this criterion on the quasi-steady-state part of string tones in our experiments, the regularity of the slip-to-slip interval and appearance of irregular slips during the last 4 s of the minimum force production period were visually inspected, and compared with the waveform during the initial part with higher bow force and normal Helmholtz motion (see phases "a" and "b" in Fig. 3). The minimum force production task was accepted only if the string motion was normal (regular Helmholtz motion with one slip per period) during the last 4 s.

The mean value of the sound pressure level during the corresponding period was also computed to judge if the dynamic level was in the acceptable interval. The subjects repeated the trials until they could produce acceptable data in three trials at the predetermined ranges corresponding to *p*, *mf*, and *f* levels. The mean force during a 1-s interval at the middle of the last 4 s of the minimum string clamping force production period was computed. The minimum of the three determined mean values was taken as the minimum string clamping force for that subject.

4. Maximum finger force measurement

After the measurement of the minimum string clamping force, the maximum isometric voluntary force for each finger was measured in all subjects when pressing the fingerboard at the D-tone position. This task was performed on another violin (4/4-size Carlo Giordano violin, model: VS-2s) with a 3D miniature force transducer (model: USL06–100N-C, Tecgihan Co., Kyoto, Japan). This transducer could measure a maximum of 100 N in the vertical direction and 50 N in the lateral and longitudinal directions. All strings were removed to permit the subjects to apply the force by the target finger directly on the fingerboard above the transducer. The subjects were asked to press with the index, middle, ring, and little fingers for 3 s each with maximal force, keeping a similar finger posture for all fingers. The maximum force data were collected three times for each finger with adequate rests between measurements, and the highest value was used as the maximum finger force for each finger.

E. Experimental procedure

A few weeks before the experiment, each subject was carefully briefed on the purpose of the study, the tasks and experimental conditions (tempi, dynamics, and fingers), and test procedure. The scores were also shown to the subjects. On the experiment day, the tasks were again explained, and then each subject participated in a practice session for about 30 min to acclimate to the experimental violin and tasks. The subjects were instructed to play as usual, keeping the speed of finger attack for pressing down the string and the magnitude of the pressing force against the fingerboard as in their normal playing. Pre-recorded violin sounds at the target sound pressure levels were presented from a speaker placed in front of the subject. The subject then performed the task at the same level. With the help of an experimenter providing feedback on differences in the produced and presented sound pressure levels, each subject practiced the task until he/she could reduce the error to within ± 1 dB of 71, 76, and 81 dB at *p*, *mf*, and *f*, respectively. Help in maintaining the specified tempo was given by a silent metronome with a light emitting diode flashing at 1 Hz in front of the subject.

After the practice session, subjects performed the first experimental tasks (the repetitive A-D tone production). The order of all conditions was randomized for each subject. The subjects then performed the remaining three tasks: melody, minimum string clamping force measurement, and maximum finger force measurement. After completion of these, each subject was requested to answer a set of questions concerning their violin training background and playing habits. The time required to complete the experiment was typically 90 min, with an adequate rest between tasks.

At the end of the experiment each subject was interviewed, answering questions related to the experimental violin, the experimental tasks, and normal musical performance. Questions were asked to assess whether they had to change their fingering action substantially from playing their own violin compared to the experimental violin, and what their perceptions were about the magnitude of the force applied by the fingers at different tempi, loudness, and finger used. Subjects were also asked about the differences between fingers regarding the ease of the fingering action.

F. Parameters evaluated

The string clamping forces in the three directions, the string velocity, and the sound were stored on the computer and later analyzed off-line. From the three directional components of force, the resultant force vector was calculated by taking the root of the squared sum of the components. This resultant force was used as the string clamping force in the subsequent analysis.

For evaluation of the string clamping force for the A-D tone production and melody tasks, one temporal and three string clamping force variables were selected (see Fig. 3). These were (1) force-application time, which was the duration between the onset (>0.03 N) and termination (<0.03 N) of the string clamping force; (2) peak force; (3) average force during the force-application time; and (4) a total value of the impulse as defined by integration of the string clamping force during the application time. The threshold criterion of 0.03 N for force onset and termination was based on the resolution of the force sensor.

All variables were computed for each finger. For each variable, the mean of the data from the 20 repetitions by each subject was calculated. For the second set of tasks, the means of the eight force values from the four repetitions of the etude and the melody, respectively, were computed.

G. Statistical tests

Using the mean data for each subject for each experimental task, one-way or two-way analysis of variances (ANOVAs) with repeated measures were performed depending on the purpose of the comparison. Significance was accepted at $p<0.05$.

III. RESULTS

A. String clamping force at different tempi

Figure 4 shows representative time-history curves for string clamping forces and string velocity signals for one subject at different tempi at *mf* level, when the ring finger was used. At the slower tempi of 1 and 2 Hz, the curves show an initial force pulse with a distinct peak, followed by a level force at a lower magnitude. This level force lasts until near the end of the duration of the note. At tempi faster than 4 Hz, only a pulsed force profile appears. At the fastest

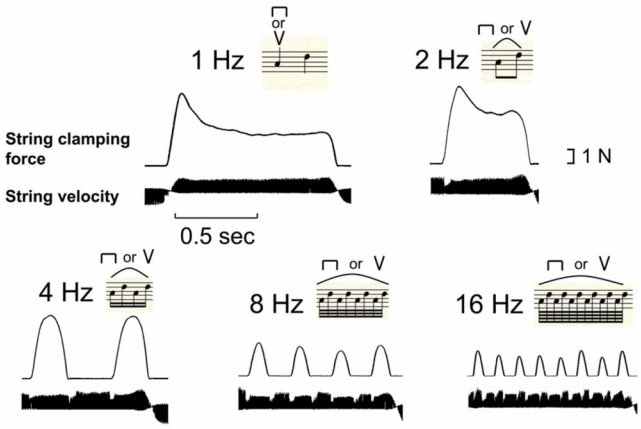


FIG. 4. (Color online) Representative time-history curves of the string clamping force and string velocity from one trial by one subject performed at each of the five tempi examined. Scores of the experimental tasks with notation of the bowing direction at each tempo are also shown. The finger used was the ring finger and the dynamic level *mf*.

tempo of 16 Hz, the force exhibits a continuous wave-like pattern with markedly lowered peaks. Similar force profiles were observed for all subjects.

The mean values of the peak and average forces across all subjects at each tempo were computed. The mean peak force exceeded 4.5 N at 1 and 2 Hz, and the corresponding average forces were around 3 N [see Fig. 5(a)]. At 16 Hz, the peak and average forces decreased to 1.7 and 1.0 N, respectively. The mean impulse value was 2.6 N s at 1 Hz, which decreased to 0.08 N s at 16 Hz [see Fig. 5(b)]. One-way ANOVA with repeated measures revealed significant differences between the means of the peak force [ANOVA's F value with degrees of freedom of 4 (the between group variation) and 40 (the within group variation) ($F_{4,40}$)=21.56, $p<0.001$], average force ($F_{4,40}$=18.29, $p<0.001$), and impulse ($F_{4,40}$=23.43, $p<0.001$). Tukey's *post-hoc* tests revealed that for the peak and average forces, the means at 8 and 16 Hz were significantly different from those at lower frequencies ($p<0.001$). For the impulse, the 1-Hz and 2-Hz data were significantly different from the slower tempo data ($p<0.001$). The differences among the means at 4, 8, and 16 Hz were non-significant.

The sign of the string velocity pulses provided information about the changes in bowing direction (see Figs. 2 and 4). The timing of finger force application relative to the bow motion could be assessed by computing the duration between the onset or termination of string clamping force and the

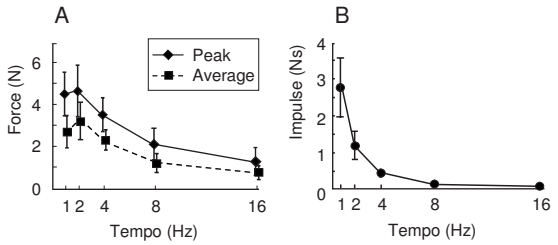


FIG. 5. Mean values of (a) peak string clamping force and average string clamping force, and (b) impulse across all subjects at different tempi. The vertical bars represent ±1 SD. The finger used was the ring finger and the dynamic level *mf*.

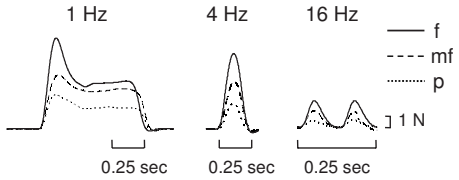


FIG. 6. Representative time-history curves of the string clamping force averaged across 20 trials by one subject at *p*, *mf*, and *f* levels. The finger used was the ring finger.

development (attack) of string velocity for the target tone. Data were obtained from three subjects at *mf* level at 1, 2, and 4 Hz. At 1 Hz, the string velocity for D-tone started on average 16 ms (range 8–31 ms) after the onset of the finger force. The string velocity for the following A-tone started on average 35 ms (range 19–56 ms) after the termination of the string clamping force for the D. For the 2- and 4-Hz cases, only the duration associated with the force termination was available, and their means (37 and 36 ms, respectively) were similar to that of the 1-Hz trials. These findings indicate that for the current three subjects' left hand fingering always preceded the start of the bow strokes.

B. String clamping force at different dynamic levels

Figure 6 shows the mean string clamping force profiles of one subject averaged across the 20 repetitions at *p*, *mf*, and *f* levels for three tempi (1, 4, and 16 Hz), using the ring finger. Note that the overall feature of the force profiles at different dynamic levels was similar even though the peak forces were higher when generating louder sounds. Similar relations between force profiles and dynamic level were observed in all subjects. The loudness effect in relation to tempo was examined using the mean values of the peak force and impulse across all subjects (see Fig. 7). Two-way ANOVA with repeated measures revealed a significant loudness × tempo interaction for the peak ($F_{4,40}$=6.51, $p<0.001$) and impulse ($F_{4,40}$=8.50, $p<0.001$), as well as the main effect of loudness for the peak ($F_{2,20}$=19.00, $p<0.001$) and impulse ($F_{2,20}$=39.75, $p<0.001$).

C. Finger difference in string clamping force

Figure 8(a) shows the mean string clamping force profiles of the index, middle, ring, and little fingers when one subject performed at 2- and 8-Hz tempi at *mf* level. The force profiles were similar for all fingers at each tempo, but the force magnitudes differed. The mean values of the peak

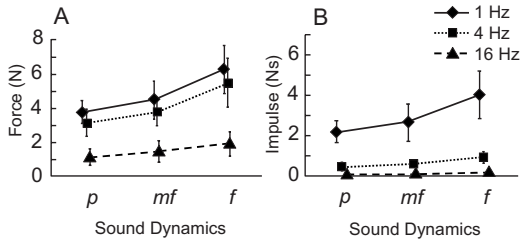


FIG. 7. Mean values of (a) peak force, and (b) impulse across all subjects at different dynamic levels and tempi. The vertical bars represent ±1 SD. The finger used was the ring finger.

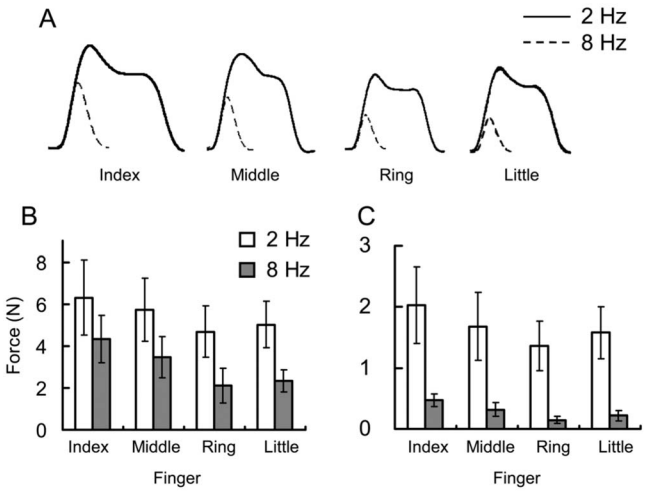


FIG. 8. (a) Illustration of the effect of different fingers on string clamping force at tempi of 2 and 8 Hz at *mf*. (a) String clamping force curves averaged across 20 repetitions by one subject. (b) Mean values of peak string clamping force, and (c) impulse for different fingers across all subjects.

force and impulse across all subjects were largest for the index finger, followed by the middle, little, and ring fingers in decreasing order [see Fig. 8(b)]. One-way ANOVA with repeated measures followed by Tukey's *post-hoc* analysis revealed that the index finger showed a significantly larger peak force ($p<0.05$) and impulse ($p<0.05$) than the ring and little fingers. The peak force for the middle finger was also significantly larger than that for the ring finger ($p<0.05$).

D. String clamping force when playing a musical piece

String clamping forces resulting from the repetitive A-D tone production and the musical pieces were compared. Table I shows the mean values of the peak and average forces, force-application time, and impulse across all subjects when playing the etude at 1-Hz tempo and the nursery rhyme at 2 Hz. Corresponding data from the A-D tone task are also given in Table I. Note that the peak and average forces, as well as the impulse, were higher when playing the melodies than in the repetitive A-D tone production. ANOVA performed on each variable at each tempo, however, revealed that only the impulse ($F_{1,10}=18.47$, $p<0.001$) and force-application time ($F_{1,10}=18.47$, $p<0.001$) at 2 Hz were significantly different. The difference in the mean values of the other variables did not reach the level of significance possibly due to relatively large inter-subject variability (see the SD values of peak and average forces in Table I).

TABLE I. Mean values and SDs (in parentheses) of peak string clamping force, average string clamping force, force-application time, and impulse across all subjects when playing melodies and repetitive A-D tones at the same tempi.

Tempo (Hz)	1		2	
Task	Etude	A-D tone	Nursery rhyme	A-D tone
Peak force	5.3 (2.6)	4.5 (2.3)	5.4 (2.6)	4.7 (2.6)
Average force	3.4 (1.9)	2.7 (1.7)	3.7 (2.0)	3.1 (1.9)
Force-application time	1024 (71)	979 (19)	527 (30)	441 (26)[a]
Impulse	3.4(1.8)	2.6 (1.8)	2.0 (1.1)	1.3 (0.9)[a]

The units for force, impulse, and time are N, N s, and s, respectively. For each variable, one-way ANOVA with repeated-measures was performed to test the difference between the melody and A-D tone tasks at each tempo.
[a]$p<0.01$.

E. Minimum string clamping force

The mean values of the minimum string clamping force for production of the D-tone at *p*, *mf*, and *f* across all subjects were 0.6 ± 0.2, 0.9 ± 0.2, and 1.1 ± 0.2 N. ANOVA revealed significant differences among these means ($F_{2,20}=36.94$, $p<0.001$) Tukey's *post-hoc* test revealed that the differences among all mean values were significant ($p<0.001$).

F. Maximum finger force for individual fingers

The mean values of the maximum finger force possible to produce by the index, middle, ring, and little fingers across all subjects were 30.1 ± 0.5, 28.9 ± 0.4, 22.5 ± 0.4, and 19.1 ± 0.3 N, respectively. ANOVA revealed that the index and middle fingers had significantly higher than the ring and little fingers ($p<0.001$).

G. Subjective responses regarding fingering action

All subjects reported that although they initially experienced some differences in playing the experimental violin compared to their own, they became accustomed to it after an adequate period of practice. They commonly stated, therefore, that the forces generated by the fingers with the experimental violin represented a performance similar to that with their own violin. Concerning the effect of tempo, five subjects reported that their finger press would be lighter with increasing tempo. The others stated that they did not perceive any tempo-related difference in finger pressing. Related to dynamics, five subjects stated that they used a slightly higher finger force to clamp the string when generating a louder sound. Four subjects stated that they experienced the same level of finger force independent of dynamic level. The remaining reported that they were not conscious of any loudness-related difference in the force.

Concerning differences between fingers in the magnitude of the force used to clamp the string, all subjects reported that they did not perceive any differences in the simple tasks in the experiment. Seven added that a slightly greater muscular effort was needed for applying the same force with the little finger as with the other fingers because of its weakness. All subjects also reported that at fast tempi, the little finger was most difficult to use. The index finger was easiest to use followed by the middle finger.

IV. DISCUSSION

A. Experimental violin with force transducer

To the best of our knowledge, our study is the first attempt to install a force transducer on a violin for objective assessment of left hand fingering force during performance. The present system allowed measurement of three orthogo-

nal force components, from which the net resultant force could be computed. This is an important feature for accurate evaluation of the magnitude of the finger force applied, because there were always measurable lateral forces (x- and y-directions) due to the tilt of the finger relative to the fingerboard.

When considering the total force generated by the left fingers in violin playing, we must also take the effects of bow force into account (see Introduction). As described in Sec. II, the string reaction force at the D position on the A string was 2.2 N without bow force, and with bow force applied about 2.1 N at *p* and 1.9 N at *f*. These values, however, will vary depending on the type of string used, the setting of the string-fingerboard distance (string height), and bow-bridge distance.

The string reaction force without bow force was also assessed for the 11 violins brought by the subjects. The force ranged from 1.9 to 2.4 N, with a mean of 2.1 N. For some of the subjects, the string reaction forces of their own instruments were thus slightly different from that of the experimental violin. These subjects indeed reported a perceived difference in string height. However, they also stated that the difference was too small to influence their fingering when playing the experimental violin after an adequate practicing period. Therefore, it seems reasonable to assume that force profiles similar to those observed in the experiments would have been obtained also if they had used their own instruments.

B. Effects of tempo

The profiles of the string clamping force were clearly dependent on tempo. At 1 and 2 Hz, the profiles were characterized by an initial pulse, followed by a level force lasting to the end of the D-tone. At tempi of 4 Hz and above, only the initial force pulse remained. This effect was consistent across subjects, also when playing at different dynamic levels and with different fingers. The initial force pulse seems to be a common feature during tone production in violin playing, which resembles the force pattern observed in fast finger tapping (Aoki *et al.*, 2003). The level force was observed only in sustained notes with durations longer than 250 ms (4 Hz). The initial rapid increase in the string clamping force also for these longer notes can therefore be explained by the fact that players strive to make the duration of the attack as short as possible at all tempi.

The peak force at tempi slower than 4 Hz exceeded 4.5 N at *mf* level, and the average force was around 3 N. These values should be set in relation to the minimum force required to produce *mf* tones, which was 0.9 N on average. At slower tempi, therefore, the players clamp the string against the fingerboard with a force, which is five times higher than actually required. The safety margin in string clamping force in this case was 3.6 N. This is much higher than the estimations by Askenfelt and Jansson (1992) and Galamian (1962), who hypothesized that violinists would use little excess force to clamp the string once the string has begun vibrating. Their descriptions match the present findings at the faster tempi of 8 and 16 Hz much better; there the peak force was clearly lower. However, at these fast tempi, the magnitude of the clamping force probably reflects the temporal constraint of maintaining the target tempo to a much greater extent than the players' intention to use a fair amount of force. It therefore seems most likely that the peak and average forces observed at slower tempi represent the violinists' intended forces for attacking and clamping the string against the fingerboard.

The magnitudes of the string clamping force in this study, in general, might reflect the selection of only young violinists as subjects. A similar study using mature professional orchestra players with much longer playing experience (up to 40 years) should be conducted to obtain more information on common magnitudes of string clamping force.

Somewhat related to the effect of tempo, playing a nursery rhyme led to a longer duration of force application, resulting in larger impulse compared to the repetitive A-D tone production. The longer force-application time suggests a locally reduced tempo, allowing a longer duration of the D-tone. We hypothesize that this is caused by the player's intention to emphasize the tone as part of the melody line. The effect of musical expression on kinematics and kinetics of the left hand fingers thus merits further investigation.

C. Loudness control and string clamping force

Loudness control in bowed string instruments is accomplished by varying three bowing parameters. The bow velocity and the bow-bridge distance control the amplitude of the string motion. The bow force controls the high-frequency spectral content, which contributes substantially to the perceived loudness (Askenfelt, 1986; Schoonderwaldt 2009). All these parameters are associated with the motion of the right hand, not the left. Interestingly, we found that the string clamping force co-varied with loudness, suggesting that the left hand in some way is influenced by the loudness control by the right hand. At slower tempi, the increase in the peak force from *mf* to *f* was much larger than that from *p* to *mf*, indicating that the relationship between loudness and string clamping force is complex.

Why do players use such an excessive force for clamping the string? One reason could be to quickly and securely clamp the string against the fingerboard at loud dynamic levels; there high bow accelerations and high bow forces are used in order to obtain clean and short attacks (Guettler and Askenfelt, 1997). The observed higher force pulse peaks at the onset of the string clamping force at loud levels support this notion. Another somewhat related reason may be that players unconsciously learn to control the dynamic level through activation of the whole body, including more active string clamping by the left fingers. Indeed, the subjects commonly reported that stronger clamping of the string when playing loudly was a spontaneous behavior along with vigorous bowing by the right arm, and that changing this relationship would be uncomfortable.

The string clamping force used for clamping the string can also be discussed in relation to the maximum force possible for the finger used. At the slowest tempo examined, the peak string clamping force ranged from 3.8 N at *p* to 6.3 N at

f level when using the ring finger (see Fig. 8). When string reaction forces at these dynamics are added, these peak values can be increased to 5.9 and 8.2 N, respectively, which correspond to 26% and 36% of the maximum finger force possible (22.5 N) by the ring finger. Comparing the average forces (3.2 N at *p* and 5.5 N at *f*) for the same finger, the corresponding values are 5.3 N (24% of the maximum force) and 7.4 N (33% of the maximum force), respectively. According to our results, the range in string clamping force commonly used by violinists in regular performance would be one-third or less than their maximum finger force possible.

D. Finger differences in string clamping force

The post-experimental interview revealed that the subjects experienced that the string clamping force they used was similar for all fingers. This may suggest that their mental effort for clamping the string against the fingerboard was the same. Contrary to the self-reported data, the transducer data showed that the string clamping force of the index finger was significantly higher than that of the other fingers (see Fig. 8). The middle finger also produced a higher string clamping force than the ring and little fingers. The orderly relationship in the string clamping force across the fingers thus resembled that of the maximum finger force possible. The subjects also reported the same order in their subjective judgment of the ease of clamping.

We earlier reported a similar ordering effect of individual fingers in the production of the maximum voluntary force as well as ease of tapping movements for nonmusicians (Kinoshita *et al.*, 1996; Aoki *et al.*, 2003). We therefore hypothesize that even for trained violinists, innate anatomical and musculo-physiological factors determining the motor function of individual fingers play an important role in string clamping in string playing. It is also possible that the overwriting of these innate factors is limited even with extensive violin training. However, these hypotheses may need to be re-examined in studies including players with a much longer professional career (e.g., 30–40 years). Difficulty in the effective use of the ring and little fingers has also been reported by highly trained young pianists (Aoki *et al.*, 2005).

ACKNOWLEDGMENTS

We thank Mr. H. Nakaya of the String Instrument Design Section, Mr. H. Nakayama and Mr. M. Hara of the Music Research Institute of Yamaha Music Foundation, and Mr. T. Iwai of Iwai's Violin Studio Cremona, Osaka, Japan, for their excellent technical support. We also thank Dr. A. Askenfelt of KTH, Sweden for his generous assistance in editing the manuscript, and Dr. E. Schoonderwaldt of KTH for kindly providing us with valuable information on the bow-string interaction. Appreciation is extended to anonymous reviewers for their valuable and thoughtful comments to improve the manuscript. This study was supported by a research grant from the Yamaha Music Foundation.

[1]String design: The E string was tin-plated carbon steel, the A and D strings had a hydronalium-wound composite core, and the G string was silver-wound on composite core (see Thomastic-Infeld Co.: http://www.thomastik-infeld.com/strings/index.html).

[2]To assess the string reaction force for the A string at the position of D5 on the experimental violin and the violins brought by the subjects, the string was slowly pressed using a hand-held uniaxial strain gauge force transducer (maximum force range=20 N) until it touched the fingerboard. The force at the moment of fingerboard contact was recorded via an A/D converter interfaced to a PC. The vertical distance between the under side of the A string (diameter=0.7 mm) and the fingerboard at the position of D5 was 2.4 mm on the experimental violin. The corresponding value on the subjects' violins ranged from 2.3 to 2.8 mm, with a mean value of 2.5 ± 0.2 (SD) mm.

[3]The bow force would act together with the finger force to press down the string during playing. A typical magnitude of the normal force applied to the string by the bow during violin playing range from 0.5 N at *p* to 2.0 N at *f* level (Askenfelt, 1989). We therefore applied a 0.5- and 2-N weight to the A string at a position of 30 mm from the bridge, and computed the associated vertical displacement of the string at the D5 position relative to the no-weight condition from a digital camera recording with calibrated scale. The string was displaced by 0.07 and 0.29 mm for the 0.5- and 2-N weights, respectively, corresponding to 3% and 12% of the original string-fingerboard distance (2.4 mm). A proportional reduction in the string reaction force of 2.2 N may be expected with these bow forces, amounting to 0.07 N at *p* and 0.26 N at *f*.

Aoki, T., Francis, P. R., and Kinoshita, H. (**2003**). "Differences in the abilities of individual fingers during the performance of fast repetitive tapping movements," Exp. Brain Res. **152**, 270–280.

Aoki, T., Furuya, S., and Kinoshita, H. (**2005**). "Finger-tapping ability in male and female pianists and nonmusician controls," Motor Control **9**, 23–39.

Askenfelt, A. (**1986**). "Measurement of bow motion and bow force in violin playing," J. Acoust. Soc. Am. **80**, 1007–1015.

Askenfelt, A. (**1989**). "Measurement of the bowing parameters in violin playing. II: Bow-bridge distance, dynamic range, and limits of bow force," J. Acoust. Soc. Am. **86**, 503–516.

Askenfelt, A., and Jansson, E. V. (**1992**). "On vibration sensation and finger touch in stringed instrument playing," Music Percept. **9**, 311–350.

Baader, A. P., Kazennikov, O., and Wiesendanger, M. (**2005**). "Coordination of bowing and fingering in violin playing," Brain Res. Cognit. Brain Res. **23**, 436–443.

Brandfonbrener, A. G. (**1990**). "The epidemiology and prevention of hand and wrist injuries in performing artists," Hand Clin. **6**, 365–377.

Flesch, C. (**1930**). *The Art of Playing the Violin: Artistic Realization and Instruction. VII*, translated by F. H. Martens (Carl Fischer Inc., New York).

Galamian, I. (**1962**). *Principles of Violin Playing and Teaching* (Prentice-Hall, Englewood Cliffs, NJ).

Guettler, K., and Askenfelt, A. (**1997**). "Acceptance limits for the duration of pre-Helmholtz transients in bowed string attacks," J. Acoust. Soc. Am. **101**, 2903–2913.

Kinoshita, H., Murase, T., and Bandou, T. (**1996**). "Grip posture and forces during holding cylindrical objects with circular grips," Ergonomics **39**, 1163–1176.

Lederman, R. J. (**2006**). "Focal peripheral neuropathies in instrumental musicians," Phys. Med. Rehabil. Clin. N. Am. **17**, 761–779.

Schoonderwaldt, E. (**2009**). "The violinist's sound palette: Spectral centroid, pitch flattening and anomalous low frequencies," Acta Acust. Acust. (in press).

Fundamental frequency influences the relationship between sound pressure level and spectral balance in female classically trained singers

Sally Collyer[a] and Pamela J. Davis[b]
National Voice Centre, The University of Sydney, New South Wales 2006, Australia

C. William Thorpe[c]
School of Communication Sciences and Disorders (C42), The University of Sydney, New South Wales 2006, Australia

Jean Callaghan
School of Contemporary Arts, The University of Western Sydney, Penrith, New South Wales 2750, Australia

(Received 31 July 2008; revised 27 March 2009; accepted 19 April 2009)

The influence of fundamental frequency (F0) on the relationship between sound pressure level (SPL) and spectral balance (SB) has been largely unexplored in the female singing voice. Five classically trained females performed a *messa di voce* across their musical F0 range. Average maximum SB rose with F0 by 0.27 dB/semitone (ST) to B4 and then decreased, while average minimum SB fell by 0.5 dB/ST to E5 and then generally rose. Of 318 tokens, 208 showed a linear SPL:SB relationship ($R^2 \geq 0.5$), but F0 affected SPL:SB slope and intercept and their interaction above and below B4. The possibility that this reflects a change from subglottal inertance to compliance is discussed. Consistency of SB behavior change at B4 and E5 contrasted with variability in first-formant frequency. Nonlinear SPL:SB relationships did not arise from SB saturation. The presence of low-SPL "tails" may reflect the challenge in modifying vocal-fold adduction during crescendo and decrescendo. The results show that analysis of the SPL:SB relationship must take F0 into consideration.
 [DOI: 10.1121/1.3132526]

PACS number(s): 43.75.Rs [DAB] Pages: 396–406

I. INTRODUCTION

Rising sound pressure level (SPL) in the human voice is known to alter the spectral balance (SB), the ratio of higher-partial to lower-partial energy. Rising SPL leads to a greater increase in higher- than in lower-partial energy (Sundberg *et al.*, 1993; Sjölander and Sundberg, 2004), which in this paper will be called an increase in SB. Studies of the relationship between SPL and SB change have, however, shown certain complexities in speech (Alku *et al.*, 2006; Ternström *et al.*, 2006) and in male classically trained singers (Sjölander and Sundberg, 2004). Female classically trained singers have shown complex and fundamental-frequency (F0) dependent behavior (Collyer *et al.*, 2007), but the SPL:SB relationship in female singing is largely unexplored. This study investigates the effect of F0 on SB behavior and on the relationship between SPL and SB change in female classically trained singers.

Before reviewing literature, it should be noted that various measures have been used to quantify relative change in high- and low-frequency bands. First, measures have been based on spectral peak within a band [e.g., the singing power ratio used by Omori *et al.* (1996)] or on energy within a band [e.g., the energy ratio used by Thorpe *et al.* (2001)]. Second, different bands have been employed, such as 0–2:2–4 kHz (Collyer *et al.*, 2007; Thorpe *et al.*, 2001), 0.1–1:2–6 kHz (Ternström *et al.*, 2006), 0–1:1–5 kHz (Frøkjær-Jensen and Prytz, 1976), 0–1:1–6 kHz (Sundberg and Nordenberg, 2006), and the level of the first formant to the level of the singer's formant (Sjölander and Sundberg, 2004). A major difference between these band choices is the treatment of the second formant. Lastly, some studies have compared long-term average spectra across 30 or more seconds with equivalent sound level (e.g., Sundberg and Nordenberg, 2006), whereas others have compared instantaneous SB with instantaneous SPL (e.g., Collyer *et al.*, 2007; Ternström *et al.*, 2006). Appropriate choice of measurement, bands, and duration depends on the question being addressed. Comparison of study results needs to take such choices into consideration.

The increase in SB with rising SPL has been attributed to vocal tract inertance effects on the glottal-flow waveform (Titze, 1994). In soft phonation, the source flow waveform is generally symmetrical, with approximately equal opening and closing rates, as illustrated by Alku *et al.* (1998) (their Fig. 3). As subglottal pressure rises, the amplitude of the flow waveform increases, but vocal tract inertance causes the waveform to skew to the right as the closing rate increases. This raises the maximum flow declination rate (MFDR), which normally represents primary excitation of the vocal

[a]Author to whom correspondence should be addressed. Electronic mail: sallycollyer@yahoo.com.au
[b]Present address: School of Communication Sciences, La Trobe University, Australia.
[c]Present address: Cantovation Technology, Auckland, New Zealand.

tract. The increase in SB with rising SPL is generally linear (Gauffin and Sundberg, 1989; Sjölander and Sundberg, 2004; Sundberg, 2001; Ternström, 1993), but F0 and SPL are known to affect the relationship.

Studies have observed a moderating effect of F0 on the relationship for singers between SPL and the high-pass (2–4 kHz) signal (Sundberg, 2001) and between SPL and the level of the singer's formant (Bloothooft and Plomp, 1986). Sjölander and Sundberg (2004) calculated SB as the difference in level of the first formant and of the singer's formant, in classical baritones. They confirmed that the relationship between SPL and the natural log of subglottal pressure was substantially linear at a given F0, as was the relationship between the natural log of subglottal pressure and SB, supporting a substantially linear relationship between SPL and SB at a given F0. However, they found that rising F0 was associated with a statistically significant increase in slope, and decrease in intercept, of the linear regression between SPL and the natural log of subglottal pressure, indicating an influence of F0 on the relationship between SPL and SB.

Linearity in the SPL:SB relationship breaks down at SPL extremes. Sundberg and Nordenberg (2006) (p. 456) observed that the relationship between 0–1:1–6kHz and equivalent sound level (in a long-term average spectrum) is linear "if softest possible phonation is excluded." Sundberg *et al.* (1993) observed that, at low values of SPL (and concomitantly of subglottal pressure and MFDR), increasing subglottal pressure raised SPL by shortening the closing phase until the closed quotient (Qc; calculated as the duration of the closed phase/total cycle duration) saturated at 0.4. Thence, increasing subglottal pressure raised peak glottal flow, in turn raising MFDR. Alku *et al.* (1999) found a greater increase in SB at lower than higher SPL.

Sundberg *et al.* (1993) noted that highest subglottal pressure (and thus SPL) on a given F0 was associated with more symmetrical source waveform, i.e., less skew and relatively lower MFDR, in baritones. Holmberg *et al.* (1988) observed this also in female speakers. The apparent decrease in MFDR at high SPL has been attributed to biomechanical limits to increasing glottal closing speed (Pabon, 1991). Alku *et al.* (2006) examined changes in the glottal-flow waveform associated with rising SPL in speech. They found that, at high effort levels (>85 dB SPL at 40 cm), the amplitude of the source waveform continued to increase but the duration of the glottal closing phase asymptoted to a minimum >0, so that at high effort level the source waveform skewed no further but only increased scale of amplitude, as observed also by Alku *et al.* (1998). Ternström *et al.* (2006) reasoned that flow-change and biomechanical limits in the human voice should mean that, beyond some SPL, the SB of the source spectrum should be unable to rise further. Using individual phonemes obtained from running speech, they found that SB did indeed saturate, at around 6–8 dB less than the individual's maximum SPL, so that SB remained unchanged or even fell at the speakers' highest SPL.

However, Alku *et al.* (2006) and Ternström *et al.* (2006) did not control F0, so the effect of F0 change on the SPL:SB relationship as noted previously remains unclear. The principal regulator of SPL is subglottal pressure (Rubin *et al.*, 1967; Sundberg *et al.*, 1993), which in turn tends to raise F0, to varying degrees depending on the location of the F0 within the overall F0 range (Titze, 1989; Vilkman *et al.*, 2002). Alku *et al.* (2002) reported that raising F0 as a strategy for increasing SPL gained an average 4 dB in loud speech. The female speakers studied by Ternström *et al.* (2006) raised average F0 with SPL by 7.8 STs, from 228 Hz (SD 23 Hz) to 357 Hz (SD 52 Hz) in the loudest condition (as reported in their companion paper Södersten *et al.*, 2005). It is possible that the significant amount of scatter, which Ternström *et al.* (2006) found in their SPL:SB plots might be due in part to F0 influence.

Influence of F0 has a particular significance for female classically trained singers because much of their F0 range is affected by first-formant (F1) interaction. Pabon (1991) noted that, as F0 rises, airflow becomes more important in controlling SPL, and it becomes harder to sustain asymmetry of the glottal waveform as F0 approaches F1. Sundberg *et al.* (1993) observed that, as F0 fell below one-third of F1, pulse symmetry (calculated as closing phase/opening phase of the flow glottogram) asymptoted to 0.6. As F0 rose above one-third of F1, pulse symmetry approached unity, i.e., equal durations of closing and opening phases. This would imply that, at high F0, SB might remain substantially unaltered as SPL changed. An example of this was illustrated in Collyer *et al.*, 2007 ("F5" in their Fig. 3), but it was also noted that the singers differed markedly in the point (F0) at which the relationship between SPL and SB became nonlinear. For one singer, linearity ceased only at B5 (988 Hz), well above the point at which the singer would have been expected to begin raising F1 for an /ɑ/ vowel (Sundberg, 1987; Joliveau *et al.*, 2006). These effects on the SPL:SB relationship in female singers are unclear.

Mode of phonation also affects the SPL:SB relationship. Sundberg *et al.* (1993) reported that the relationship between SPL and subglottal pressure altered with the degree of glottal adduction. They observed that a given subglottal pressure resulted in decreasing peak glottal flow as phonation moved from breathy to pressed, i.e., as adductive force increased. Bloothooft and Plomp (1986) observed that the relationship between SPL and the level of the singer's formant altered with changes in vocal-fold adduction, as well as with F0, as noted previously. Titze (1988) noted that a major challenge in the *messa di voce* is to maintain a constant abduction quotient (defined as prephonatory glottal halfwidth at the vocal processes/vibration amplitude at the center of the glottis). Glottal adduction must vary inversely with amplitude of vibration during the *messa di voce* in order to avoid pressed voice at high SPL and breathy voice at low SPL, conditions which would imply a nonlinear SPL:SB relationship.

This study investigated changes in SB behavior across the musical F0 range of female classically trained singers. The study also investigated whether F0 influenced the relationship between SPL change and SB change. The first of three questions addressed the overall behavior across F0 range and asked the following: How does F0 alter SB behavior? The second question considered SB behavior within the *messa di voce* and asked the following: Do female singers exhibit a linear SPL:SB relationship affected by F0? The

third question asked the following: Is nonlinearity due to greater variance in the SPL:SB relationship at SPL extremes?

II. METHOD

Details of participants and the method for data-gathering and for preparation of SPL and SB vectors are contained in a companion paper (Collyer *et al.*, 2007) and are summarized here. The study was approved by the Human Ethics Committee of The University of Sydney. Participants were five female classically trained singers (three lyric sopranos, one lyric coloratura soprano, and one lyric mezzo), ranging in age from 24 to 30 years. Audio recordings were made in a sound-treated room 3.7 m wide × 3.5 m long × 2.7 m high, using a cardioid microphone (Pearl Mikrofon-laboratorium, model TL6C) at a distance of 30 cm from the singer's mouth. A calibration signal was recorded with a 1 kHz tone from a signal generator (Power Acoustik CP 500C) and a Rion NL 06 SPL meter set to slow damping with no weighting.

The study used the *messa di voce*, a crescendo and decrescendo on a single F0 with a single vowel. This standardized each token for F0 and for vowel. Two *messa di voce* tokens with a target duration of 8 s were obtained on /ɑ/ on each F0 throughout the singer's musically acceptable phonational range, beginning at A4 and descending by ST, then repeating A4 and rising by ST. A SPL (rms amplitude) vector for each token was obtained (MATLAB 5.3, MathWorks) at a sampling frequency of 300/s, and the SPL vectors were calibrated. For SB analysis, the sound signal of each token was filtered into low band-pass (0–2 kHz) and high band-pass (2–4 kHz) signals (COOL EDIT 96), from which were derived low-band and high-band SPL vectors, respectively. The low-band vector was subtracted from the high-band vector to obtain a SB vector over the course of each token. Plotting and analysis of the SPL and SB vectors are described with the results.

Linear mixed-effects models used the NLME package (Pinheiro *et al.*, 2008) in R (v2.8.1, R Development Core Team, 2008). First-formant (F1) frequency was estimated by linear predictive coding using PRAAT (Boersma and Weenink, 2007). F1 was averaged across six tokens per singer, on A#3, B3, and C4. In soft phonation, amplitude of the first harmonic generally exceeds the amplitude of the harmonic closest to F1. To ensure this did not distort F1 prediction, estimation was made over the section 2.5–4.5 s.

III. RESULTS

A total of 318 *messa di voce* tokens were recorded from E3 (165 Hz) to E6 (1319 Hz), with F0 range per singer averaging 30.8 semitones (STs) [standard deviation (SD) 3.8 STs].

A. Individual SPL:SB plots

Each *messa di voce* token was plotted with SPL on the abscissa and SB (the difference in dB of 2–4 kHz less 0–2 kHz) on the ordinate. All SPL:SB plots were standardized to a SPL range of 40–120 dB at 30 cm and a SB range of −60 to 20 dB. Linear regression was applied to each token, and

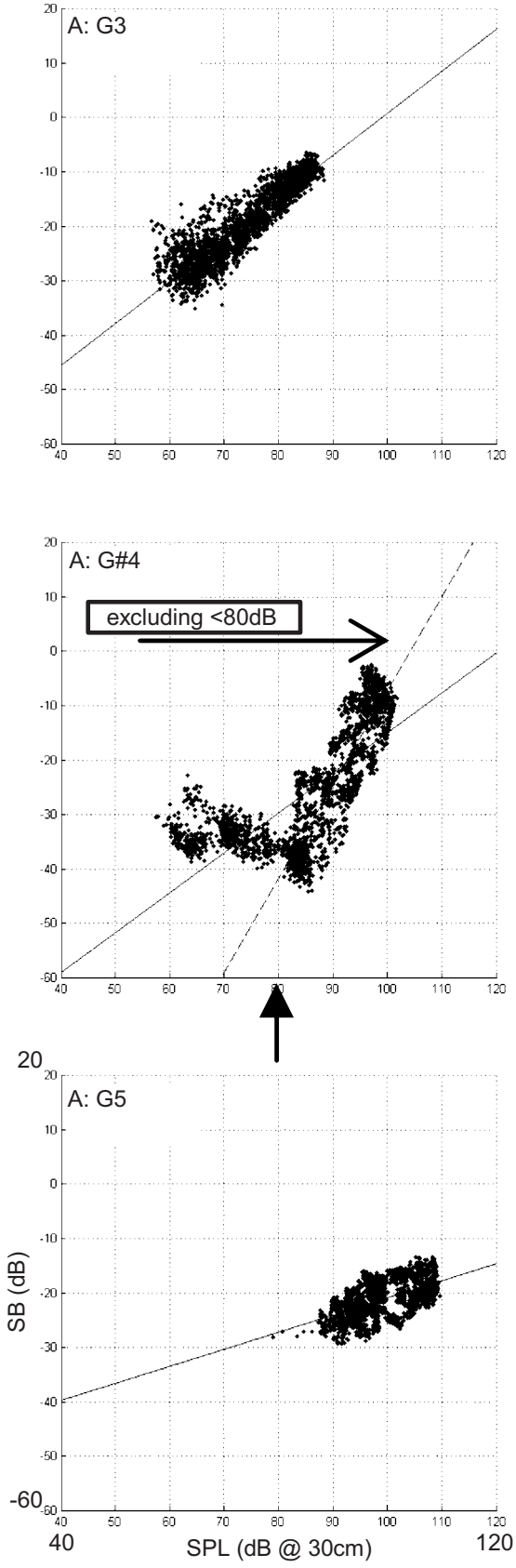


FIG. 1. Examples of three SPL:SB plots (G3, G#4, and G5 by singer A), with sound pressure level (SPL dB at 30 cm) on the abscissa and SB on the ordinate, with linear regression. SB was calculated as high band-pass (2–4 kHz) less low band-pass (0–2 kHz). All SPL:SB plots were standardized to a SPL range of 40–120 dB at 30 cm and a SB range of −60 to 20 dB. The dotted line on G#4(1) represents the regression line for points ≥80 dB SPL, as discussed under Sec. III.

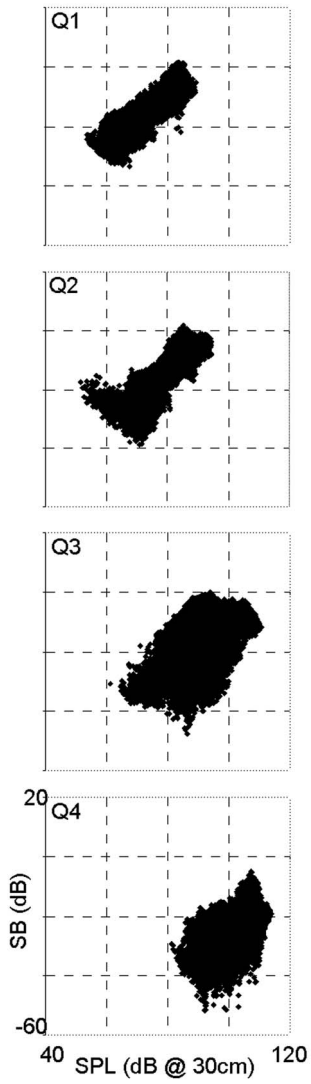


FIG. 2. SPL:SB data points for singer E, demonstrating the change in the SPL:SB relationship with rising F0. Singer E's results were representative of the results for all singers. Each plot contains data points from one quartile of singer E's range. (Q1 = 0–25% F0 range; Q2 = 26–50%; Q3 = 51–75%; Q4 = 76–100%).

the slope, intercept, and coefficient of determination (R^2) were recorded. Figure 1 sets out examples of SPL:SB plots obtained, including the regression line that will be addressed later. The plots suggested a change in SPL:SB relationship from linear and tight at low F0s to highly scattered at high F0s.

B. Combined SPL:SB plots

Next, four *x*-*y* plots of the SPL:SB data were drawn for each singer, one plot per quartile of the singer's F0 range. These plots were similar to those in Ternström *et al.*, 2006 but illustrated the effects of F0 change. The four quartile plots for singer E are presented in Fig. 2 and are representative of the results for all singers. The plots show that SPL rose but SB decreased with rising F0, so that the groups of data points shift horizontally rightwards and downwards with rising quartile. Thus, as F0 rose, a given SB value was associated with higher SPL values.

The effect of F0 quartile grouping was analyzed as a linear mixed-effects model, with SPL (covariate) and F0 quartile (block) as fixed effects and a random intercept for singer. F0 quartile was a statistically significant effect after adjusting for SPL ($F_{(3,251\,665)}=41\,682.58$, $p<0.0001$). Post-hoc contrasts of adjacent F0 quartile groups found that SB decreased significantly with rising F0 after adjusting for SPL (all $p<0.0001$). Mean SB differences between F0 quartile groups, with 95% confidence intervals and t-tests, are set out in Table I. Mean SB fell by 19.0 dB from the first to the fourth quartile.

C. SPL and SB across F0

Figure 2 identified that there were changes in the SPL:SB relationship with rising F0. Figure 3 details the trajectory of the minimum and maximum curves for (a) SPL and for (b) SB across F0 range for each singer. To draw Fig. 3, the SPL and SB vectors for each *messa di voce* token (four tokens on A4 and two on the remaining F0s) were separately sorted into ascending order, and the 97th and 3rd percentiles for each token were plotted as maxima and minima, respectively. [The slope measurements defined by the filled diamonds in Fig. 3(b) are discussed later.] Both SPL curves showed a strong linear trend between rising F0 and increasing SPL, averaging $R^2=0.92$ (SD 0.04) for maximum SPL and $R^2=0.89$ (SD 0.04) for minimum SPL. However, straight lines were a poor fit to the SB curves, averaging $R^2=0.29$ (SD 0.14) for maximum SB and $R^2=0.26$ (SD 0.24) for minimum SB. Optimal fit (defined as the lowest-order polynomial regression for which the next higher-order improved average R^2 by <0.03) was a fourth-order polynomial regression for maximum SB (average $R^2=0.75$, SD 0.08) and a third-order polynomial regression for minimum SB (average $R^2=0.67$, SD 0.13).

In other words, whereas the SPL curves generally in-

TABLE I. Post-hoc contrasts of adjacent F0 quartile groups (as illustrated in Fig. 2) found that SB decreased significantly with rising F0 after adjusting for SPL (all $p<0.0001$). Mean SB fell by a total of 19.0 dB from the first to the fourth quartile.

Quartile contrast	Mean difference (dB)	95% confidence interval		*t*-statistic
		Lower	Upper	
Q1 vs Q2	−3.495	−3.573	−3.417	$t_{(251\,665)}=-87.999$, $p<0.001$
Q2 vs Q3	−8.813	−8.893	−8.732	$t_{(251\,665)}=-213.816$, $p<0.001$
Q3 vs Q4	−6.733	−6.812	−6.654	$t_{(251\,665)}=-167.134$, $p<0.001$

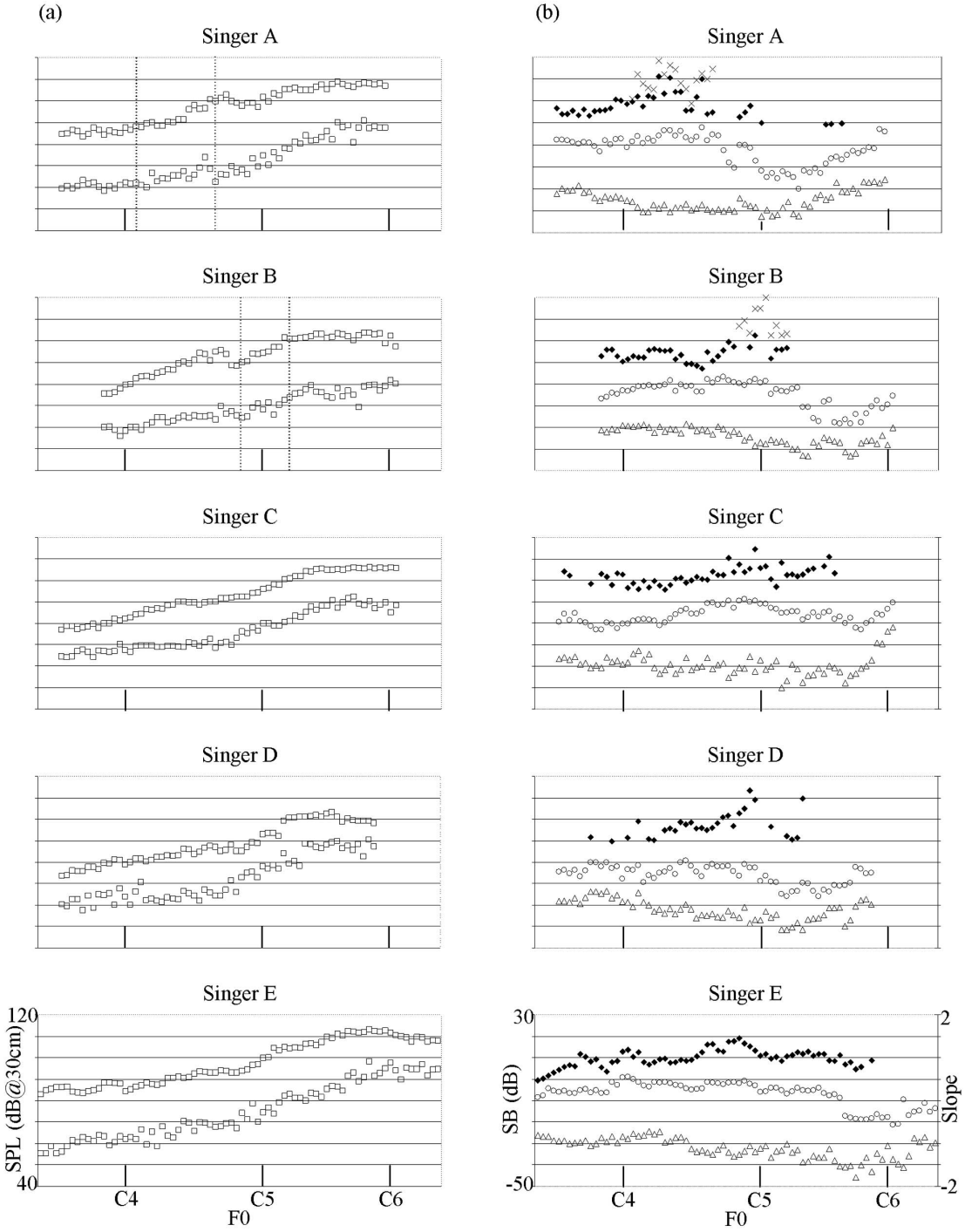


FIG. 3. SPL and SB in the *messa di voce* across fundamental frequency (F0) range (two tokens per F0). (a) Maximum and minimum SPLs at 97th and 3rd percentiles, respectively. The F0 range of tokens with low-SPL tails (as illustrated in Fig. 1 "A: G#4") is indicated by dashed vertical lines for singers A and B. (b) Maximum SB at 97th percentile (open circles); minimum SB at 3rd percentile (open triangles) and SPL:SB slope where $R^2 \geq 0.5$ (filled diamonds). For singers A and B, SB was also calculated after removing low-SPL tails. Resulting slopes are marked with a cross.

creased with rising F0, the SB curves varied trajectory. Maximum SB increased as F0 rose to B4, decreased from C5 to a point within the range F#5-A5, then increased again for the singer's highest F0s. Minimum SB decreased to E5, then increased. These patterns are clear in Fig. 4, which pools tokens from all singers on F0s common to all (A#3–A#5). Figure 4 plots the maxima and minima of SPL and of SB, averaged across the 10 tokens at each F0 in this range (20 tokens on A4). Figure 4 also plots SPL and SB ranges (calculated as average maximum minus average minimum). Regression lines (linear or polynomial) were optimally fit to each measure, and the results are set out in Table II. Again, optimal fit was defined as the lowest-order regression for which the next higher-order improved R^2 by <0.03.

The increase in average minimum SPL was 1.20 dB/ST ($r=0.975$), slightly steeper than the increase in average maximum SPL of 1.14 dB/ST ($r=0.986$). Optimal fit for the average maximum and average minimum SB curves required

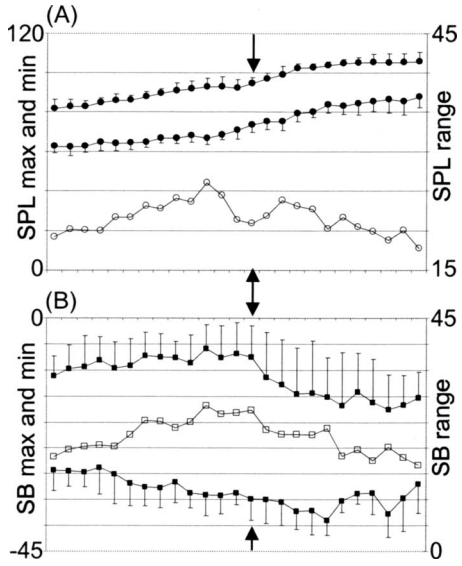


FIG. 4. The five singers performed two *messa di voce* tokens at each F0 (four tokens each on A4) in the range A#3–A#5. The arrows identify B4. SPL and SB measures were averaged across all 10 tokens at each F0 (20 tokens on A4). (a) Solid circles mark average SPL maxima and minima on the left ordinate ranging from 0 to 120 dB (at 30 cm). Open circles mark SPL range, calculated as maximum SPL-minimum SPL, on the right ordinate ranging from 15 to 45 dB. (b) Solid squares mark average SB maxima and minima on the left ordinate ranging from −45 to 0 dB. Open squares mark SB range (maximum SB-minimum SB) on the right ordinate ranging from 0 to 45 dB. T bars mark SD on both charts.

fourth- and third-order polynomial regressions, respectively, which are detailed in Table II. However, as suggested by Fig. 4, average maximum SB and average minimum SB showed strong linear trends up to and including B4 and E5, respectively.

Average maximum SB increased with rising F0 ($r=0.796$) up to B4 inclusive, with a slope of 0.27 dB/ST. It is noted for subsequent discussion that Fig. 4 shows average maximum SPL at B4 was 94.34 dB at 30 cm, which was 11.31 dB less than the peak of average maximum SPL on A#5. It is also noted that Fig. 4 shows a dip in SPL range centering on B4. Above B4, maximum SPL continued to rise but maximum SB generally declined.

Average minimum SB decreased to E5 inclusive ($r=-0.966$) at a rate of −0.50 dB/ST. The nadir around E5 occurred for all five singers, with a second dip at G#5 for four singers. These characteristics of minimum SB were not reflected in any of the average SPL curves (maximum, minimum, or range), except that at E5 the rate of increase in minimum SPL with rising F0 notably reduced. Above G#5, both average maximum SB and average minimum SB rose; Fig. 3 showed that this was the case for all singers up to their individually highest F0.

D. Linear SPL:SB relationships

Examples in Fig. 1 suggested that SPL:SB plots at low F0s were highly linear but became scattered and nonlinear at higher F0s. The extent of linearity across F0 change was tested. As previously mentioned, a straight line was fitted to the SPL:SB plot for each token using linear regression, and the slope, intercept, and R^2 were recorded. For $R^2<0.5$, a straight line was deemed unrepresentative of the SPL:SB trace, and the token was excluded from analysis of slope and intercept. These cases are demonstrated in Fig. 1, where singer A's tokens on G3 and G#4 were linear but her token on G5 was not. Of 318 tokens, $R^2\geq 0.5$ for 208 tokens (65.4%), but individual singers' results varied (singer A =60%, B=57%, C=73%, D=48%, and E=83%). Figure 3(b) (filled diamonds) charts the slope of the regression line for those tokens where $R^2\geq 0.5$.

Figure 5 offers a clearer and more detailed representation of the gradual change in the SPL:SB relationship with rising F0 crudely mapped in Fig. 2 and described by the curves in Fig. 3. Figure 5 plots the regression lines of those tokens by singer E for which $R^2\geq 0.5$. (Again, her results are representative of those of the other singers.) For each slope recorded in Fig. 3(b) (filled diamonds), a line with that slope and intercept was drawn in Fig. 5 between the 3rd and 97th percentiles of SPL and of SB.

The slope of the SPL:SB regression line of linear tokens and its relationship to F0 varied widely between singers, as can be seen in Fig. 3(b). Slope averaged 0.936 (SD 0.263),

TABLE II. Optimally fitting linear or higher-order regression lines of SPL and SB curves in Fig. 4. Correlations (r) for linear regressions are enclosed in square brackets, to indicate polarity. Maximum and minimum SPL and SB were averaged from 10 tokens per fundamental frequency (F0) (20 tokens on A4), and range was calculated as maximum-minimum. The maximum SB and minimum SB curves showed strong linear trends up to and including B4 and E5, respectively, and results for these have been included.

Curve	Type of line	Regression equation	R^2 [r]
SPL maxima	Linear	$y=1.1378x+80.006$	0.972 [0.986]
SPL minima	Linear	$y=1.1952x+57.643$	0.951 [0.975]
SPL range	Second-order	$y=-0.0335x^2+0.8138x+18.442$	0.693
SB maxima	Fourth-order	$y=0.0004x^4-0.018x^3+0.1845x^2-0.1196x-10.679$	0.910
(≤B4)	Linear	$y=0.2668x-10.314$	0.634 [0.796]
SB minima	Third-order	$y=0.0023x^3-0.0643x^2+0.0116x-29.19$	0.808
(≤E5)	Linear	$y=-0.5009x-28.17$	0.934 [−0.966]
SB range	Second-order	$y=-0.0599x^2+1.4465x+16.673$	0.791

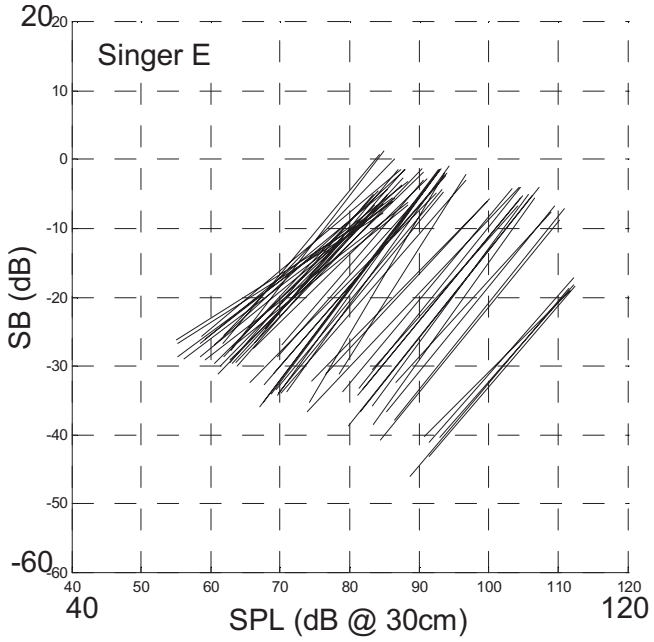


FIG. 5. Linear regression for singer E for each *messa di voce* token where $R^2 \geq 0.5$, illustrating the decrease in minimum SB and the shift rightwards along the SPL axis. Each line ranges from the point of minimum SPL:minimum SB to maximum SPL:maximum SB with slope as presented in Fig. 3.

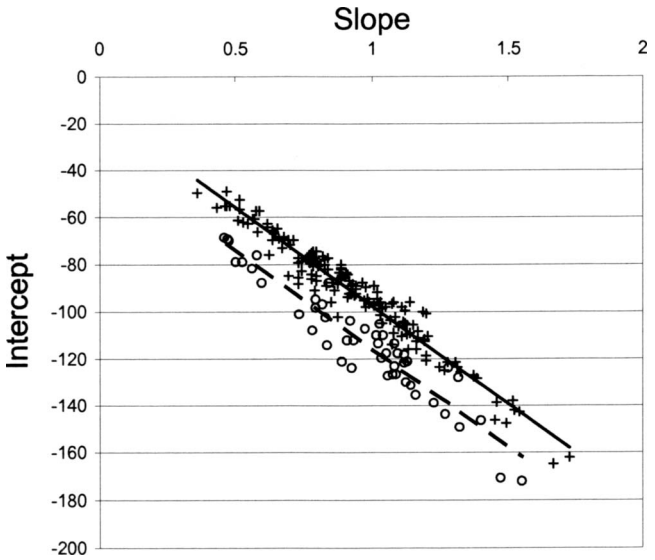


FIG. 6. Slope (abscissa, ranging from 0 to 2) against intercept (ordinate, ranging from 0 to −200) of the regression lines illustrated in Fig. 5. (Symbols: +=tokens≤B4; o=tokens>B4) Intercept was 19.98 dB lower for tokens above B4, after adjusting for slope.

ranging from 0.359 to 1.730. Although slope peaked at B4, correlation with rising F0 was negligible across linear tokens ≤B4 (r=0.39), reflecting that the increase in slope did not begin from the lowest F0 for three of the singers (A–C). The decrease in slope above B4 occurred as the decrease in SB range proportionately outstripped the decrease in SPL range, which in turn was due to decrease in maximum SB exceeding the near-linear decrease in minimum SB to E5, inclusive.

Correlation of the slope and intercept of the regression lines showed a strong inverse relationship (r=−0.916; 208 tokens). However, plotting token slope against intercept for all linear tokens in Fig. 6 indicated that intercept was lower for tokens >B4. A linear mixed-effects model contrasting below/equal to B4 and above B4, with slope as covariate and a random intercept for singer, found a mean decrease in intercept of 19.98 dB (95% CI lower=−21.71 and upper=−18.26) above B4, after adjusting for slope. The difference was statistically significant ($t_{(201)}$=−22.8218, $p<0.0001$). In other words, for two tokens with a matching rate of change in SB with changing SPL (same slope), a given SPL value in the token above B4 was associated with a SB value 19.98 dB lower than in the token at or below B4.

E. Nonlinear SPL:SB relationships

Of the 318 tokens, 110 (34.6%) were defined as having a nonlinear SPL:SB relationship ($R^2<0.5$). We defined SB as the difference between the high-band (2–4 kHz) and low-band (0–2 kHz) SPL vectors. This meant that the slope in the SPL:SB plots was mathematically the slope of the SPL:high-band minus the slope of the SPL:low-band. To investigate whether nonlinear SPL:SB relationships arose because the SPL:high-band relationship was SPL-dependent, we examined the constituent vectors of high-band and low-band against SPL change. Across all 318 tokens, there was near-perfect correlation between low-band and SPL (average R^2=0.998 and SD=0.004), with low-band rising on average 0.98 dB per 1 dB SPL (SD=0.03 dB). The two vectors were substantially identical because key determinants of overall SPL (the first harmonic in soft phonation, and harmonics in the region of the first formant in moderate and loud phonation) fall within the low-band. As expected, the increase in high-band exceeded that of overall SPL, averaging 1.72 dB per 1 dB overall SPL (SD=0.38 and median=1.77 dB). However, there was also a strong correlation between high-band and overall SPL (average R^2=0.871, SD=0.105, and median R^2=0.907).

An example of nonlinearity is given in Fig. 7(a), which plots SB, high-band, and low-band against overall SPL for singer C on B5. Slopes of the regression lines were x=0.112, 0.997, and 1.109, respectively. The SPL:SB relationship in this token was defined as nonlinear (R^2=0.028), i.e., change in SPL explained <3% of the variance in SB. The SPL:low-band relationship approached unity (R^2=0.994). However, the SPL:high-band correlation was also high (r=0.850), so that change in overall SPL explained 72.3% of the variance in high-band SPL. Thus, nonlinearity in the SPL:SB relationship did not arise from nonlinearity in the underlying SPL:high-band relationship. That is to say, as can be seen in Fig. 7(a), high-band variability was essentially constant across the token and showed no SPL-dependency. Instead, nonlinearity arose because the small difference in slope between SPL:high-band and SPL:low-band (x=1.109−0.997=0.112) reduced the proportion of SB variance which could be explained by overall SPL. By contrast, Fig. 7(b) plots the same data for singer C on B3. The greater difference in SPL:high-band and SPL:low-band slopes (x=2.153−0.983=1.17) meant that more than 50% of the variance in SB could still be explained by change in SPL (R^2=67.3%).

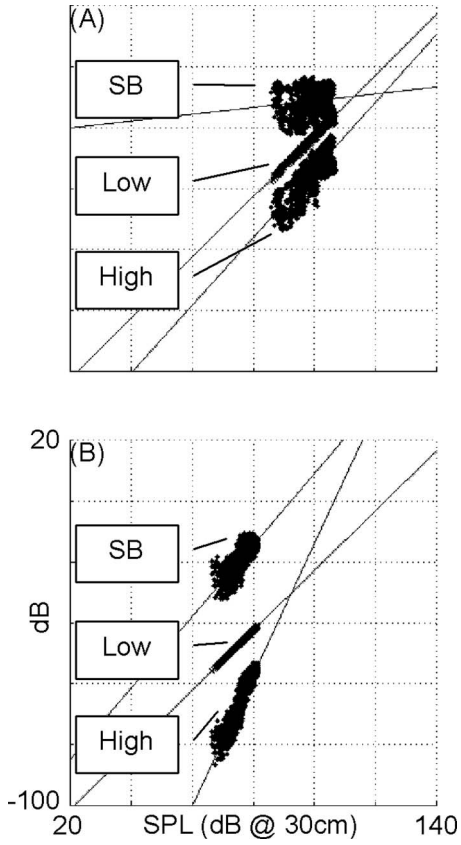


FIG. 7. Plotting SPL (abscissa) against SB and low-band and high-band SPL vectors for singer C on (a) B5 and (b) B3, with linear regression lines for each relationship. The nonlinear SPL:SB relationship on B5 (R^2 =0.028) was not due to SPL-dependency on the SPL:high-band relationship (R^2=0.723) but depended on whether the difference in slope between SPL:low-band and SPL:high-band reduced the proportion of SB variance, which could be explained by change in SPL below the criterion of 50%. By comparison, the larger difference in slope on B3 resulted in a SPL:SB slope of 1.17; change in SPL accounted for 67.3% of SB variance, defining B3 as a linear token.

In terms of SB saturation, visual examination of the nonlinear plots found maximum SB coincided with maximum SPL in 107 of the 110 nonlinear tokens, as detailed in Table III. Only three tokens (two on G#5 and one on A5, all by singer A), showed maximum SB increase then plateau.

The onset of nonlinearity was compared to first-formant (F1) frequency, estimated by linear predictive coding and averaged across the six *messa di voce* tokens on A#3, B3, and C4 (refer Table III). The results varied markedly between singers, ranging from 620Hz (SD 54 Hz) for singer D to 893 Hz (SD 61 Hz) for singer A, which approximates the range D#5 (622 Hz) to A5 (880 Hz). F1 frequency at lower F0s did not predict the onset of nonlinearity, nor did it predict where maximum SB would become essentially unchanged across the token, as illustrated in Fig. 7(a).

F. Low-SPL tail

An apparent SPL-dependency in the SPL:high-band relationship was observed at low SPL on some SPL:SB plots for singers A and B. These showed a low-SPL tail, where SB initially fell with rising SPL (giving a negative slope), then began to rise above a "cut-off" SPL. An example is given in Fig. 1 (A: G#4), where SB decreased up to ~80 dB SPL, then rose as SPL exceeded ~80 dB. Plots following the format of Fig. 7 showed that below the cut-off SPL the slope of SPL:high-band was still positive but was less than SPL:low-band (which approximated unity), i.e., these marked a crossover in SPL:high-band slope from <1 to >1.

The low-SPL tail occurred on all tokens from C#4 to G#4 inclusive for singer A and from A#4 to D5 inclusive for singer B, as marked by vertical dashed lines in Fig. 3(a). Although these F0 ranges did not overlap, the maximum SPL ranges in which the tail occurred were strikingly similar: 87.8–100.6 dB for singer A and 89.9–101.3 dB for singer B. On the other hand, other tokens by these singers with maximum SPL falling within this SPL range did not exhibit a tail. For interest's sake, a second regression line was fitted to data points above the SPL cut-off (drawn as a dotted line in Fig. 1), and the recalculated slope was included in Fig. 3(b) (crosses). Four additional tokens by singer B now met the linearity requirement of $R^2 \geq 0.5$, but there were no additions for singer A. The cut-off SPL itself increased with rising maximum SPL (and therefore also with F0). Despite being crudely estimated, i.e., visually and in steps of 5 dB, cut-off SPL showed a clear relationship to maximum SPL, being on average 20.1 dB (SD 1.4 dB) less than maximum SPL for singer A and 15.4 dB (SD 1.2 dB) for singer B. However, the strong linear relationship between maximum SPL and F0 made it unclear from this study whether the tail was related to SPL, F0, or both.

TABLE III. Frequency (in Hz) of the first formant (F1) for each singer, derived by linear predictive coding and averaged over six tokens on A#3, B4, and C4, inclusive, together with the nearest ST approximation. F1 frequency did not predict the start of nonlinearity in the SPL:SB relationship (where $R^2 < 0.5$). Neither was it related to the F0 range within which maximum SB was essentially constant throughout the token.

Singer	Average (SD) of F1 frequency (Hz)	Nearest ST (Hz)	Start of nonlinearity in ST (Hz)	F0 range (number) of tokens with constant maximum SB
Singer A	893 (61)	A5 (880)	A4 (440)	D5–A#5 (8)
Singer B	689 (49)	F5 (698)	C5 (523)	F5–C6 (11)
Singer C	684 (28)	F5 (698)	G5 (784)	A#5–C6 (5)
Singer D	620 (54)	D#5 (622)	C5 (523)	A5–A#5 (4)
Singer E	780 (43)	G5 (784)	A#5 (932)	Nil

IV. DISCUSSION

This study investigated the effect of F0 on SB behavior across the musical F0 range of five female classically trained singers. The study also investigated whether the relationship between SPL change and SB change was linear and influenced by F0.

The first question was as follows: How does F0 alter SB behavior? Clear F0 influence was found on all SB measures: maximum, minimum, and range. Maximum SB averaged over the five singers increased steadily by 0.27 dB/ST to peak at B4 and then decline, while minimum SB decreased steadily by −0.50 dB/ST to a nadir at E5. By contrast, maximum SPL and minimum SPL rose steadily throughout F0 range at 1.14 and 1.20 dB/ST, respectively. The differences in behavior and proportion between SPL and SB meant that combining all tokens by one singer resulted in a highly scattered plot, which showed no correlation between SPL and SB. However, separating the tokens by F0 identified a gradual horizontal shift rightwards along the SPL axis as F0 rose. Thus, a given SB value was associated with higher SPL values as F0 rose.

The second question considered SB behavior within the *messa di voce* and asked the following: Do female singers exhibit a linear SPL:SB affected by F0? Linearity was defined as a straight-line regression for which $R^2 \geq 0.5$. The SPL:SB relationship within the *messa di voce* was linear for 208 of 318 tokens (65.4%). However, the differences in SPL and SB behaviors across F0, discussed above, led to F0-dependent changes in the SPL:SB ratio. As F0 approached B4, slope increased (i.e., there was a greater gain in SB for a given rise in SPL) and intercept decreased. Above B4, slope decreased and intercept increased, but their relationship also altered: Intercept above B4 was in the order of 19.98 dB less than for a comparable slope at or below B4. Thus, while female singers did exhibit a linear SPL:SB relationship in nearly two-thirds of tokens, the SPL:SB ratio was affected by F0 in slope, in intercept, and in their interaction. The change around B4 is discussed in more detail below.

The third question asked is the following: Is nonlinearity due to greater variance in the SPL:SB relationship at SPL extremes? Correlation of SPL with low-band was almost unity, as would be expected, but correlation of SPL with high-band was also strong. Visual inspection of plots such as Fig. 7 confirmed that SPL:SB nonlinearity was not due to altered rates of high-band rise at SPL extremes within the *messa di voce* (such as greater increase in SB at low-SPL and lesser or no increase at high-SPL). Instead, SPL:SB nonlinearity was a factor of the difference in slope between SPL:high-band and SPL:low-band and of the variance in the SPL:SB relationship throughout the entire *messa di voce*. It was noted that, although linearity ceased above a particular F0, that F0 varied greatly between singers and could not be related to first-formant (F1) frequency, as estimated from the F0 range A#3–C4. However, formant frequencies are known to rise with F0 (Sundberg, 1987), and formant estimation using linear predictive coding becomes unreliable above approximately E4 (Monsen and Engebretson, 1983). While F1 frequency at lower F0s did not predict the onset of linearity, it cannot be assumed that there is no relationship between the rise in F1 frequency and nonlinearity. This requires further study.

Because each *messa di voce* was sung on one F0 with a constant vowel /ɑ/, variation in SPL within each *messa di voce* can be taken to be strongly reflective of changes in the vocal-fold (source) flow waveform due to changing subglottal pressure. Changes in the relationship between SPL and SB observed in this study allow certain inferences to be made about the underlying glottal flow. First, as F0 rose to B4, SB range (maximum SB-minimum SB) increased, as seen in Fig. 4. This suggests that the change in the source flow waveform with rising subglottal pressure followed the full continuum as illustrated in Alku *et al.*, 1998 (their Fig. 3). This continuum begins at soft phonation with a waveform which is substantially symmetrical, i.e., has equivalent opening and closing durations. Increasing subglottal pressure raises flow amplitude and raises the MFDR, which raises SB. If the continuum were truncated with rising F0, e.g., if phonation commenced with a source waveform at an increasingly skewed point along the continuum, SB range would be expected to decrease as F0 rose.

Second, the decrease in minimum SB as F0 rose suggests that the amplitude of the flow waveform in soft phonation increased but remained substantially symmetrical. This is consistent with the increase in minimum SPL, since in soft phonation SPL is highly correlated with the amplitude of the F0 partial (Gauffin and Sundberg, 1989). However, it is contrary to the concept of a fixed relationship as implied by attempts to correlate SPL and SB without considering F0.

Third, as minimum SB decreased with rising F0 but the slope of the regression line remained constant, the regression lines in Fig. 5 can be seen to shift to the right along the SPL axis (abscissa). This suggests that the increase in the amplitude of the flow waveform and in the MFDR with rising subglottal pressure during the *messa di voce* remained proportional as F0 rose, despite commencing with a higher flow amplitude, i.e., the increase in amplitude of the flow waveform with rising SPL was not lessened by an initially higher amplitude in soft phonation.

Sub- and supraglottal effects on the glottal-flow waveform limit the inferences that may be drawn about underlying vocal-fold vibratory behavior (Titze, 2008). However, the consistency of effects observed in our study at B4 and E5 merits further consideration.

An unexpected result was the consistency among the singers for the SPL:SB relationship to alter in linear tokens above B4: The rate of gain (slope) in SB decreased, and intercept for a given slope was in the order of 19.98 dB less. This consistency occurred despite the singers' considerable variability in F1 frequency and in the point (F0) where nonlinearity began. One possibility is a first subglottal resonance ($F_{sub}1$) of 510 Hz, lying between B4 (494 Hz) and C5 (523 Hz) (Cranen and Boves, 1987). For all singers except singer A (who had few linear tokens above G#4), there was a marked increase in slope (i.e., sharper increase in SB for a given increase in SPL) in tokens in the range F4–C5. This would be consistent with increasing skew resulting from combined supraglottal and subglottal inertance (Titze, 1988,

2008). From C5, subglottal compliance would be unfavorable for the divergent glottis of vertically thinner vocal folds associated with higher F0. Thus it might be that the peak in slope around B4 was primarily an effect of the change from subglottal inertance to compliance. Švec *et al.* (2008) observed a register change from chest to head in the region of B4 in an untrained female singer. Their key observations in the transition were the appearance of a posterior glottal chink (which they attributed to decreased arytenoid adduction) and decreased depth of vocal-fold vibration (which they attributed to decreased thyroarytenoid contraction). Their observations may reflect sub- and/or supraglottal inertance effects in a level 2 interaction as proposed by Titze (2008).

Any subglottal resonance effect at B4 did not appear to affect averaged minimum SB, which showed a very strong negative correlation (−0.966) with rising F0 up to E5, inclusive. All five singers showed this nadir of minimum SB in the region of E5, and four singers showed a second dip at G#5. Again, this consistency contrasts with the inter-singer variability in F1 frequency. Above E5, Švec *et al.* (2008) reported a register change from head to whistle, marked by a sudden strong increase in SPL (not apparent in our SPL profiles in Figs. 3 and 4) and cessation of vocal-fold closure and suggesting a vocal tract resonance effect. Inspection of their spectrogram (their Fig. 6) suggests that H1 and H2 simultaneously aligned with F1 and F2. Closer examination of the frequency spectra in our data may yield more information on these changes at B4 and E5. Key points for further analysis include frequency increase of F1, as has been observed with increases in F0 (Sundberg, 1987; Joliveau *et al.*, 2006) and SPL (Gramming and Sundberg, 1988; Holmberg *et al.*,1988); the point (in F0 and SPL) where the fundamental partial becomes the highest-amplitude partial throughout the entire *messa di voce*; and the role of the second and higher formants.

Our results show that F0 affects the SPL:SB relationship. Ternström *et al.* (2006) observed in speech a "spectrum balance saturation level" at which SB stopped increasing and sometimes began to decrease. Only 3 of 318 tokens in our study exhibited some SB saturation (maximum SB rise followed by plateau or fall) within the *messa di voce*. Ternström *et al.* (2006) found that SB saturation occurred at an average of 93.2 dB SPL at 30 cm for females and that the average highest F0 for females was F4 (as reported in Södersten *et al.*, 2005). In our study, maximum SPL at F4 averaged only 89.3 dB. Södersten *et al.* (2005) noted that vocal roughness and instability increased significantly at high SPL, and it is not surprising that classically trained singers might keep below any such saturation level in order to maintain a high musical standard of singing. However, Ternström *et al.* (2006) reported a high degree of scatter in their plots, as was also the case in our plots combining all F0s. Our results suggest that the SB saturation observed by them might have been due at least in part to F0, with speakers gaining the very highest SPL by, in effect, "jumping" across F0 lines. Alku *et al.* (2002) observed that raising F0 as a strategy for increasing SPL gained an average 4 dB in loud speech. It might be that maximum SB at a given F0 represents the point beyond which further increases in SPL by raising subglottal pressure without altering the mode of phonation force the speaker to raise F0 and thus "jump slope," an option not open to the singer constrained by pitch or melody.

Interestingly, the peak of averaged maximum SB on B4 occurred at 94.3 dB SPL at 30 cm, which is very similar to the 93.2 dB (but on F4) of Ternström *et al.* (2006) for female speakers. Vilkman *et al.* (2002) noted that subglottal pressure increased abruptly for male and female speakers above approximately 90 dB SPL at 40 cm (equivalent to 92.5 dB at 30 cm). Such similarity in maximum SPL but difference in F0 between the speakers studied by Ternström *et al.* (2006) and our singers could be due to difference in the mode of phonation. If speakers at high SPL were using strong vocal-fold adductive force (pressed phonation), as observed by Vilkman *et al.* (2002) and Holmberg *et al.* (1988), then a decrease in adductive force would alter the glottal waveform toward so-called "flow phonation," increasing SPL but decreasing SB for a given subglottal pressure (Gauffin and Sundberg, 1989). Higher glottal-flow peak amplitude is characteristic of classically trained singers (Laukkanen and Sundberg, 2008; Titze, 1992). Mode of phonation might thus constitute quasi-parallel lines per F0, with flow phonation having a lower SB intercept (i.e., being displaced rightwards along the SPL axis) similar to the effect of rising F0, making it a further variable in the SPL:SB relationship. In that case, maximum SB would represent the point beyond which any further increase in subglottal pressure would cause a rise in F0 or a change in mode of phonation resulting in lower SPL, both forbidden in the *messa di voce*.

It should be remembered that glottal adduction is not constant during the *messa di voce*. Titze (1988) noted that the singer needs to vary glottal adduction inversely with changing amplitude of vibration to avoid pressed voice at high SPL and breathy voice at low SPL. The low-SPL tails would seem to be an example of the challenge in maintaining this balance, but their apparent relationship to an interaction of maximum SPL and F0 requires further investigation.

A final, though peripheral, observation in our study relates to the proportion of tokens for which SPL:SB achieved $R^2 \geq 0.5$. Singer E achieved linearity on 83% of her tokens, singer C 73%, singer A 60%, singer B 57%, and singer D 48%. This study did not seek to assess the standard of singing. However, it is interesting to note that ranking the singers by proportion of linearity coincided, not only in order but also in degree, with general opinion of their vocal ability in the performing community at the time of data collection. Singer E was regarded as a leading young professional, as evidenced by an already well established international career. Singer C was seen as having an outstanding natural voice. The remaining singers were considered to be of lesser vocal ability, with A and B at a similar vocal standard. This coincidence is interesting because of the difficulty in finding acoustical measures, which equate with perceived vocal quality (Kenny and Mitchell, 2006). Barnes *et al.* (2004) observed a tendency for female opera singers of professional standing to exhibit greater energy in the high-frequency band (2–4 kHz) at high F0s. The small sample size of the present study can only suggest the possibility that a key challenge for female opera singers might be to learn to maintain a

linear relationship between SPL and SB against phonatory and articulatory changes as F0 rises, delaying the transition to a purely flow-regulated SPL mechanism (Pabon, 1991). This might also be related to the singer's ability to delay a shift from (favorable) nonlinear to linear source-filter interaction (Titze, 2008). The *messa di voce* has for centuries been esteemed as the ultimate vocal test, and evenness of timbre is an essential challenge of the exercise. It seems plausible that linearity of the SPL:SB relationship across F0 range could reflect the perception of the singer's vocal skill, and this possibility merits further investigation.

V. CONCLUSION

The results of this study showed that the relationship between SPL and SB altered with F0. In nearly two-thirds of tokens, the relationship was linear at a given F0, but maximum SB and slope (SB gain) peaked at B4 and intercept fell in the order of 19.98 dB above B4. The consistency of the B4 peak among the singers contrasted with their variability in F1 frequency and might reflect a change from subglottal inertance to compliance. Singers were also consistent in a nadir of minimum SB around E5. Nonlinearity in the SPL:SB relationship resulted not from increased variance in the relationship at high SPL but from a combination of reduced SB slope and variance throughout the *messa di voce*. More detailed analysis of the spectral properties underlying SB change may shed light on the changes occurring at B4 and E5.

Alku, P., Airas, M., Björkner, E., and Sundberg, J. (**2006**). "An amplitude quotient based method to analyze changes in the shape of the glottal pulse in the regulation of vocal intensity," J. Acoust. Soc. Am. **120**, 1052–1062.

Alku, P., Vilkman, E., and Laukkanen, A.-M. (**1998**). "Parameterization of the voice source by combining spectral decay and amplitude features of the glottal flow," J. Speech Lang. Hear. Res. **41**, 990–1002.

Alku, P., Vintturi, J., and Vilkman, E. (**1999**). "On the linearity of the relationship between the sound pressure level and the negative peak amplitude of the differentiated glottal flow in vowel production," Speech Commun. **28**, 269–281.

Alku, P., Vintturi, J., and Vilkman, E. (**2002**). "Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation," Speech Commun. **38**, 321–334.

Barnes, J. J., Davis, P., Oates, J., and Chapman, J. (**2004**). "The relationship between professional operatic soprano voice and high range spectral energy," J. Acoust. Soc. Am. **116**, 530–538.

Bloothooft, G., and Plomp, R. (**1986**). "The sound level of the singer's formant in professional singing," J. Acoust. Soc. Am. **79**, 2028–2033.

Boersma, P., and Weenink, D. (**2007**). "Praat: Doing phonetics by computer" (version 5.0.01) (computer program), retrieved Dec. 18 from http://www.praat.org/.

Collyer, S., Davis, P. J., Thorpe, C. W., and Callaghan, J. (**2007**). "Sound pressure level and spectral balance linearity and symmetry in the *messa di voce* of female classical singers," J. Acoust. Soc. Am. **121**, 1728–1736.

Cranen, B., and Boves, L. (**1987**). "On subglottal formant analysis," J. Acoust. Soc. Am. **81**, 734–746.

Frøkjær-Jensen, B., and Prytz, S. (**1976**). "Registration of voice quality," Brüel and Kjaer Technical Review **3**, 3–17.

Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**, 556–565.

Gramming, P., and Sundberg, J. (**1988**). "Spectrum factors relevant to phonetogram measurement," J. Acoust. Soc. Am. **83**, 2352–2360.

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.

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.

Kenny, D. T., and Mitchell, H. F. (**2006**). "Acoustic and perceptual appraisal of vocal gestures in the female classical voice," J. Voice **20**, 55–70.

Laukkanen, A.-M., and Sundberg, J. (**2008**). "Peak-to-peak glottal flow amplitude as a function of F_0," J. Voice **22**, 614–621.

Monsen, R. B., and Engebretson, A. M. (**1983**). "The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction," J. Speech Hear. Res. **26**, 89–97.

Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., and Blaugrund, S. M. (**1996**). "Singing power ratio: Quantitative evaluation of singing voice quality," J. Voice **10**, 228–235.

Pabon, J. P. H. (**1991**). "Objective acoustic voice-quality parameters in the computer phonetogram," J. Voice **5**, 203–216.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Core Team. (**2008**). "nlme: Linear and nonlinear mixed effects models," R package version 3.1-89, retrieved Jan. 14 from http://cran.r-project.org/.

R Development Core Team. (**2008**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, retrieved Jan. 14 from http://cran.r-project.org/.

Rubin, H. J., LeCover, M., and Vennard, W. (**1967**). "Vocal intensity, subglottic pressure and air flow relationships in singers," Folia Phoniatr (Basel) **19**, 393–413.

Sjölander, P., and Sundberg, J. (**2004**). "Spectrum effects of subglottal pressure variation in professional baritone singers," J. Acoust. Soc. Am. **115**, 1270–1273.

Södersten, M., Ternström, S., and Bohman, M. (**2005**). "Loud speech in realistic environmental noise: Phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers," J. Voice **19**, 29–46.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, IL).

Sundberg, J. (**2001**). "Level and center frequency of the singer's formant," J. Voice **15**, 176–186.

Sundberg, J., and Nordenberg, M. (**2006**). "Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech," J. Acoust. Soc. Am. **120**, 453–457.

Sundberg, J., Titze, I., and Scherer, R. (**1993**). "Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source," J. Voice **7**, 15–29.

Švec, J. G., Sundberg, J., and Hertegård, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**, 347–353.

Ternström, S. (**1993**). "Long-time average spectrum characteristics of different choirs in different rooms," Voice **2**, 55–77.

Ternström, S., Bohman, M., and Södersten, M. (**2006**). "Loud speech over noise: Some spectral attributes, with gender differences," J. Acoust. Soc. Am. **119**, 1648–1665.

Thorpe, C. W., Cala, S. J., Chapman, J., and Davis, P. J. (**2001**). "Patterns of breath support in projection of the singing voice," J. Voice **15**, 86–104.

Titze, I. R. (**1988**). "A framework for the study of vocal registers," J. Voice **2**, 183–194.

Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.

Titze, I. R. (**1992**). "Phonation threshold pressure: A missing link in glottal aerodynamics," J. Acoust. Soc. Am. **91**, 2926–2935.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.

Vilkman, E., Alku, P., and Vintturi, J. (**2002**). "Dynamic extremes of voice in the light of time domain parameters extracted from the amplitude features of glottal flow and its derivative," Folia Phoniatr Logop **54**, 144–157.

Rapid pitch correction in choir singers

Anke Grell[a)]
Department of Speech, Music and Hearing, Royal Institute of Technology (KTH), SE-10044 Stockholm, Sweden and Institute for Music Physiology and Musicians' Medicine (IMMM), Hannover University of Music and Drama, Hohenzollernstrasse 47, DE-30161 Hannover, Germany

Johan Sundberg and Sten Ternström
Department of Speech, Music and Hearing, Royal Institute of Technology (KTH), SE-10044 Stockholm, Sweden

Martin Ptok
Department of Phoniatry and Paediatric Audiology, Hannover Medical School, Carl-Neuberg-Strasse 1, DE-30625 Hannover, Germany

Eckart Altenmüller
Institute for Music Physiology and Musicians' Medicine (IMMM), Hannover University of Music and Drama, Hohenzollernstrasse 47, DE-30161 Hannover, Germany

(Received 3 June 2008; revised 6 May 2009; accepted 11 May 2009)

Highly and moderately skilled choral singers listened to a perfect fifth reference, with the instruction to complement the fifth such that a major triad resulted. The fifth was suddenly and unexpectedly shifted in pitch, and the singers' task was to shift the fundamental frequency of the sung tone accordingly. The F0 curves during the transitions often showed two phases, an initial quick and large change followed by a slower and smaller change, apparently intended to fine-tune voice F0 to complement the fifth. Anesthetizing the vocal folds of moderately skilled singers tended to delay the reaction. The means of the response times varied in the range 197– 259 ms depending on direction and size of the pitch shifts, as well as on skill and anesthetization.
 [DOI: 10.1121/1.3147508]

PACS number(s): 43.75.Rs [DD] Pages: 407–413

I. INTRODUCTION

The voice is probably the most important tool for interhuman communication. Concerning the acoustic properties of the voice, several aspects contribute to the information transfer. Among them pitch may be one of the most prominent, being used to code emotional arousal, stress, prosodic and grammatical sentence structure. A deviation from the target pitch contour of an utterance therefore might affect or even destroy the meaning of a spoken sentence.

With respect to pitch control, singers are faced with a specific problem. Not only do they have to reach and preserve the required pitch in a melodic phrase with high precision, but also they need to rapidly adapt their intonation to accompanying instruments or to fellow singers in an ensemble. Professional singers can be expected to be highly skilled in this task, and even semiprofessional choir singers should be able to quickly adapt their pitch, since this is a mandatory prerequisite for the joyful experience of harmonious group singing.

The mechanisms allowing such precise pitch control have been subject to several investigations. Wyke (1967) stressed the relevance of a combined acoustico-laryngeal reflex system. According to Burnett *et al.* (1998) auditory information from the spiral ganglion in the inner ear reaches, via the ventral and dorsal cochlear nuclei, a collection of nuclei constituting the superior olive. From the superior olive, a fiber tract termed the lateral lemniscus projects to the inferior colliculus. The brainstem pathway then projects directly from the inferior colliculus to the periaqueductal gray and then via the nucleus retroambiguus and nucleus ambiguus to the motoneurons of the respiratory system in the spinal cord and to the laryngeal motor system in the nucleus ambiguus.

The long cortical auditory-motor pathway follows the ascending auditory path from the inferior colliculus to the medial geniculate nucleus. The ventral region of the medial geniculate nucleus projects to the primary auditory cortex (area 41), the dorsal region of the medial geniculate nucleus to the secondary auditory cortex (area 42, 22). From there, a pitch control system could project from the anterior cingulate gyrus and from the dorsolateral prefrontal motor region via the limbic system to the periaqueductal gray and finally follow the same structures as the reflex-like pathway to reach the vocal folds and the respiratory system (Duus, 1995; Deetjen *et al.*, 2005; Friedrich *et al.*, 2004; Schönweiler and Ptok, 2004).

Most evidence for these pathways comes from animal research (e.g., Jürgens, 2006). However, empirical findings in humans are also available. Sapir *et al.* (1982), for example, asked their subjects to vocalize with constant pitch and intensity, while receiving auditory stimulation with clicks of different intensities. The electromyography (EMG) of the cricothyroid muscle and the vocal output was assessed and averaged using the respective clicks as triggers. The au-

[a)]Author to whom correspondence should be addressed. Present address: Kleiner Schäferkamp 16E, DE-20357 Hamburg, Germany. Electronic mail: ankegrell@hotmail.com

thors found short latency responses in fundamental frequency (F0) (50 ms) and EMG (11 ms) in response to auditory stimulation, which definitely supports the above outlined brainstem path.

Brown *et al.* (2008) found two sets of activation peaks in their functional magnetic resonance imaging study while subjects were phonating and performing glottal stops. One ventromedial peak was located deep in the central sulcus and one dorsolateral peak in area 6, which is more superficial. These two peaks outline a kind of wedge, extending from a sulcal position ventrally to a gyral position dorsally and anteriorly. They referred to this area as the "larynx/phonation area" and claim that it is the major region for vocal control in the human motor cortex. They further suggested a connection between this area and the nucleus ambiguus, which is involved in activating intrinsic laryngeal muscles.

Over the past decades several investigations have studied reactions to suddenly occurring pitch shifts in the auditory feedback. Burnett *et al.* (1997, 1998) studied the relevance of the feedback of one's own voice to the reaction time for pitch corrections in trained singers. Subjects were instructed to phonate at a constant pitch. Without warning, the auditory feedback was manipulated such that the pitch that the subjects heard of their own voices suddenly shifted by about a semitone. Analyzing the F0 change curves, Burnett *et al.* (1997, 1998) identified two components in most reactions, one early after about 160 ms on average and a longer latency response after approximately 300 ms.

Kestler *et al.* (1999) used the same experimental approach and analyzed the influence of the singers' expertise by comparing opera singers to laymen. When measuring the reaction times, i.e., the time between the change in the auditory feedback pitch and the singer's pitch change, they found in professional singers two peaks in the distribution, one early at about 113 ms and one late appearing at 261 ms, while non-singers showed one peak only at 135 ms.

In both these investigations the results were interpreted as evidence for two pathways involved in pitch control in singing, a rapid mechanism that may operate via the brain stem and a slower mechanism that may operate via the cerebral cortex. Kestler *et al.* (1999) proposed that the slower mechanism includes analysis, e.g., of the precise pitch change required to stay in tune.

Zarate and Zatorre (2008) recently found neurophysiological support for the idea that different regions of the brain are recruited by singers and non-singers for the purpose of pitch control. Thus, when subjects were asked to compensate for a pitch shift in the auditory feedback of their singing, the singers recruited bilateral auditory areas and left putamen, while non-musicians recruited the left supramarginal gyrus and primary motor cortex.

The technique of pitch-shifted auditory feedback has also been applied to experiments with speech production. Xu *et al.* (2004) analyzed the effect of such shifts on the production of Mandarin tone sequences and found that the majority of the compensatory pitch changes occurred at 143 ms. Applying the same pitch-shifted auditory feedback technique, Chen *et al.* (2007) studied the effects in the production of English speech. They found a mean latency of 122 ms in syllable production and a somewhat longer latency in vowel production.

Singers are likely to rely not only on auditory but also on proprioceptive feedback for reaching this level of pitch control. Wyke (1974) found that there are three different types of laryngeal mechanoreceptors: stretch-sensitive myotactic receptors in the intrinsic muscles of the larynx, mucosal receptors in the subglottic mucosa, and articular receptors in the fibrous capsules of the intercartilaginous joints. At any change in tension, subglottic pressure, or posture, the information reaches the brain stem from Ia-afferents, and then motor impulses go back to the larynx via efferent alpha-motoneurons as a reflex. From a longitudinal study of pitch accuracy under different conditions Mürbe *et al.* (2004) concluded that kinesthetic feedback substantially contributes to singers' pitch control.

Larson *et al.* (2008) also demonstrated the influence of kinesthetic feedback on vocal pitch control by anesthetizing singers' vocal folds. They found larger response latencies to pitch-shifted voice auditory feedback in the anesthetic than in the pre-anesthetic condition. Additionally, they developed a mathematical model suggesting that "early in response, kinesthesia alone provides feedback control, but after about 100 ms, auditory feedback also participates."

However, up to now, there is incomplete information available concerning the role of the kinesthetic system in pitch control of the voice.

The aim of the present investigation was to elucidate the physiological mechanisms that underlie pitch control; the authors measured the distribution of reaction times observed when subjects were required to adjust their F0 in response to a sudden shift of an external tuning reference. Measurements were made under three conditions. In experiments 1 and 2 the effect of training was studied by comparing the response time in highly and moderately skilled singers, respectively. Experiment 3 aimed at studying the importance of the laryngeal kinesthetic system by analyzing the effect of anesthetizing the vocal folds.

II. MATERIAL AND METHODS

The *subjects* in experiment 1 consisted of a group of 13 highly skilled female choir singers, mean age 28.9 years (range 23–39 years). All had extensive musical experience. The mean choral experience was 19.3 years, range 5–30 years, and all had been taking solo singing lessons for 10.2 years on average, range 4–17 years.

Eleven female moderately skilled singers constituted the subject group in experiment 2, age range 14–24 years, mean 19.9. On average, they had 10 years of experience of choral singing.

Five female moderately skilled choral singers participated as subjects in experiment 3. They were all experienced singers from different choirs in Hannover, aged 23–27 years, mean 24.8. Two of them participated also in experiment 2.

At the time of the experiments none of the subjects reported any voice disease, hearing loss, or other medical problems.

The *procedure* was basically the same in all three experiments although they were run in different sound insulated rooms. Experiment 1 was run at the Department of Speech Music Hearing, KTH, Stockholm, experiment 2 at Institute for Music Physiology and Musicians' Medicine Hannover, Germany, and experiment 3 at the Department of Phoniatrics and Pediatric Audiology in the Hannover Medical School, Germany.

The subjects were presented with reference dyad stimuli representing the sound of two fellow choral singers singing a fifth interval. Their task was to complement this dyad by singing the missing major third such that a complete major triad resulted.

The two tones constituting the reference stimuli were prepared in the following way. The tones were sung by a female singer on the vowel /u:/ and recorded digitally one by one on a Yakumo 166 MHz computer. The F0 values were adjusted in COOL EDIT (1996) so that they produced a perfect fifth dyad (D4=293.7 Hz and A4=440 Hz). The two tones were then mixed and edited to 5 s duration. Using the COOL EDIT program the F0 values of the two tones were shifted up or down by either a quarter tone or a semitone at a point in time that was randomly selected to occur at $t=1.5$, 2.0, 2.5, 3.0, or 3.5 s following stimuli onset. The resulting 20 stimuli (4 shifts $\times$ 5 times) were recorded in random order and were separated by 10 s silence. The set of 20 stimuli was presented three times to each subject, making a total of 60 stimuli for every participant in each of the three experiments.

The initial pitch level of the missing third was always F#4 (370 Hz). When the reference pitch shifted, the singers' task was to adapt to this shift by changing their F0 accordingly. They were asked to perform this correction as fast and accurately as possible and to phonate continuously until they reached the new target F0.

The procedures were the same in experiments 1 and 2 while the subjects in the former were more experienced as singers than those in experiment 2.

The anesthesia used in experiment 3 was applied by an experienced phoniatrician. Prior to local anesthesia it was ruled out that the subjects suffered from an allergy against lidocain. Then, they were asked to phonate a neutral vowel /e/. During phonation and while the tongue was gently fixed by the examiner, lidocain was sprayed onto the vocal cords using a commercially available spraying device (Xylocain Pumpspray, AstraZenica, Wedel, Germany). Two sprays of 20 mg lidocain 5% were applied. Due to the position of the tip of the spray device not only vocal folds but also the whole aditus ad laryngis was anesthetized. The recording was started when the subjects reported a throat numbness and swallowing difficulties. According to clinical experience these effects are reliable signs of an anesthesia, sufficient for performing phonosurgery.

The anesthesia was administered in a room close to the experimental setup, so that the subjects could start the experiment immediately after the anesthesia was applied. The experimental procedure was approved by the local ethical committee.

As the subjects' ability to perceive frequency differences was crucial to the outcome of the main experiment, two additional tests were run in order to determine their *just noticeable difference* (JND) for F0. These tests were conducted after the first and after the second set of 20 stimuli. However, as in experiment 3, the anesthetization effect lasted for no longer than about 20 min, all three experimental blocks of singing were performed in one sequence, and the JND tests were carried out afterwards.

In the first JND test, the vowel /u:/ sung at an F0 of 293,8 Hz was presented followed by the same vowel sound with an F0 increased by $n \times 1.73$ cent, $20 \geq n \geq 0$. Thus, the smallest stimulus difference was 0 cent, the second smallest was 1.73 cent, the third was 2×1.73 cent, and so on, and the largest was 20×1.73 cent, corresponding to 299.7 Hz. In the second test, the vowel sounds were replaced with sine tones. The subjects were sitting in front of a computer screen listening to two consecutive tones, each 2 s long and separated by a 1.16 s pause. The tones were presented monophonically over two loudspeakers. The singers were asked in a forced choice condition to decide whether or not they heard the same stimulus twice and to click accordingly on a yes or a no button on the screen. If three consecutive answers were correct, the frequency difference between the following two tones was decreased, and if an answer was wrong, the frequency difference in the following stimulus was increased. After about 10 min the test was stopped, and the subject's JND was automatically calculated, thus specifying the subject's ability to discriminate very small pitch differences.

In all three experiments the two stimulus loudspeakers and the microphone were located at the corners of an equal-sided triangle with sides of 1 m. The distance between the microphone and the subject's mouth was 5 cm. However, the *equipment* used in experiment 1 differed from that used in experiments 2 and 3. In experiment 1 the stimuli were played from a Dell Optiplex GX 240 computer and presented monophonically over two loudspeakers (Fostex Personal Monitor 6301B). The subjects' responses were picked up by an omnidirectional microphone (TCM 110). The subjects' responses were recorded on one channel of a digital audio tape (DAT) recorder while the computer stimulus was recorded on the other channel. The stimuli and responses were then digitized and transferred to a Laptop (Targa Visionary N251C2) in COOL EDIT by means of a custom-made program (DINODAT, S. Granqvist). Translations between sound file formats were performed by another custom-made program (AUDIOFIL, S. Granqvist). The F0 analysis was performed by PRAAT (version 3.8.27).

In experiments 2 and 3 a Yakumo computer (166 MHz) and Typhoon PS 56 loudspeakers were used. The responses were picked up by a Sennheiser Black Line microphone connected to a Viscount professional (MM 8) mixing console and an Aiwa HD-S200 DAT recorder. The responses were digitally transferred from the DAT recorder directly into the COOL EDIT program of the Yakumo computer.

For the *analysis* it was important to select a method suitable for the material collected. A common way to identify the onset of an F0 reaction is to define a variance estimate for the F0 signal prior to the change and then to identify the point in time when the signal exceeds this average variation window. However, several subjects were singing with wide

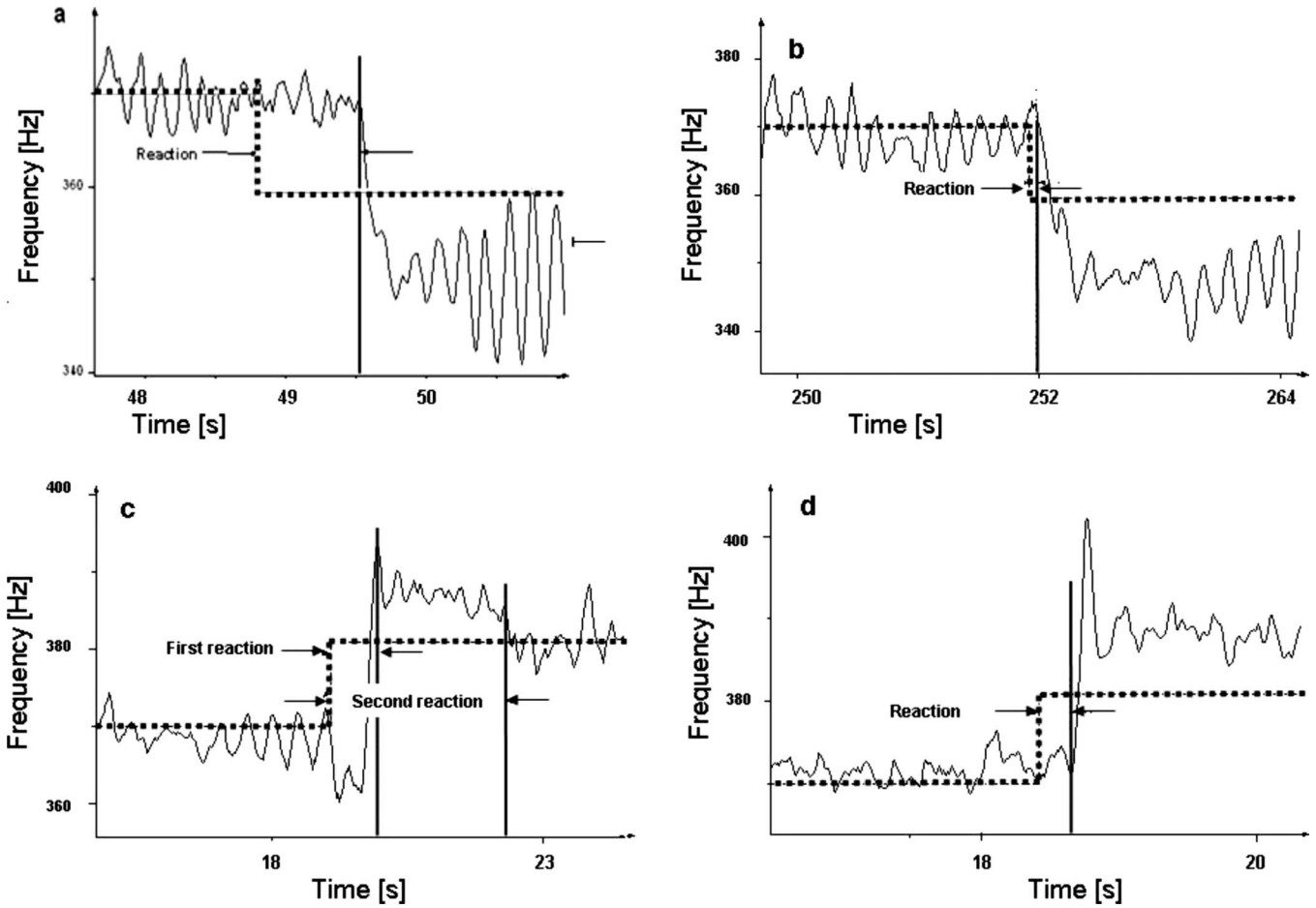


FIG. 1. Examples of frequently observed patterns of pitch shift corrections: (a) slow reaction after 723 ms, (b) quick reaction after 50 ms, (c) double pitch change after 220 and 1290 ms, and (d) overshoot reaction after 220 ms. These examples were taken from the professional and the moderately skilled singers' groups.

vibrato extent that generated a very large average variation of F0. Other subjects sang with no or quite small vibrato, producing a quite narrow average variation window. Applying the above definition of the onset of F0 change would produce a systematic delay of the reaction in subjects who were singing with a wide vibrato. Therefore, another definition of the onset of F0 shift had to be applied.

The reaction time, defined as the interval between the F0 shift of the reference and onset of the subject's response, was measured using the ORIGIN 6 program. This program displayed the subject's F0 contour together with the target F0 in the same graph [see Figs. 1(a)–1(d)]. The onset of the response was manually identified in these graphs. The criterion used for this onset was an F0 shift that approached the new target without interruption, disregarding occasional vibrato undulations during the F0 change. The time coordinates of the onset of the F0 shifts were then collected in an Excel file. To check data reliability a different experimenter (co-author JS) measured 75 randomly chosen reaction times. The mean difference amounted to 41 ms [standard deviation (SD) 45 ms].

The choice of *statistical analysis* was complicated by the fact that there were different numbers of subjects in the different experiments. Also, some subjects participated both in experiments 2 and 3. Moreover, the data were not normally distributed according to a Kolmogorov–Smirnov test. This suggested the application of non-parametric tests. A Kruskal–Wallis test showed that the ranks were significantly different. Therefore, the averages of the different conditions were tested pairwise by means of a Mann–Whitney statistics.

As the observed values showed considerable scatter, particularly for the highly skilled subjects, it was necessary to exclude outliers. All values lying above three times the interquartile range, counted from the median, were considered as outliers. The minimum latency value for outliers was just above 740 ms. To eliminate outliers in all groups from the computation of means, all values greater than 740 ms were excluded from the calculation of means. This implied that a total of 17 values were excluded, 11 for the highly skilled, 5 for the moderately skilled, and 1 for the anesthetized subjects. The total number of observations thus became 1381.

III. RESULTS

The JND for frequency, averaged across all subjects in all three experiments, amounted to 10.5 cent (SD 4.4) for the voice stimulus and to 14.2 cent (SD 5.5) for the sine tone. Thus the subjects were able to hear much smaller pitch differences than the quarter tone (50 cents) used for the stimulus shift in the main experiment.

In the singing task, subjects performed in accordance with the instructions in most cases, i.e., they changed their

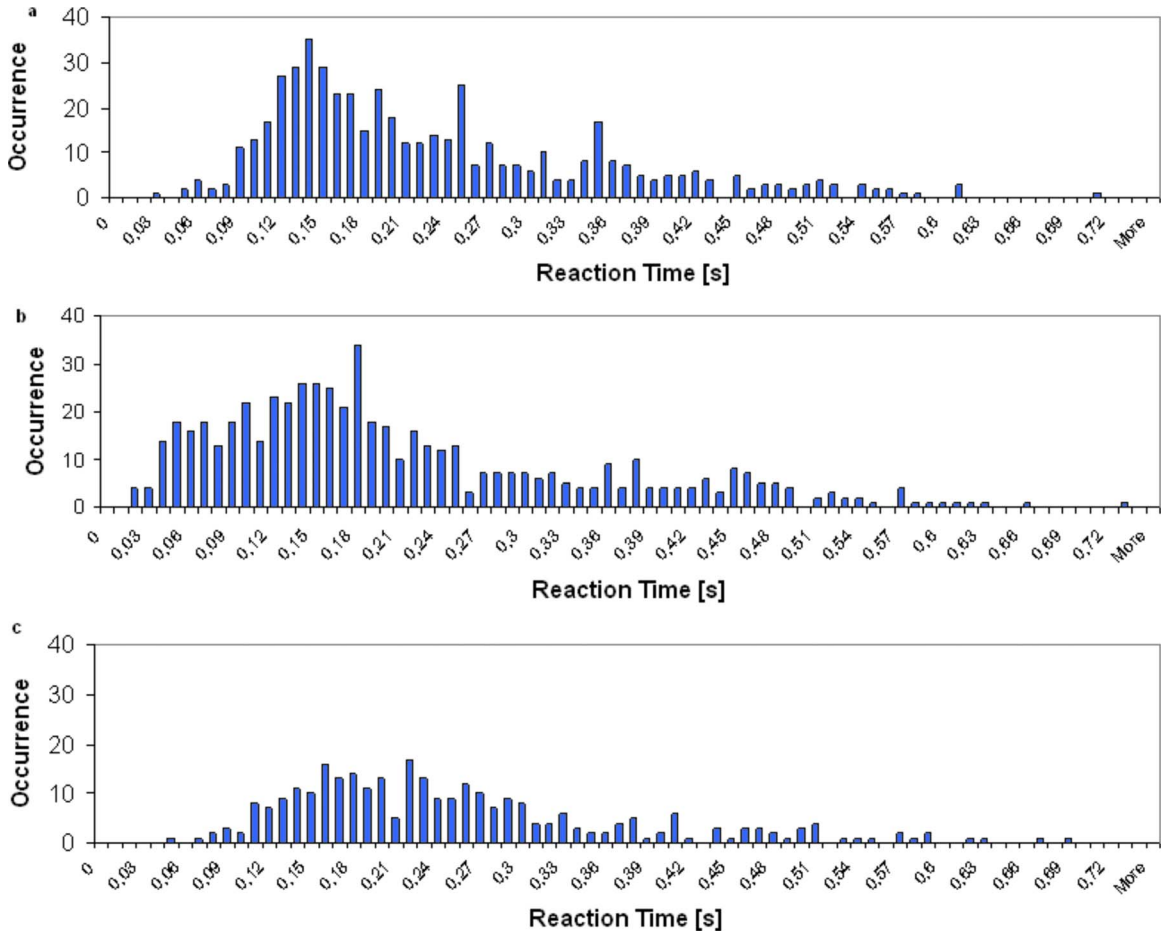


FIG. 2. (Color online) Distribution of reaction times for the three subject groups: (a) highly skilled singers, (b) moderately skilled singers, and (c) moderately skilled singers with anesthetized vocal folds.

F0 quickly and continuously until the new target was reached. In some cases, however, they interrupted their singing when the reference stimulus changed, and in some cases the onset of F0 change was very slow, thus preventing determination of the onset of the F0 shift. This problem was particularly frequent in the group of highly skilled singers. These cases were eliminated from the subsequent analysis (253, 82, and 7 in experiments 1, 2, and 3, respectively).

The curves shown in Figs. 1(a)–1(d) show typical examples of different types of responses, all observed for a quarter-tone shift of the stimulus. Figures 1(a) and 1(b) illustrate a slow (723 ms) and a quick (50 ms) reaction time. Interestingly, in both these examples of a descending F0 shift, the rising phase of the vibrato cycle was interrupted by the F0 drop. Figure 1(c) presents an example of a pitch change containing two parts. The first one, appearing after 220 ms, is large (+152 cent peak-to-peak, which was three times as large as the stimulus shift). The second part is small (−32 cent) occurring after 1290 ms. Figure 1(d) shows an example of a marked overshoot, amounting to +134 cent.

The reactions of the two subjects who participated in both experiments 2 and 3 did not differ notably from the rest of these respective groups. Experiment 3 was run several weeks later than experiment 2. Thus there was no indication of a learning effect.

The distribution of reaction times for the different groups is shown in Fig. 2 and in terms of box plots in Fig. 3.

The means and the standard deviations are listed in Table I. As the Kruskal–Wallis test indicated that the ranks were significantly different, the means of the groups' latencies were compared pairwise. A Mann–Whitney test for two

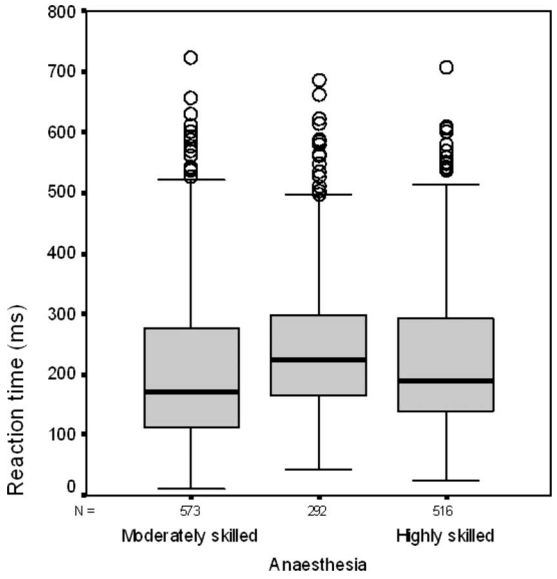


FIG. 3. Box plots of the results for the three indicated subject groups. The boxes represent the interquartile range, the heavy horizontal lines show the medians, and the bars ±1 standard deviation. Unfilled circles show values exceeding the standard deviation.

TABLE I. Means in milliseconds and standard deviations for all subject groups and all conditions. *N* is the number of trials.

	All conditions			Semitones			Quarter tones			Ascending			Descending		
	Mean (ms)	SD	*N*	Mean (ms)	SD	*N*	Mean (ms)	SD	*N*	Mean (ms)	SD	*N*	Mean (ms)	SD	*N*
Highly skilled singers	227	120	516	216	113	265	238	126	251	236	112	246	219	126	270
Moderately skilled singers	206	135	573	197	135	283	216	136	290	199	127	285	214	143	288
Anesthetized vocal folds	251	122	292	243	129	147	259	114	145	246	110	147	257	133	145
Groups pooled	223	128	1381	214	127	695	233	129	686	222	120	678	224	136	703

independent samples showed that the difference between the means for the moderately skilled and the anesthetized groups was significant ($p<0.001$). The same was true both for the difference between the anesthetized and highly skilled groups ($p<0.001$), and the difference between the moderately and highly skilled groups ($p<0.001$). The mean values showed that anesthesia slowed down the reaction time in all conditions. Moreover, high skill tended to be associated with slightly longer latencies than moderate skill in all conditions. The Mann–Whitney test further revealed that the direction of the pitch change did not have a significant effect on the latency ($p>0.1$) comparing the mean latencies of the three subject groups pooled. Looking at the subject groups separately, the moderately trained subjects, with and without anesthesia, showed the same nonsignificant result as the pooled group (both $p>0.5$), while highly skilled singers showed significantly shorter latencies in performing descending intervals ($p<0.01$). With regard to interval size, the quarter-tone shifts were associated with a significantly slower reaction in all subject groups as a pool ($p<0.01$), as well as the separated groups (in all groups $p<0.05$).

IV. DISCUSSION

The authors studied how quickly choir singers adapted their intonation to a change in a reference pitch. Burnett *et al.* (1997) carried out a somewhat related experiment in which the pitch of singers' auditory feedback was suddenly shifted while the subject was instructed to sustain a tone and to keep the pitch constant. Under these conditions they observed a mean latency of 159 ms (variation range 104–223 ms) between the onset of the pitch shift of the auditory feedback and the onset of the subject's attempt to correct F0. This value is smaller than the mean values of 227 and 206 ms observed in experiments 1 and 2. It is probable that the responses reported by Burnett *et al.* (1997) were unconscious while the authors analyzed a deliberate F0 change. In their experiment, the subjects were frequently even unaware of the pitch changes that they were performing. Our subjects were performing a far more complex task that required analysis of both the direction and the magnitude of the pitch shift.

The highly skilled group showed greater mean latency values than the moderately skilled group. It is tempting to speculate that this may be related to the neurophysiologic difference between singers and non-singers observed by Zarate and Zatorre (2008). Quoting these authors: "Through years of training and experience, singers have learned that they need to monitor their auditory feedback closely to ensure that their notes are produced correctly." Moderately skilled singers may be less perfectionistic than highly skilled singers in their attempts to produce intended pitches. Such perfectionism may require recruiting the motor cortex and hence take more time. Anesthetizing the vocal folds slowed down the reaction time, thus showing that the kinesthetic feedback represents an important part of singers' pitch control system. Mürbe *et al.* (2004) found experimental evidence for the same conclusion.

The semitone intervals showed a shorter latency than the quarter-tone intervals. This may simply be an effect of training, since the singers are more accustomed to halftone than to quarter-tone intervals.

Burnett *et al.* (1997) and Kestler *et al.* (1999) suggested the involvement of a double pathway for the control of vocal pitch. Also other investigations have suggested the existence of such a double pathway. Furthermore, as mentioned, Zarate and Zatorre (2008) observed that different subjects used different brain areas for the purpose of pitch control. One possible manifestation of a double pathway in our results would be that some reactions were early and some late, i.e., a double-peaked distribution. The histograms in Fig. 2, however, fail to show any distributions of this type.

On the other hand a double pathway could also be manifested in a different way in our data, such that several individual F0 curves showed two such phases in many cases. A typical example of such a curve was shown in Fig. 1(c). Constructing models of the human pitch control system seems a promising avenue for further elucidating the pathways involved. For example, Guenther *et al.* (2006) presented a neural network model of the components which corresponded to regions of the cerebral cortex and the cerebellum, including premotor, motor, auditory, and somatosensory cortical areas. Another attempt was presented by Xu *et al.* (2004).

Several mechanisms are likely to have been involved in the subjects' behaviors, some conscious and others unconscious. Reflexes would cause quick, though imprecise reactions. The authors may speculate that very quick F0 shifts such as the one illustrated in Fig. 1(b) (50 ms) are produced by such an imprecise reflex system, while slow reactions such as in Fig. 1(a) (723 ms) and the second F0 change in Fig. 1(c) (1290 ms) rely on conscious control.

Also Hain *et al.* (2000) observed an early and a late component in subjects' F0 changes. Interestingly, subjects showed such double responses more often when they were instructed to produce specific voluntary responses. The first response was often incorrect, whereas the second, later one

almost always followed the instruction. This seems another support for the idea of an early subcortical unconscious reaction and a consciously controlled cortical mechanism.

Our results seem relevant also to choral singing practice. They showed that most singers have a rather quick reaction to a shift in an external auditory pitch reference. In a choir the fellow singers generally provide this reference and it is important that the entire ensemble synchronize their pitch changes. At least in many amateur choirs, one individual in each choir voice tends to act as a leader and the fellow singers mainly follow this leader when they sing. Indeed, this structure of the choral group was systematized in the Baroque period, when the leaders were called "concertisten" and the followers "riepienisten" as the German music performance expert Wilhelm Ehmann described in his work (Ehmann, 1961). Incidentally, this structure is systematically implemented also in today's orchestras in terms of the Concert Masters of each instrument group.

Xu *et al.* (2004) found in Mandarin speakers that the mean latency of 164 ms was comparable to the mean syllable duration. This seems comparable to the situation in singing. In coloratura singing typical note durations lie in the vicinity of 125 ms (Huron, 2001; Lindblom and Sundberg, 2007). This is too short to allow a slow correction of pitch. This implies that the strategy of ripienisten simply shadowing concertisten in choral ensembles is not appropriate in fast music. Instead singers need to hit the target closely enough on the first attempt. Singers must know the entire sequence of pitches before they start to sing it.

V. CONCLUSION

Our investigation of the reaction times in different groups of singers has shown that, typically, highly skilled choir singers reacted to a change in a pitch reference after 227 ms while moderately skilled choir singers' reactions appeared after 206 ms. Anesthetization of moderately skilled singers' vocal folds tended to slow down the reaction. The reaction time was typically shorter when the reference was shifted by a semitone rather than by a quarter tone. Indications of a double pathway for F0 controls were not found in the distribution of reaction times, but many singers' F0 curves during the transitions showed two phases, an initial quick and large change followed by a slower and smaller change, apparently aiming at a fine-tuning of the pitch.

ACKNOWLEDGMENTS

Mikael Bohman kindly assisted with some of the MATLAB processing. Henrik Jansson kindly discussed with us some control-theory aspects of this investigation. Friederike Lipka helped complementing the statistics. Kathrin Lüerßen did the anesthesia of the vocal folds in the Department of Phoniatry and Paediatric Audiology at Hannover Medical School. Dietrich Parlitz who supported the first idea of the study and the pilot experiment. The work on experiment 1 was done in Stockholm while participating in the Marie Curie Fellowship Program, sponsored by the European Commission.

Brown, S., Ngan, E., and Liotti, M. (**2008**). "A larynx area in the human motor cortex," Cereb. Cortex **18**, 837–845.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). "Voice F0 responses to pitch-shifted auditory feedback: A preliminary study," J. Voice **11**, 202–211.

Chen, S., Liu, H., Xu, Y., and Larson, C. (**2007**). "Voice F0 responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am. **121**, 1157–1163.

Deetjen, P., Speckmann, E. J., and Hescheler, J. (**2005**). *Physiologie (Physiology)* (Elsevier Urban Fischer, Amsterdam).

Duus, P. (**1995**). *Neurologisch-Topische Diagnostik (Neurologic-Topical Diagnostics)* (Thieme, Stuttgart).

Ehmann, W. (**1962**). *Concertisten und Ripienisten in der h-moll Messe Johann Sebastian Bachs (Concertists and Ripienists in the Mass in B Minor by Johann Sebastian Bach)* (Bärenreiter, Kassel, Germany).

Friedrich, G., Biegenzahn, W., and Zarowka, P. (**2004**). *Phoniatrie und Pädaudiologie (Phoniatrics and Paedaudiology)* (Huber, Bern).

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K., (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Huron, D. (**2001**). "Tone and voice: A derivation of the rules of voice-leading from perceptual principles," Music Percept. **19**, 1–64.

Hage, S. R., Jürgens, U., and Ehret, G., (**2006**). "Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey," Eur. J. Neurosci. **23**, 3297–3308.

Kestler, C., Parlitz, D., and Altenmüller, E. (**1999**). "Experimentelle studie zur schnellen tonhöhenkorrektur bei sängern und nichtsängern (An experimental study in rapid pitch correction of professional singers and non-singers)," Diploma thesis, Hannover University for Music and Drama, Hannover, Germany.

Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (**2008**). "Interactions between auditory and somatosensory feedback for voice F0 control," Exp. Brain Res. **187**, 613–621.

Lindblom, B., and Sundberg, J. (**2007**). "The human voice in speech and singing," in *Springer Handbook of Acoustics*, edited by T. Rossing (Springer, Heidelberg, Germany), Chap. 16, pp. 669–712.

Mürbe, D., Pabst, F., Hofmann, G., and Sundberg, J. (**2004**). "Effects of a professional solo singer education on auditory and kinesthetic feedback—A longitudinal study of singers' pitch control," J. Voice **18**, 236–241.

Sapir, S., McClean, M. D., and Larson, C. R. (**1983**). "Human laryngeal responses to auditory stimulation," J. Acoust. Soc. Am. **73**, 315–321.

Schönweiler, R., and Ptok, M. (**2004**). *Phoniatrie und Pädaudiologie (Phoniatrics and Paedaudiology)* (Hannover Medical School, Hannover, Germany).

Wyke, B. D. (**1967**). "Advances in the neurology of phonation: Phonatory reflex mechanisms in the larynx," British J. Communic. **2**, 2–14.

Wyke, B. D. (**1974**). "Laryngeal neuromuscular control systems in singing. A review of current concepts," Folia Phoniatr Logop **26**, 295–306.

Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifter auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Zarate, J. M., and Zatorre, R. J. (**2008**). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," Neuroimage, **40**, 1871–1887.

Singing in congenital amusia

Simone Dalla Bella[a)]
Department of Cognitive Psychology, University of Finance and Management in Warsaw, 01-030 Warsaw, Poland and BRAMS, H2V 4P3 Montreal, Canada

Jean-François Giguère and Isabelle Peretz
Department of Psychology, University of Montreal and BRAMS, H2V 4P3 Montreal, Quebec Canada

(Received 1 October 2008; revised 3 April 2009; revised 16 April 2009)

Congenital amusia is a musical disorder characterized by impaired pitch perception. To examine to what extent this perceptual pitch deficit may compromise singing, 11 amusic individuals and 11 matched controls were asked to sing a familiar tune with lyrics and on the syllable /la/. Acoustical analysis of sung renditions yielded measures of pitch accuracy (e.g., number of pitch errors) and time accuracy (e.g., number of time errors). The results revealed that 9 out of 11 amusics were poor singers, mostly on the pitch dimension. Poor singers made an anomalously high number of pitch interval and contour errors, produced pitch intervals largely deviating from the score, and lacked pitch stability; however, more than half of the amusics sang in-time. Amusics' variability in singing proficiency was related to their residual pitch perceptual ability. Thus, their singing deficiency might be a consequence of their perceptual deficit. Nevertheless, there were notable exceptions. Two amusic individuals, despite their impoverished perception, sang proficiently. The latter findings are consistent with the existence of separate neural pathways for auditory perception and action.
 [DOI: 10.1121/1.3132504]

PACS number(s): 43.75.Rs, 43.75.Yy, 43.75.Cd [DD] Pages: 414–424

I. INTRODUCTION

Singing is widespread in the general population. In spite of the general belief that most people are inept at singing, there is increasing evidence that the majority can carry a tune. Nonmusicians' sung performance is consistent both within (Bergeson and Trehub, 2002; Halpern, 1989) and across subjects (Levitin, 1994; Levitin and Cook, 1996) in terms of starting pitch and tempo. In addition, occasional singers can proficiently sing a well-known song provided that they perform it at a slow tempo (Dalla Bella *et al.*, 2007).

However, a significant proportion of the population have notorious difficulties with singing in-tune. 10%–15% of the population was classified as poor-pitch singers when singing a familiar song (Dalla Bella *et al.*, 2007) and when imitating unfamiliar pitch patterns (Pfordresher and Brown, 2007). Many factors (e.g., poor perception, poor motor planning and execution, and poor sensory-motor integration) can cause inadequate singing (see Pfordresher and Brown, 2007). One obvious cause of poor-pitch production is a faulty pitch perceptual system. Accurate singing requires fine perceptual monitoring of the vocal output, especially when learning melodies and when singing along with others. Through feedback analysis, singers can correct and adjust vocal performance. A poor-pitch perceptual system is likely to affect this feedback mechanism, hence leading to poor-pitch singing. This deficient perceptual monitoring of vocal performance is likely to be found in tone deafness. Self-defined "tone-deaf" or "unmusical" individuals consider poor singing as a hallmark of their musical deficiencies (see Sloboda *et al.*, 2005, for a discussion). Yet evidence is scant about singing proficiency in these individuals. This condition, more recently referred to as "congenital amusia," has been mostly studied and defined in terms of poor perceptual abilities (Ayotte *et al.*, 2002; Foxton *et al.*, 2004; Peretz, 2001; Peretz *et al.*, 2002, 2007; Peretz and Hyde, 2003).

Congenital amusia can be traced to degraded pitch perception abilities (Foxton *et al.*, 2004; Hyde and Peretz, 2004). Amusic individuals exhibit lower accuracy than matched controls in detecting pitch changes that are smaller than one semitone[1] and ones that are out-of-key whereas they are normal at detecting time changes in the same sequences (Hyde and Peretz, 2004; Peretz *et al.*, 2007). To the extent that singing reflects perceptual abilities, we predict poor performance for small pitch intervals and little sensitivity to key distance in singing in amusia.

In one prior study (Ayotte *et al.*, 2002), amusics' singing was judged by peers as impaired with regard to normal performance. The deficit mostly concerned, but was not limited to, the pitch dimension. This supports the notion that amusics' poor singing may result from an impoverished perceptual system (Ayotte *et al.*, 2002). Nonetheless, one congenital amusic was judged to sing accurately. This case raised the intriguing possibility that perceptual disorders may not completely account for singing impairments. This possibility finds some support in the recent discovery that congenital amusic individuals are able to reproduce the pitch direction of two successive single tones despite being unable to judge pitch direction (Loui *et al.*, 2008). It is noteworthy that spared production was confined to pitch direction; amusics' reproduction of pitch interval size was inaccurate. The reverse dissociation (i.e., impaired performance with spared

a)Author to whom correspondence should be addressed. Electronic mail: sdallabella@vizja.pl

TABLE I. Congenital amusics' characteristics, individual scores on the MBEA (percents of correct responses), and on a pitch change detection task (from Hyde and Peretz, 2004). For the pitch change detection task, average percents of Hits-F.A. across 25-, 50-, and 100-cent pitch changes are reported.

	AG	AM	AS	EL	FA	GC	IC	MB	PT	SR	TC	Controls (SD)
Gender	F	M	F	F	F	F	M	F	F	F	M	9F 2M
Age (years)	49	66	63	54	64	59	61	55	63	53	35	56 (6)
Education	17	16	15	19	16	19	19	17	16	18	15	15.8 (2)
Handedness	R	R	R	R	R	R	R	R	R	R	R	R
I.Q	N/A	116	117	110	N/A	128	107	120	108	N/A	N/A	N/A
M.Q.	N/A	135	134	114	N/A	137	112	130	114	N/A	N/A	N/A
						MBEA						
Scale	53.3[a]	60[a]	63.3[a]	53.3[a]	66.7[a]	56.7[a]	50[a]	46.7[a]	53.3[a]	53.3[a]	66.7[a]	92.1 (6)
Contour	70[a]	60[a]	63.3[a]	53.3[a]	70[a]	56.7[a]	50[a]	46.7[a]	53.3[a]	56.7[a]	66.7[a]	87 (8)
Interval	50[a]	56.7[a]	60[a]	53.3[a]	70	73.3	50[a]	73.3	53.3[a]	73.3	70	88.2 (9)
Rhythm	76.7	73.3[a]	76.7	63.3[a]	66.7[a]	96.7	50[a]	93.3	63.3[a]	76.7	90	90 (7)
Metric	70	66.7	60[a]	73.3	66.7	70	56.7[a]	70	66.7	50[a]	76.7	87.6 (9)
Memory	76.7	53.3[a]	73.3	66.7[a]	76.7	73.3	50[a]	76.7	50[a]	43.3[a]	76.7	85.5 (9)
Composite score	66.1[a]	61.7[a]	66.1[a]	60.5[a]	69.5[a]	71.1[a]	51.1[a]	67.8[a]	56.7[a]	58.9[a]	74.5[a]	89 (6)
Pitch change detection (Hyde and Peretz, 2004)	N/A	35.8	71.2	64.2	66	59.9	50.8	70	81.2	N/A	N/A	N/A

F=female, *M*=male, and N/A=not available.
[a]Below cut-off score (as indicated in Peretz *et al.*, 2003).

perception) is more common. Poor-pitch singing can occur in cases of normal perception (Bradshaw and McHenry, 2005; Dalla Bella *et al.*, 2007; Pfordresher and Brown, 2007; Wise and Sloboda, 2008). Similarly, brain damage (e.g., lesion of the right fronto-temporal regions) can selectively impair sung performance without affecting perception (Schön *et al.*, 2003, 2004). In sum, these findings point to the possibility of two separate streams for auditory perception and action (Griffiths, 2008), thus extending to the auditory modality the idea of independent perceptual and action systems previously observed in vision (i.e., the dorsal and ventral neural pathways, Goodale *et al.*, 1991).

In the study of Ayotte *et al.* (2002), singing proficiency was assessed by peer judgments. Such judgments are common in studies on singing (e.g., Alcock *et al.*, 2000a, 2000b; Hébert *et al.*, 2003; Racette *et al.*, 2006; Schön *et al.*, 2004). However, discrepancies among subjective ratings of impaired singing are frequent (e.g., Kinsella *et al.*, 1988; Prior *et al.*, 1990). This problem is likely the effect of music notation and perceptual constraints, which may impinge on judgments. For example, experts tend to integrate pitch and time information when embedded in a musical context (Jones and Pfordresher, 1997; Peretz and Kolinsky, 1993). In other words, judges cannot provide fine estimates of accuracy in terms of pitch and time, while keeping these two dimensions separate. Acoustical methods represent a powerful alternative to perceptual judgments (Dalla Bella *et al.*, 2007; Murayama *et al.*, 2004; Terao *et al.*, 2006). Based on acoustical features such as tone onset and pitch height, objective and reliable measures of singing proficiency on pitch and time dimensions can be obtained. With this method, we showed that occasional singers can sing proficiently a well-known melody from memory, provided that they sing at a slow tempo (Dalla Bella *et al.*, 2007). Nevertheless, a minority of individuals cannot sing proficiently: Their difficulty is confined to poor-pitch production with no evidence of a concomitant perceptual deficit, as assessed by a task that required the detection of pitch and time incongruities in unfamiliar melodies (Peretz *et al.*, 2008).

In the present study, the singing abilities of 11 adults with congenital amusia were examined with an acoustically-based method. Congenital amusics were asked to sing a well-known song in Quebec (*Gens du pays*, by Gilles Vigneault) with lyrics and on the syllable "ta" or "la."[2] Measures of pitch and time accuracy were yielded by an acoustical analysis of sung performance (as in Dalla Bella *et al.*, 2007). Since amusics' perceptual deficit affects mostly the pitch dimension (Hyde and Peretz, 2004; Peretz *et al.*, 2007), we predicted that amusics would sing out of tune while being able to sing in-time. Furthermore, we expected poor singing to be related to the severity of the music perceptual difficulties (as assessed by the Montreal Battery of Evaluation of Amusia; Peretz *et al.*, 2003), and to the degree of impairment in detecting pitch changes (as assessed in Hyde and Peretz, 2004). Nonetheless, in line with previous suggestions, we also expected to find some amusic individuals who would be able to sing accurately without awareness. Finally, in keeping with the findings by Loui *et al.* (2008), we predicted that amusics, despite impaired production of pitch interval sizes, would be able to produce the correct pitch contour.

II. METHOD

A. Participants

Eleven congenital amusics aged between 35 and 66 (M=57 years) participated in the study (see Table I). Amusics had no neurological or psychiatric history. Eight participants were assessed in previous studies (Ayotte *et al.*, 2002;

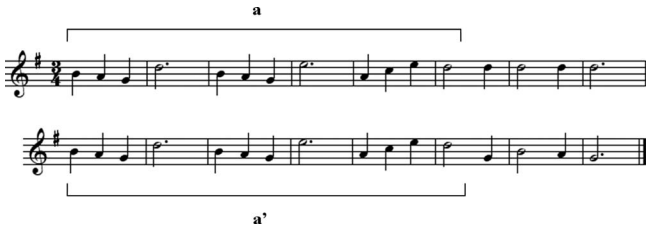


FIG. 1. Score of the chorus of *Gens du Pays*.

Hyde and Peretz., 2004). Amusics obtained a composite score between 55.1 and 74.5 on the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz *et al.*, 2003), hence below the cut-off score for amusia (77.6; Peretz *et al.*, 2003).

As can be seen in Table I, congenital amusics exhibited impaired music perception mostly affecting the melodic dimension (i.e., the scale, contour, and interval tests of the MBEA). A deficit in perceiving acoustical pitch was confirmed in each of the eight amusics tested on a pitch change detection task (Hyde and Peretz, 2004). In this task, participants had to detect a pitch change in five-tone standard and comparison sequences. In standard sequences (no change), all the tones had the same pitch level (1047 Hz). In comparison sequences (with change), the fourth tone was displaced at one of five pitch distances (from 25 to 300 cents) upward or downward from the pitch of the other tones. The results in the pitch change detection task are reported in Table I. At pitch distances smaller than one semitone, the amusics obtained a lower score than controls, whose performance was above 92% correct.

A control group ($n=11$) matched to amusics for age, gender, education, and musical training but with no musical difficulties participated in the study. Participants were remunerated for participating in the experiment.

B. Material and procedure

Participants were asked to sing the chorus of the song *Gens du pays* (Vigneault and Rochon, 1976), well-known in Quebec and typically sung to celebrate birthdays. The same tune was studied in our prior work on singing proficiency in the general population (Dalla Bella *et al.*, 2007). As can be seen in Fig. 1, the chorus of *Gens du pays* comprises 32 notes with a vocal range of less than an octave and a stable tonal center. Each note is associated with a syllable. The segment a' is a repetition of a; this characteristic of the melody served to assess pitch stability within the same performance (see below).

Participants sang the chorus of *Gens du pays* twice (*singing with lyrics* condition): at the beginning of the experiment (test 1) and immediately afterwards (test 2). Then, after a short break, participants were asked to sing the same melody twice on the syllable /ta/ or /la/ (*singing on /la/* condition). Participants did not receive cues (e.g., first notes of the melody) or indications about the beginning pitch of the melody. Performances were recorded in a laboratory setting with a Shure 565SD microphone (sampling frequency =44.1 kHz) directly onto an IBM-compatible computer using COOLEDIT software.

C. Acoustical analysis of sung performance

Only complete performances (i.e., with 32 notes, as indicated in the score) were analyzed. Acoustical analyses of sung renditions were performed on the vowel groups (e.g., /i/ in "mi"), determined by visual inspection of the waveform and of the spectrogram. Vowel groups are the best targets for acoustical analysis, given that vowels carry the maximum of voicing (e.g., Murayama *et al.*, 2004). Moreover, the initiation of the vowel group is well-suited to indicate the onset of musical tones, because vowel onsets, rather than consonant onsets, are typically synchronized with the beat in singing (Sundberg and Bauer-Huppmann, 2007). The onset of vowel groups was considered as the *note onset time*. The median of the fundamental frequencies within the vowel group served to measure *pitch height*. Note onset times and pitch heights were used to compute various measures of pitch and time accuracy.

1. Pitch dimension variables

Initial pitch is the pitch of the first note used to assess absolute pitch level.

Pitch stability is the difference between the produced pitch in the melody segment a and in the repetition a'. The absolute difference in semitones between the 12 corresponding notes (e.g., note 1 in segments a and a', note 2 in segments a and a', and so forth) was computed. Pitch stability is the mean of these absolute differences. The larger this mean difference, the more instable the pitch.

Number of contour errors refers to the number of produced intervals that deviated in direction from their respective notated intervals. Pitch direction was counted as ascending or descending if the sung interval between two notes was higher or lower by more than one semitone. If pitch direction was different to that noted in the musical score, it was counted as an error.

Number of pitch interval errors indicates the number of produced intervals that deviated in magnitude from their respective notated intervals. An error was scored when the sung interval was larger or smaller by one semitone than the interval prescribed by the notation. It is noteworthy that pitch interval errors were coded irrespectively of pitch direction (e.g., if a singer produced a one-tone ascending interval instead of a one-tone descending interval, this was not scored as a pitch interval error).

Interval deviation measures the size of the pitch deviations, by averaging the absolute difference in semitones between the produced intervals and the intervals prescribed by musical notation. Small deviation reflects high accuracy in relative pitch.

2. Time dimension variables

Tempo is the mean inter-onset-interval (IOI) of the quarter note.

Number of time errors indicates duration deviations from the score. When the duration of the sung note was 25% longer or shorter than its predicted duration based on the preceding note, as prescribed by the musical notation, this

TABLE II. Mean values for pitch and time variables obtained in the singing with lyrics and singing on /la/ conditions for congenital amusics and their controls.

	Singing with lyrics		Singing on /la/	
Variable	Amusics *M* (SE)	Controls *M* (range)	Amusics *M* (SE)	Controls *M* (range)
		Pitch dimension		
Initial pitch (Hz)				
Males	122.0 (8.1)	123.7 (117.5–129.8)	115.8 (−)[a]	132.1 (128.0–136.3)
Females	214.1 (13.3)	218.9 (175.2–273.9)	226.8 (24.8)	216.6 (168.2–294.5)
Pitch stability (semitones)	1.3 (0.2)[b]	0.5 (0.3–0.7)	1.3 (0.2)[c]	0.5 (0.2–1.0)
No. of contour errors	6.1 (2.0)[c]	1.0 (0–3.5)	3.4 (2.0)	0.8 (0.0–2.5)
No. of pitch interval errors	13.0 (1.9)[b]	3.6 (0–8.0)	11.0 (2.7)[b]	3.4 (0.5–8.5)
Interval deviation (semitones)	1.3 (0.2)[b]	0.5 (0.3–0.8)	1.1 (0.2)[d]	0.5 (0.3–0.8)
		Time dimension		
Tempo (mean IOI, ms)	314.4 (10.9)	299.0 (257.8–343.0)	334.5 (19.1)[c]	291.6 (254.2–332.7)
No. of time errors	3.2 (0.5)	2.4 (0–5)	2.2 (1.4)	1.2 (0.0–3.5)
Temporal variability (CV IOIs)	0.18 (0.02)[c]	0.12 (0.08–0.16)	0.17 (0.05)	0.10 (0.06–0.17)
Rubato	0.6 (0.08)	0.7 (0.4–0.9)	0.6 (0.1)	0.7 (0.1–0.9)

[a]Only one participant.
[b]$p<0.01$.
[c]$p<0.05$.
[d]Marginally significant ($p=0.07$).

was considered as a time error. The first and last notes were not used to compute time errors.

Temporal variability is the coefficient of variation (CV) of the quarter-note IOIs, calculated by dividing the standard deviation of the IOIs by the mean IOI.

Rubato consistency is an additional measure referring to variations in the timing of onsets of subsequent musical notes as compared with the musical notation. An example of rubato is observed when musicians speed up at the beginning of a musical phrase and slow down toward the end of it (e.g., Todd, 1985). Rubato consistency was obtained from the correlation of the quarter-note IOIs for the segment *a* with the IOIs for corresponding notes in segment *a′* (a similar measure was proposed by Timmers *et al.* (2000) in piano performance). High correlation reflects high consistency in the rubato pattern. Throughout the paper, for simplicity, the term *rubato* will refer to rubato consistency.

III. RESULTS AND COMMENTS

A. Singing with lyrics: Group results

All amusics and controls were able to produce complete renditions (i.e., 32 notes) with lyrics. Means and variability of pitch and time variables in the singing with lyrics condition for amusics and their controls are reported in Table II. The reported values are averaged across repetitions (i.e., test 1 and test 2). The measures of pitch and time accuracy were highly correlated across repetitions in both amusics and controls (with Spearman rho values between 0.60 and 0.98, average $p<0.01$)[3] with the exception of rubato (for amusics, Spearman rho=0.41, p=n.s.; for controls, rho=0.29, p=n.s.), pitch interval deviation, and pitch interval errors for controls (with rhos=0.49 and 0.50, respectively, ps=n.s.).

Amusics were impaired on the pitch dimension showing a large number of pitch interval errors, contour errors, lower pitch stability, and average pitch interval deviation larger than 1.2 semitones. However, amusics' difficulties were not confined to the pitch dimension. Amusics exhibited larger temporal variability (i.e., CV of the IOIs) than controls. It is noteworthy that amusics did not sing at a faster tempo than controls. Thus, it is unlikely that amusics' poor singing is due to tempo differences. Nevertheless, the amusics who performed at a slower tempo were more accurate than those singing at fast tempi, as revealed by the significant negative correlations between tempo (mean IOI) and pitch interval deviation (rho=−0.62, $p<0.05$), number of pitch interval errors (rho=−0.61, $p<0.05$), temporal variability (rho=−0.64, $p<0.05$). In controls, only a positive correlation between tempo (mean IOI) and temporal variability reached significance (rho=0.64, $p<0.05$), thus suggesting rather an increase in temporal variability at slower tempi. These discrepancies are probably due to the fact that the range of tempi in controls was much smaller (85 ms) than in amusics (134 ms). Yet, in general, these results are in keeping with the speed-accuracy trade-off previously found in occasional singers (Dalla Bella *et al.*, 2007).

Further analyses were conducted on pitch and time errors made by amusics and controls. In amusics, the number of pitch interval errors increased with the number of time errors (rho=0.62, $p<0.05$). This correlation did not reach significance in controls (rho=−0.25, p=n.s.). However, only 10% of pitch interval errors made by the amusics co-occurred with time errors (4% of errors in controls). Thus, errors on the pitch and time dimensions were relatively independent. To examine whether pitch interval errors led to produce notes in-key or out-of-key, the tonality of the sung

melody was inferred based on the starting pitch; the notes in- and out-of-key were detected by approximating the produced pitches to the closest notes in the chromatic scale. Amusics produced on average 7.3 pitch errors that were in-key (i.e., 55.9% of total number of errors), and 5.7 notes that were out-of-key (44.1%). These productions did not differ from chance performance (as revealed by binomial tests; note that chance level differs for in-key—6 out of the 11 chromatic pitches in the octave, or 55% of the possible tones—and for out-of-key—5 out of 11 possible pitches, or 45%). Pitch interval errors occurred more often on strong beats ($M=7.4$ errors, 58.6% of the total number of errors) than on weak beats ($M=5.5$ errors, 41.4%) ($t(10)=4.48$, $p<0.01$) in amusics. This difference is significantly above chance (=50% for errors on strong beats, and 50% for errors on weak beats), as attested by a binomial test ($p<0.05$). Thus, even if about half the pitch errors were in-key, these are likely to be noticed because strong beats are the most salient events in melodies (Jones *et al.*, 2002). This effect was not observed in controls. Time errors always occurred on weak beats in both amusics and controls.[4]

In order to examine whether amusics are more impaired on small than large pitch intervals, we examined pitch interval deviations for each of the 31 pitch intervals from the chorus (spanning from the unison to nine semitones; see Fig. 1 for the musical notation of the correct intervals and Fig. 2 for the data, with panel (a) referring to amusics and panel (b) to controls). Positive and negative deviations indicate interval expansion and compression, respectively. The produced intervals were analyzed in 2 groups (amusics vs controls) by 5 interval sizes (2, 3, 4, 7, and 9 semitones)[5] by 2 deviation types (compression vs expansion) repeated-measures Analysis of Variance (ANOVA), taking intervals as the random factor.[6] Group and deviation types were considered as the within-item factors, and interval size as the between-item factor. As can be observed in Fig. 2, the effect of interval size was different in each group, as revealed by a significant group × interval size × deviation type triple interaction ($F(4,22)=15.33$, $p<0.001$). Separate interval sizes by deviation type ANOVAs were run for amusics and controls. Amusics exhibited a monotonic dependency of interval deviation on interval size, with a tendency to compress large intervals ($F(4,22)=40.88$, $p<0.001$). In controls, interval deviation did not significantly vary as a function of interval size. Thus, contrary to expectations, amusics' large pitch deviations from target intervals are not limited to small intervals (e.g., zero and two semitones). Amusic singing cannot be explained by a fine-grained perceptual pitch deficit alone. This finding will be examined in more detail in the discussion.

B. Singing with lyrics: Individual results

The individual data for pitch and time accuracy are presented in Figs. 3 and 4, respectively. As can be seen, individual amusics were more deviant from controls on the pitch than on the time dimension. Nonetheless, not all amusics exhibited impaired performance. For example, the amusic PT performed within the range of controls on all variables, and

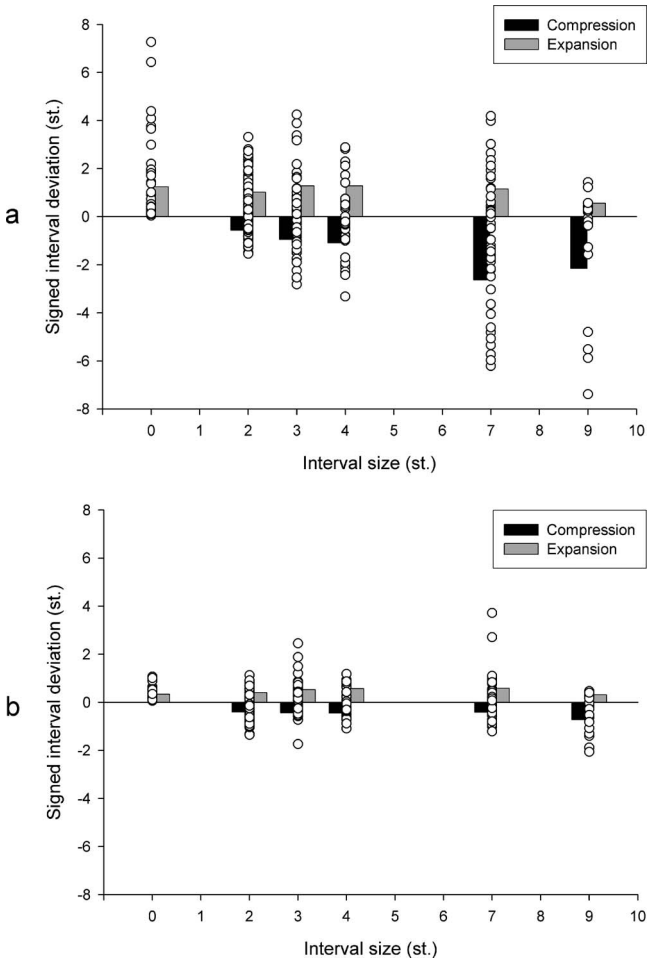


FIG. 2. Average pitch interval deviations in terms of compression and expansion for each interval of the chorus of Gens du pays (from unison to nine semitones) produced by amusics (a) and controls (b). A positive deviation indicates an extension of the target interval and negative deviation, a compression of the interval. The dots indicate individual performances.

GC's performance fell outside the range of controls in terms of pitch stability only. To assess more thoroughly amusics' individual performance, we examined cases in which performance departed from the average obtained from the control group by more than 2 standard deviations (SD) (mildly impaired) or 3 SD (very impaired) on each variable (see Table III). The most common deficits observed in amusics, with the exception of PT, affected the pitch dimension (i.e., impaired pitch stability, increased number of pitch interval errors and contour errors, and larger pitch interval deviation from the score). In four cases (AM, EL, FA, and IC), poor-pitch singing was associated with large time variability. In no cases, however, did deficits of the time dimension occur in isolation. When ranked in terms of singing proficiency, AM and IC appear as the most impaired, and PT and GC as the least impaired (for examples of productions see Dalla Bella *et al.*, 2009).[7] AM and IC made numerous pitch interval errors (22.5 and 20, respectively) and contour errors (20 and 16); their renditions were characterized by large pitch interval deviations from the target (by 2.6 and 2.1 semitones on average; see also Fig. 2), low pitch stability (1.2 and 2.9 semitones), and high temporal variability (CVs of the IOIs =0.33 and 0.30, respectively). In contrast, PT and GC sang

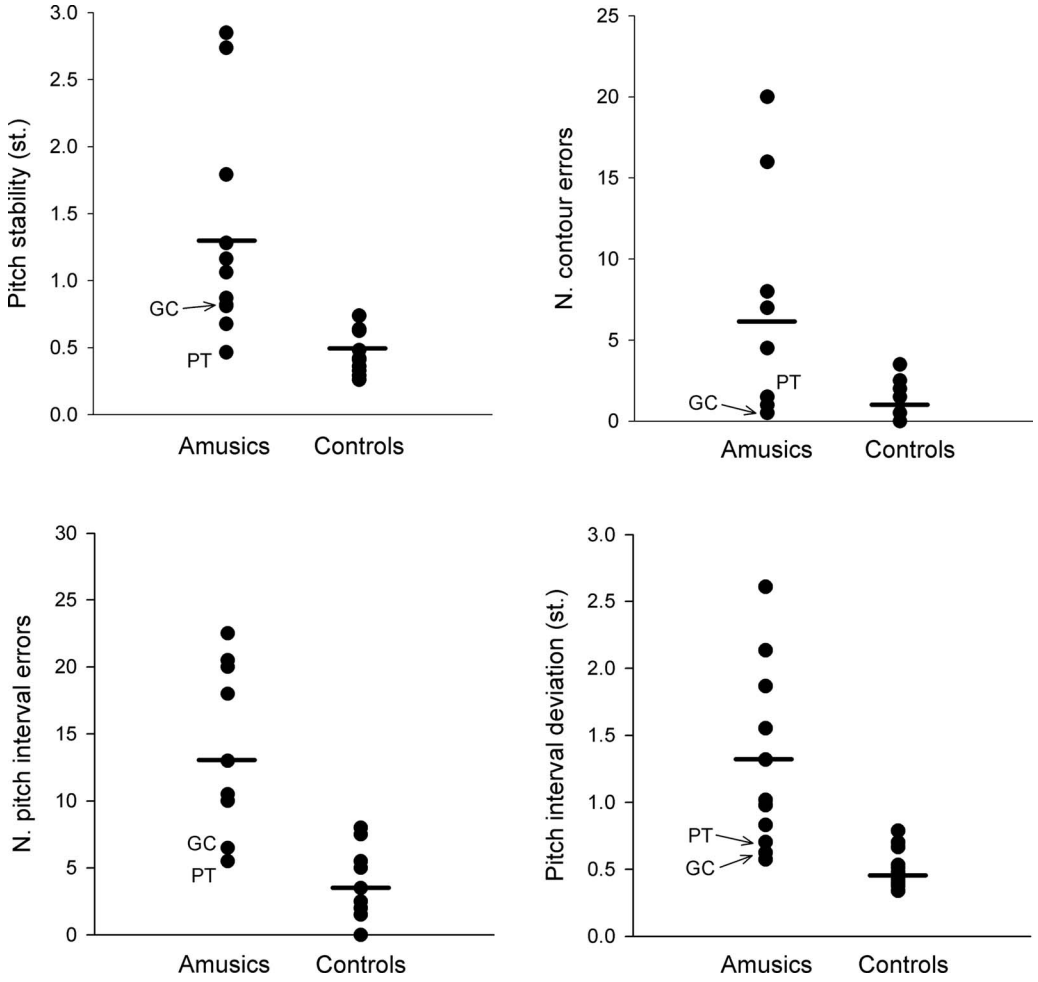


FIG. 3. Amusics' and controls' individual results for pitch stability, number of contour errors, number of pitch interval errors, and pitch interval deviation in the singing with lyrics condition. Horizontal lines indicate group averages.

quite proficiently. Five other amusics (AG, AS, EL, GC, and PT) did not make more pitch contour errors than controls (with less than two contour errors). These five amusics were impaired in perceiving melodic contour, as attested by the MBEA (see Table I). Hence, they were able to produce the correct pitch direction while being unable to perceive it, in line with the results of Loui *et al.* (2008) with single intervals. However, there are important differences between the results reported by Loui *et al.* (2008) and those obtained here. First, at least three amusics (AG, GC, and PT) were not impaired in producing interval sizes. Second and more important, six amusics were inaccurate in producing both pitch direction and pitch interval size. In sum, production in interval imitation tasks (Loui *et al.*, 2008) does not seem to predict performance in musical tasks.

C. Singing on /la/

Singing on /la/ turned out to be an extremely difficult task for amusics. Only 5 amusics out of 11 were able to produce complete performances (i.e., including 32 notes) when asked to sing on /la/ while all controls succeeded in producing complete performances. The other six amusics could only produce a few notes when asked to sing on /la/. The amusics who produced complete performances (i.e., AS, AG, GC, MB, and TC) were among the least severely impaired on the MBEA (mean composite score=69.1), as can be seen in Table I. In addition, the five amusics who produced complete performances had memory scores that lied in the low but normal range on the MBEA (see Table I). However, PT, who was able to sing proficiently with lyrics, failed to sing on /la/. It is noteworthy that the observed difficulty in amusics to sing on /la/ as compared to singing with lyrics was not found in controls. Pitch accuracy was comparable in the two conditions in controls; in addition, controls were more accurate on the time dimension when singing on /la/ than when singing with lyrics ($t(10)=2.41$, $p<0.05$).

Means and variability of pitch and time variables in the singing on /la/ condition for the minority of amusics who performed the complete song on /la/ and their controls are reported in Table II. The reported values are averaged across repetitions (i.e., test 1 and test 2). As can be seen, amusics' impairment was limited to the pitch dimension, as shown by their reduced pitch stability and larger number of pitch interval errors than controls. In addition, amusics sang slower than controls. To examine amusics' individual performance, we indicated in Table IV cases in which performance departed from the average obtained from the control group by more than 2 SD (mildly impaired) or 3 SD (very impaired) on each variable. As can be observed, two of the five amusics

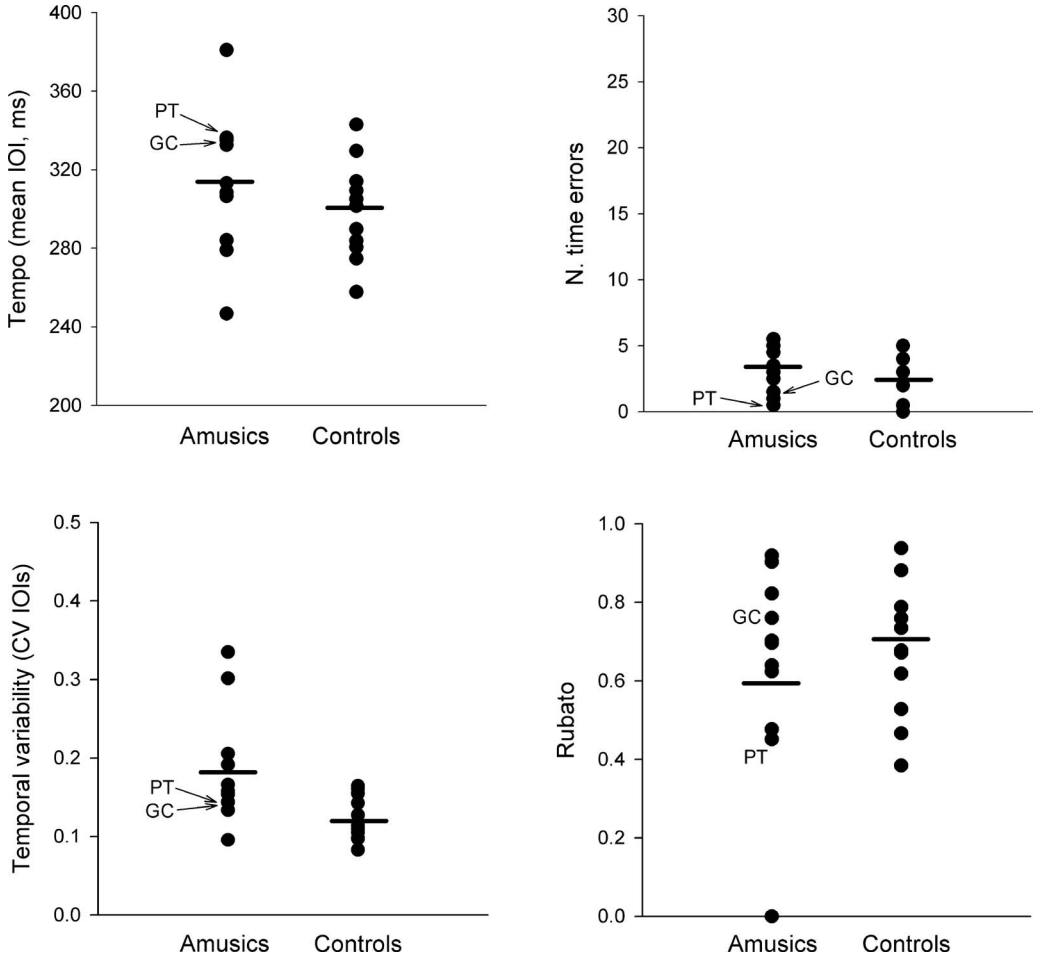


FIG. 4. Amusics' and controls' individual results for tempo, number of time errors, temporal variability, and rubato in the singing with lyrics condition. Horizontal lines indicate group averages.

(AS and MB) who were able to sing on /la/ were impaired on both the pitch and time dimensions. Deficits on the time dimension were always associated with poor-pitch singing (excluding the case of GC, for whom a difference in tempo cannot be considered as a real deficit). In addition, in keeping with what was found in the singing with lyrics condition, AS and MB were among the most impaired. Finally, we compared the performance of amusics who were able to sing on /la/ ($n=5$) in the two conditions (singing with lyrics and on /la/) using non-parametric tests (Wilcoxon). No significant differences were found.

D. Amusics' perceptual abilities

As noted above, the two amusics who were the most impaired on the MBEA were also the ones who sang most

TABLE III. Amusics' individual performance for pitch and time accuracy measures in the singing with lyrics condition ($n=11$).

Variable	AG	AM	AS	EL	FA	GC	IC	MB	PT	SR	TC
					Pitch dimension						
Pitch stability	--	--	--	−	--	−	--	−	+	--	+
No. of contour errors	+	--	+	+	−	+	--	--	+	--	--
No. of pitch interval errors	+	--	−	−	−	+	--	--	+	--	--
Interval deviation	+	--	--	−	--	+	--	--	+	--	--
					Time dimension						
Tempo	--	+	+	+	+	+	+	+	+	+	+
No. of time errors	+	+	+	+	+	+	+	+	+	+	+
Temporal variability	+	--	+	−	−	+	--	+	+	+	+
Rubato	--	+	+	+	+	+	+	+	+	+	+

+=normal, −=mildly impaired (>2 SD from controls), and −−=severely impaired (>3 SD).

TABLE IV. Amusics' individual performance for pitch and time accuracy measures in the singing on /la/ condition (n=5).

Variable	AG	AS	GC	MB	TC
Pitch dimension					
Pitch stability	--	--	+	+	--
No. of contour errors	+	+	+	--	−
No. of pitch interval errors	+	−	+	--	--
Interval deviation	+	--	+	--	--
Time dimension					
Tempo	--	+	−	+	+
No. of time errors	+	--	+	--	+
Temporal variability	+	--	+	+	+
Rubato	+	+	+	+	−

+=normal, −=mildly impaired (>2 SD from controls), and −−=severely impaired (>3 SD).

poorly (i.e., AM and IC). Similarly, amusics with mild perceptual deficits (i.e., with MBEA scores closest to cut-off) exhibited little impairment in sung performance (e.g., GC). However, there are notable exceptions. TC, the least impaired of the amusics (composite score=74.5), was very poor at singing (i.e., with 13 pitch interval errors, interval deviation=1.32 semitones on average, and 7 contour errors when singing with lyrics). Conversely, PT is one of the most severe cases of amusia (composite score=56.7), and yet sang as proficiently as controls. The observation that a severe amusic can sing proficiently, both in terms of pitch interval size and pitch direction, is very intriguing and is reported here for the first time. Indeed, the amusic individuals described by Loui *et al.* (2008) could correctly reproduce pitch direction of isolated intervals. Yet these same amusics were inaccurate in imitating pitch interval sizes and hence cannot be considered as proficient singers.

The puzzling case of PT may, in fact, be best explained by relatively spared pitch perception. PT obtained the best score on the pitch change detection task (81.2%, see Table I; this score is significantly lower than controls but clearly above chance). The pitch change detection task is a perceptual task that does not engage short-term memory whereas the MBEA is highly loaded in working memory demands. Most MBEA tests require the subject to hold a melody in memory in order to compare it to the same melody or to a slightly modified one. The fact that PT suffers from severe amusia (as indicated by her MBEA scores) may be due to poor short-term memory. In sum, her spared abilities to perceive small pitch changes may be sufficient to support feedback analysis, and ultimately, to support proficient singing. This possibility deserves further enquiry.

We also examined whether the singing results obtained in the other seven congenital amusics (i.e., AM, AS, EL, FA, GC, IC, and MB) who completed the pitch detection task could be related to their performance in pitch perception. Correlations were computed between accuracy in detecting pitch changes and pitch stability, number of contour errors, number of pitch interval errors, and interval deviation, as obtained in the singing with lyrics condition. Parametric correlation tests revealed that lower pitch change detection scores were associated with larger pitch instability (r=−0.87, $p<0.01$), more contour errors (r=−0.83, $p<0.05$), larger interval deviation (r=−0.74, $p<0.05$), and more pitch interval errors (r=−0.67, p=0.07, marginally significant). Thus, amusics' difficulties in pitch production were tied to their impairment in detecting pitch changes in an acoustical context. However, there was one exception. GC, one of the most proficient amusic singers performed quite poorly (<60% of Hits-F.A.) when asked to detect a pitch change. Thus, it seems possible to find cases who are able to sing in-tune despite the presence of a severe perceptual pitch disorder.

E. Analyses of melodic complexity

The musical material used in the production task (Chorus of "Gens du pays") might be easier than the musical selections presented in the MBEA. These differences in stimulus complexity may account for the pitch perception-performance mismatch observed here in some of the amusics. For example, cases such as PT (who exhibit poor scores on the MBEA but good singing abilities) might be due to differences in the musical material presented in the two testing situations. This possibility was examined by computing melodic complexity, based on both pitch and rhythm-related factors, for stimuli used in the three melodic tests from the MBEA (scale, contour, and interval tests), and for the chorus of Gens du pays, using the expectancy-based model of melodic complexity (Eerola and North, 2000). Complexity, as obtained with MATLAB MIDI toolbox (Eerola and Toivianen, 2004), is a value that is based on the Essen collection (with mean=5 and SD=1). High values indicate large complexity. Gens du pays has a complexity of 4.54. Likewise, the average complexity of the melodies used in the MBEA is 4.53 (range=2.58–6.11). A further test was conducted by selecting ten melodies from the MBEA, which have complexity values in the vicinity of Gens du pays (i.e., the first five melodies above, and the first five below the complexity value 4.54). With this subset of melodies of average complexity, amusics are still performing below the cut-off score (<72.2%; see Peretz *et al.*, 2003), with the exception of GC (77%). The observed differences between perception and performance do not seem to result from differences in stimulus complexity.

IV. DISCUSSION

The results of the present study show that congenital amusia is characterized by poor singing. Amusic individuals could not maintain a stable pitch throughout singing and were inaccurate at producing pitch intervals; however, many succeeded in singing in-time. This singing pattern is consistent with the amusics' perceptual profile, which is characterized by impaired melodic pitch perception. As predicted, amusics' variability in singing proficiency was related to their ability to detect pitch changes (Hyde and Peretz, 2004). Amusics with markedly impaired ability to detect pitch changes were the most unstable in pitch production, made numerous pitch interval and contour errors, and exhibited significant pitch interval deviation from the score. However,

amusics' impairment was not confined to pitch intervals of one semitone, as one would expect from their deficient detection of such pitch changes; the implications of this finding will be discussed below. Poor singers' deficits were very consistent across repetitions, thus indicating stable impairment. However, there were a few notable exceptions. PT, despite severely impaired melodic pitch discrimination on the MBEA, was able to sing quite proficiently with lyrics. In addition, GC, one of the best amusic singers, had severely impaired pitch perception. Thus, it seems possible to find cases who are able to sing relatively in-tune despite the presence of a severe perceptual pitch disorder. We will return to this paradoxical dissociation below.

When singing the same song without lyrics but on /la/, more than half of the amusics failed to sing more than a few notes. None of the controls experienced this difficulty. On the contrary, normal singers tend to sing more in-time when singing without words. This striking finding fuels the long-held debate as to whether lyrics and melody in songs are represented in a separate or integrated fashion. Lyrics and melody in songs have been previously treated as parts of an integrated representation (e.g., Serafine *et al.*, 1984, 1986); yet, neuropsychological evidence points toward separate codes in perception, memory, and performance (e.g., Besson *et al.*, 1998; Hébert *et al.*, 2003; Peretz, 1996; Samson and Zatorre, 1991).

This dissociation between singing with and without lyrics can be explained by weak memory traces of the musical component of songs. This hypothesis is consistent with the observation that five out of the six amusics who could not sing on /la/ were also impaired on the incidental memory test of the MBEA. Severe amusics might be able to produce complete performances with lyrics due to the benefit of the strong association between melody and text in memory or by relying on an integrated representation of melody and lyrics. When the task requires the association of a well-known melody to new speech segments, such as the repeated syllable /la/, retrieval of melodic information from memory alone may become impossible. Faulty memory for musical information may encourage amusics to rely on a melody-lyrics compound code. This faulty memory representation of melodies cannot be explained by melody complexity since complexity was comparable in the singing and memory tasks. Further work is needed to understand the origins of this poor memory for melodies.

Most congenital amusics sang out of tune but a few sang in-time. This finding mirrors neuropsychological dissociations between pitch and time previously uncovered in the perceptual domain with patients suffering from acquired and developmental music disorders (e.g., Peretz, 1990; Peretz and Kolinsky, 1993; Peretz *et al.*, 1994). This dissociation supports the notion that pitch and time processing may be governed by separable mechanisms both in perception and in performance.

Another intriguing observation relates to the apparent separability between perception and production. In the present study, as mentioned above, low pitch accuracy in singing is associated with poor-pitch discrimination, highlighting the close coupling between perception and action. Yet, amusics were inaccurate at producing pitch intervals far above one semitone, whereas such large pitch intervals lie well above the anomalously high threshold for detection of pitch changes in amusics (see Hyde and Peretz, 2004). Therefore, deficient low-level pitch perception cannot be the sole cause of amusics' poor-pitch singing. Indeed, amusics are also deficient in the melodic tests of the MBEA, which require comparing pitch intervals in a melodic context, which, with a few exceptions, differ by more than a semitone (see also Foxton *et al.*, 2004, for a similar finding). Thus, a more general musical pitch perception deficit is likely to be responsible for amusics' poor-pitch singing. This possibility is confirmed by the observation that the least proficient amusic singers were also the most severe amusics, as indicated by the MBEA. In sum, these findings are in keeping with the perceptual account of poor-pitch singing in congenital amusics but the origins must lie at a higher level than acoustical processing. One likely source of the difficulty experienced by amusics in musical pitch tasks is related to their difficulty in mapping pitch onto musical scales (Peretz, 2008).

However, there are notable exceptions. In a few amusics, perception and performance seem to dissociate. For example, PT, who suffers from a severe pitch perceptual defect, sang with lyrics as proficiently as controls. Conversely, TC, who had only mild problems on the pitch dimension in perception, was a very poor singer. In addition, we found support for dissociations between perception and performance at the level of the melodic contour. All amusics were impaired in perceiving contour changes in melodies, as assessed by the MBEA. However, five of them (AG, AS, EL, GC, and PT) were able to produce the correct contour when singing with lyrics. These findings are consistent with recent evidence of patients with impaired perception but spared production of pitch direction (Loui *et al.*, 2008). Together with previous evidence of poor-pitch singing in presence of unimpaired pitch perception (Bradshaw and McHenry, 2005; Dalla Bella *et al.*, 2007; Pfordresher and Brown, 2007; Wise and Sloboda, 2008) and of selectively impaired sung performance following brain damage without perceptual disorders (Schön *et al.*, 2003, 2004), these results point toward a double dissociation between pitch perception and production mechanisms (for a discussion, see Griffiths, 2008). These findings seem to question the more dominant view that perception and action share a common representational basis (e.g., Hommel *et al.*, 2001; Prinz, 2005, for a review). The latter model is supported by neurophysiologial studies showing that neurons in the prefrontal cortex of the macaque monkey (i.e., mirror neurons) respond both during action execution (e.g., picking a nut) and during action observation (see Rizzolatti, 2005, for a review). The dissociations reported here between perception and performance rather argue for independence. However, task differences (i.e., production of well-known songs from memory vs novel melody discrimination) may account for the dissociation. Further studies with highly comparable perception/production tasks (see Loui *et al.*, 2008, for an example) are in order to clarify the degree of independence between the two musical pitch systems.

The surprising finding that two amusic cases were able to sing in-tune despite severely impaired pitch perception deserves particular attention. This dissociation is reminiscent of action-blindsight in vision (e.g., Danckert and Rossetti, 2005, for a review) where the lack of awareness for visual stimuli does not preclude implicit treatment of information by the visual system (e.g., sufficient for spatial localization by pointing or saccading toward the stimuli). Similarly, amusics are generally impaired on tasks (e.g., pitch change detection) requiring explicit analysis of pitch differences. This pitch perception deficit is associated with brain anomalies within right front-temporal cortical regions (Hyde *et al.*, 2006, 2007). Yet, in the two aforementioned amusic cases, an implicit pitch-tracking mechanism may still be functional. Such mechanism would allow the analysis of fine-grained pitch differences without conscious awareness, thus providing sufficient feedback information for proficient singing. Note that this implicit pitch-tracking mechanism cannot be studied by simply asking amusics to judge their own singing proficiency. Amusics are notoriously unaware of how they sing. Yet, implicit pitch-tracking may be uncovered by recording brain responses to pitch differences that amusics are not aware of (such as quarter-tone pitch differences; see Peretz *et al.*, in press, for supporting evidence).

In summary, the present study indicates that components of the general ability to sing fractionate as a result of a developmental anomaly. For example, the ability to produce pitch intervals can be selectively disrupted without disrupting time, thus confirming what we previously observed in normal participants (Dalla Bella *et al.*, 2007). Hence the detailed study of singing provides a rich source of information not only on the multiple processing components involved in music cognition but also on how spared knowledge can drive behavior in a more natural setting than perceptual experiments.

ACKNOWLEDGMENTS

This research was supported by an International Reintegration Grant 14847 from the European Commission to SDB and by a Canada Institute of Health Research Grant and a Canada Research Chair to IP. We thank Peter Pfordresher, and an anonymous reviewer for insightful comments on an earlier version of the manuscript.

[1]Note that the discrimination of one-semitone differences in repeating tone sequences is impaired in amusic individuals as compared to matched controls, but not abolished. This poor but residual ability to discriminate semitones in an impoverished tone context may not support reliable pitch encoding in a rich musical context where pitch intervals mostly vary between zero and three semitones (Vos and Troost, 1989; Peretz and Hyde, 2003).

[2]For simplicity, the condition in which participants sang *Gens du pays* on a syllable will be referred to as "singing on /la/" regardless of the fact that some of the patients sang the melody either on /la/ or on /ta/.

[3]Due to the small samples, non-parametric correlation coefficients (i.e., Spearman's rho) were reported instead of standard Pearson *r* coefficients.

[4]Analyses considering items (i.e., intervals) instead of participants as the random factor yielded the same results.

[5]Because compression is impossible in the case of the unison, data for this interval were not considered in this analysis.

[6]The same ANOVA considering subjects instead of intervals as the random factor yielded a main effect of interval size ($F(4,80)=3.30$, $p<0.05$). Group × interval size, and group × interval size × deviation type interactions did not reach significance. This might be due to the small sample sizes in each group.

[7]As can be heard from the examples, PT and GC may appear less confident than controls in sustaining pitch, although they sing quite proficiently. This potential difference between amusics and controls was not captured by the aforementioned measures of pitch accuracy. However, further analyses indicated that amusics did not significantly differ from controls in sustaining pitch within a vowel group.

Alcock, K. J., Passingham, R. E., Watkins, K., and Vargha-Khadem, F. (**2000a**). "Pitch and timing abilities in inherited speech and language impairment," Brain Lang **75**, 34–46.

Alcock, K. J., Wade, D., Anslow, P., and Passingham, R. E. (**2000b**). "Pitch and timing abilities in adult left-hemisphere dysphasic and right-hemisphere damaged subjects," Brain Lang **75**, 47–65.

Ayotte, J., Peretz, I., and Hyde, K. (**2002**). "Congenital amusia: A group study of adults afflicted with a music-specific disorder," Brain **125**, 238–251.

Bergeson, T. R., and Trehub, S. E. (**2002**). "Absolute pitch and tempo in mothers' songs to infants," Psychol. Sci. **13**, 72–75.

Besson, M., Faita, F., Peretz, I., Bonnel, A.-M., and Requin, J. (**1998**). "Singing in the brain: Independence of lyrics and tunes," Psychol. Sci. **9**, 494–498.

Bradshaw, E., and McHenry, M. A. (**2005**). "Pitch discrimination and pitch matching abilities of adults who sing inaccurately," J. Voice **19**, 431–439.

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (**2007**). "Singing proficiency in the general population," J. Acoust. Soc. Am. **121**, 1182–1189.

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (**2009**). "Example of renditions from congenital amusics and controls," on http://www.mpblab.vizja.pl/publication.html (Last viewed 3/29/2009).

Danckert, J., and Rossetti, Y. (**2005**). "Blindsight in action: What can the different sub-types of blindsight tell us about the control of visually guided actions?," Neurosci. Biobehav Rev. **29**, 1035–1046.

Eerola, T., and North, A. C. (**2000**). "Expectancy-based model of melodic complexity," in *Proceedings of the Sixth International Conference on Music Perception and Cognition*, edited by C. Woods, G. B. Luck, R. Brochard, S. A. O'Neill, and J. A. Sloboda (Keele, Staffordshire, UK), [CD-ROM].

Eerola, T., and Toivianen, P. (**2004**). "MIDI toolbox: MATLAB tools for music research," available at http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/ (Last viewed 1/12/2009).

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (**2004**). "Characterization of deficits in pitch perception underlying 'tone deafness'," Brain **127**, 801–810.

Goodale, M. A., Milner, A. D., Jakobson, L. S., and Carey, D. P. (**1991**). "A neurological dissociation between perceiving objects and grasping them," Nature (London) **349**, 154–156.

Griffiths, T. D. (**2008**). "Sensory systems: Auditory action streams?," Curr. Biol. **18**, R387–R388.

Halpern, A. R. (**1989**). "Memory for the absolute pitch of familiar songs," Mem. Cognit. **17**, 572–581.

Hébert, S., Racette, A., Gagnon, L., and Peretz, I. (**2003**). "Revisiting the dissociation between singing and speaking in expressive aphasia," Brain **126**, 1838–1850.

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (**2001**). "The theory of event coding (TEC): A framework for perception and action planning," Behav. Brain Sci. **24**, 849–937.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A., and Peretz, I. (**2007**). "Cortical thickness in congenital amusia: When less is better than more," J. Neurosci. **27**, 13028–13032.

Hyde, K. L., and Peretz, I. (**2004**). "Brains that are out of tune but in time," Psychol. Sci. **15**, 356–360.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (**2006**). "Morphometry of the amusic brain: A two-site study," Brain **129**, 2562–2570.

Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (**2002**). "Temporal aspects of stimulus-driven attending in dynamic arrays," Psychol. Sci. **13**, 313–319.

Jones, M. R., and Pfordresher, P. Q. (**1997**). "Tracking melodic events using joint accent structure," Can. J. Exp. Psychol. **51**, 271–291.

Kinsella, G., Prior, M. R., and Murray, G. (**1988**). "Singing ability after right and left sided brain damage. A research note," Cortex **24**, 165–169.

Levitin, D. J. (**1994**). "Absolute memory for musical pitch: Evidence from

the production of learned melodies," Percept. Psychophys. **56**, 414–423.
Levitin, D. J., and Cook, P. R. (**1996**). "Memory for musical tempo: Additional evidence that auditory memory is absolute," Percept. Psychophys. **58**, 927–935.
Loui, P., Guenther, F., Mathys, C., and Schlaug, G. (**2008**). "Action-perception mismatch in tone-deafness," Curr. Biol. **18**, R331–R332.
Murayama, J., Kashiwagi, T., Kashiwagi, A., and Mimura, M. (**2004**). "Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia," Brain Cogn **56**, 36–42.
Peretz, I. (**1990**). "Processing of global and local musical information by unilateral brain-damaged patients," Brain **113**, 1185–1205.
Peretz, I. (**1996**). "Can we lose memory for music? The case of music agnosia in a non-musician," J. Cogn Neurosci. **8**, 481–496.
Peretz, I. (**2001**). "Brain specialization for music: New evidence from congenital amusia," Ann. N.Y. Acad. Sci. **930**, 189–192.
Peretz, I. (**2008**). "Musical disorders. From behavior to genes," Curr. Dir. Psychol. Sci. **17**, 329–333.
Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., and Jutras, B. (**2002**). "Congenital amusia: A disorder of fine-grained pitch discrimination," Neuron **33**, 185–191.
Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (**2009**). "The amusic brain: In tune, out of key, and unaware," Brain (in press).
Peretz, I., Champod, A. S., and Hyde, K. (**2003**). "Varieties of musical disorders: The Montreal battery of evaluation of amusia," Ann. N.Y. Acad. Sci. **999**, 58–75.
Peretz, I., Cummings, S., and Dubé, M.-P. (**2007**). "The genetics of congenital amusia (tone deafness): A family-aggregation study," Am. J. Hum. Genet. **81**, 582–588.
Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L., Gagnon, B., Trimmer, C., Paquette, S., and Bouchard, B. (**2008**). "On-line identification of congenital amusia," Music Percept. **25**, 331–343.
Peretz, I., and Hyde, K. (**2003**). "What is specific to music processing? Insights from congenital amusia," Trends Cogn. Sci. **7**, 362–367.
Peretz, I., Kolinski, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (**1994**). "Functional dissociations following bilateral lesions of auditory cortex," Brain **117**, 1283–1301.
Peretz, I., and Kolinsky, R. (**1993**). "Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective," Q. J. Exp. Psychol. **46A**, 301–325.
Pfordresher, P. Q., and Brown, S. (**2007**). "Poor-pitch singing in the absence of 'tone-deafness'," Music Percept. **25**, 95–115.
Prinz, W. (**2005**). "An ideomotor approach to imitation," in *Perspectives on Imitation: From Neuroscience to Social Science*, Mechanisms of Imitation and Imitation in Animals Vol. **I**, edited by S. Hurley and N. Chater (MIT Press, Cambridge, MA), pp. 141–156.
Prior, M., Kinsella, G., and Giese, J. (**1990**). "Assessment of musical processing in brain-damaged patients: Implications for laterality of music," J. Clin. Exp. Neuropsychol. **12**, 301–312.
Racette, A., Bard, C., and Peretz, I. (**2006**). "Making non-fluent aphasics speak: Sing along!," Brain **129**, 2571–2584.
Rizzolatti, G. (**2005**). "The mirror neuron system and imitation," in *Perspectives on Imitation: From Neuroscience to Social Science*, Mechanisms of Imitation and Imitation in Animals Vol. **I**, edited by S. Hurley and N. Chater (MIT Press, Cambridge, MA), pp. 55–76.
Samson, S., and Zatorre, R. (**1991**). "Recognition for text and melody of songs after unilateral temporal lobe lesions: Evidence for dual encoding," J. Exp. Psychol. Learn. Mem. Cogn. **17**, 793–804.
Schön, D., Lorber, B., Spacal, M., and Semenza, C. (**2003**). "Singing: A selective deficit in the retrieval of musical intervals," Ann. N.Y. Acad. Sci. **999**, 189–192.
Schön, D., Lorber, B., Spacal, M., and Semenza, C. (**2004**). "A selective deficit in the production of exact musical intervals following right-hemisphere damage," Cogn. Neuropsychol. **21**, 773–784.
Serafine, M. L., Crowder, R. G., and Repp, B. (**1984**). "Integration of melody and text in memory for song," Cognition **16**, 285–303.
Serafine, M. L., Davidson, J., Crowder, R. G., and Repp, B. (**1986**). "On the nature of melody-text integration in memory for songs," J. Mem. Lang. **25**, 123–135.
Sloboda, J. A., Wise, K. J., and Peretz, I. (**2005**). "Quantifying tone deafness in the general population," Ann. N.Y. Acad. Sci. **1060**, 255–261.
Sundberg, J., and Bauer-Huppmann, J. (**2007**). "When does a sung tone start?," J. Voice **21**, 285–293.
Terao, Y., Mizuno, T., Shindoh, M., Sakurai, Y., Ugawa, Y., Kobayashi, S., Nagai, C., Furubayashi, T., Arai, N., Okabe, S., Mochizuki, H., Hanajima, R., and Tsuji, S. (**2006**). "Vocal amusia in a professional tango singer due to a right superior temporal cortex infarction," Neuropsychologia **44**, 479–488.
Timmers, R., Ashley, R., Desain, P., and Heijink, H. (**2000**). "The influence of musical context on tempo rubato," J. New Music Res. **29**, 131–158.
Todd, N. (**1985**). "A model of expressive timing in tonal music," Music Percept. **3**, 33–58.
Vigneault, G., and Rochon, G. (**1976**). *Gens du pays* (People from the country, Score) (Editions du vent qui tourne, Montreal).
Vos, P. G., and Troost, J. M. (**1989**). "Ascending and descending melodic intervals: Statistical findings and their perceptual relevance," Music Percept. **6**, 383–396.
Wise, K. J., and Sloboda, J. A. (**2008**). "Establishing an empirical profile of self-defined 'tone-deafness': Perception, singing performance and self-assessment," Music. Sci. **12**, 3–23.

Temperature modes for nonlinear Gaussian beams

Matthew R. Myers and Joshua E. Soneson
Center for Devices and Radiological Health, U. S. Food and Drug Administration, Silver Spring, Maryland 20993

(Received 11 December 2008; revised 11 May 2009; accepted 12 May 2009)

In assessing the influence of nonlinear acoustic propagation on thermal bioeffects, approximate methods for quickly estimating the temperature rise as operational parameters are varied can be very useful. This paper provides a formula for the transient temperature rise associated with nonlinear propagation of Gaussian beams. The pressure amplitudes for the Gaussian modes can be obtained rapidly using a method previously published for simulating nonlinear propagation of Gaussian beams. The temperature-mode series shows that the nth temperature mode generated by nonlinear propagation, when normalized by the fundamental, is weaker than the nth heat-rate mode (also normalized by the fundamental in the heat-rate series) by a factor of $\log(n)/n$, where n is the mode number. Predictions of temperature rise and thermal dose were found to be in close agreement with full, finite-difference calculations of the pressure fields, temperature rise, and thermal dose. Applications to non-Gaussian beams were made by fitting the main lobe of the significant modes to Gaussian functions. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3148204]

PACS number(s): 43.80.Gx, 43.80.Sh, 43.35.Wa, 43.25.Zx [CCC] Pages: 425–433

I. INTRODUCTION

In determining the biological effects of high-intensity focused ultrasound (HIFU) procedures, mathematical models are often called upon to provide estimates of relevant physical quantities such as acoustic pressure and tissue temperature. The mathematical analysis can often be simplified when the beam intensity profile perpendicular to the propagation axis can be approximated by a Gaussian function. Saito and Tanaka (1990) and Soneson and Myers (2007) derived a set of nonlinear equations governing the Gaussian modes representing a propagating HIFU beam. These Guassian-mode models, requiring only the solution of ordinary differential equations for the modal amplitudes, provide rapid estimates of the acoustic pressure associated with HIFU beams. The Gaussian-mode models are particularly useful for performing parametric studies, as an input to the HIFU procedure is varied over a range of values.

In order to determine thermal effects of HIFU procedures, pressure modes characterizing nonlinear acoustic propagation can be converted to temperature modes by solving the bioheat equation. Wu and Du (1990) used a Green-function approach to sum over one-dimensional modes of a sawtooth wave. We employ a similar Green-function approach in this paper. By limiting interest to on-axis locations (where the temperature rise is maximum), and by arguing that the observation location is heated primarily by nearby sources for typical HIFU procedure times, the double integrals in the Green-function approach can be performed analytically. In the limit of linear propagation (only the first mode contributing), the temperature rise at the focus matches that derived by Bacon and Shaw (1993). Bacon and Shaw (1993) derived the temperature at the focal point of a Gaussian beam by modeling the ultrasound beam as a cylinder with no axial intensity variation. This paper may be seen as extending the Bacon and Shaw (1993) result to nonlinear propagation, as well as to axial locations away from the focus.

The results derived from the Gaussian-mode model are compared against finite-difference calculations performed by the software HIFU_SIMULATOR (Soneson, 2009). This software package solves the axisymmetric Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation (Zabolotskaya and Khokhlov, 1969; Kuznetsov, 1971) in layered media using a split-step algorithm in which the linear terms are computed in the frequency domain, while the nonlinear term is calculated in the time domain. It accommodates power-law absorption vs frequency curves and incorporates absorbing boundary conditions to eliminate artifacts. The code provides pressure and intensity distributions, and can use these results to determine temperature rise and lesion volume for a specified treatment protocol by solving the axisymmetric Pennes bioheat transfer equation (Pennes, 1948). It may also be obtained without charge by contacting the authors.

In Sec. II, the foundation of the Green-function approach is provided. The details of the computations, which are performed using the method of matched asymptotic expansions, are provided in the Appendix. Section III contains numerical computations of temperature rise and thermal dose, performed using both the Gaussian-mode approach and a finite-difference solution to the energy equation. Section IV discusses accuracy of the Gaussian-mode model and the features contained in the analytical solutions.

II. METHOD

We consider a spherically symmetric transducer of radius a and focal length d radiating into a medium having a power-law dependence on frequency

$$\alpha(f) = \alpha_0 (f/f_0)^{\eta}, \tag{1}$$

where α_0 is the absorption at some reference frequency f_0. In the simulation of nonlinear beam propagation, a dispersion relation accompanies Eq. (1). We employ a local approximation of the Kramers–Kronig relation, as described in Soneson and Myers, 2007.

In our analyses we take the axial coordinate z to coincide with the transducer axis, and the radial coordinate r to be perpendicular to z. The transducer face is positioned at $z=0$. The pressure is given by $p(r,z,t)$, with the peak pressure across the transducer face denoted by p_0.

A. Gaussian-mode formulation

A time-harmonic pressure of angular frequency ω generated by the transducer will result in the production of modes of frequency $n\omega(n=1,2,3,\ldots)$ due to nonlinear propagation, and a time-harmonic representation of the pressure field can be prescribed by

$$p = \frac{p_0}{2} \sum_{n=1}^{\infty} [u_n(\rho,\zeta) e^{in\omega t} + u_n^*(\rho,\zeta) e^{-in\omega t}], \tag{2}$$

where u_n is the dimensionless complex pressure, $\zeta=z/d$, and $\rho=r/a$.

The pressure across the transducer face is assumed to have a Gaussian profile

$$u_1(\rho,\zeta=0) = \exp[(iG-1)\rho^2], \tag{3}$$

where $G=\pi a^2 f/(dc_0)$ is the linear pressure gain, f is frequency, and c_0 is the infinitesimally small-amplitude sound speed. Amplitudes of higher-order modes are taken to be zero at the transducer face, i.e.,

$$u_n(\rho,\zeta=0) = 0, \quad n=2,3,\ldots. \tag{4}$$

Soneson and Myers (2007) showed that the pressure field at a general location (ρ,ζ) may be approximated by the following Gaussian representation:

$$u_n(\rho,\zeta) = a_n(\zeta)\exp[-Gnb(\zeta)\rho^2], \tag{5}$$

where

$$b(\zeta) = \frac{1-iG}{G-(G+i)\zeta}. \tag{6}$$

The functions $a_n(\zeta)$ satisfy a system of ordinary differential equations (Soneson and Myers, 2007). The intensity for the system of Gaussian modes is

$$I(\rho,\zeta) = \sum_{n=1}^{\infty} I_n = \sum_{n=1}^{\infty} I(0,0)|a_n(\zeta)|^2 \exp(-2Gn\,\mathrm{Re}[b(\zeta)]\rho^2), \tag{7}$$

where Re denotes the real part and $I(0,0)=p_0^2/2\rho_0 c_0$ is the intensity at the center of the transducer. The parameter ρ_0 is the mean density.

B. Temperature rise

Energy in the field (7) is absorbed according to Eq. (1), manifesting as a temperature rise in the medium. We neglect perfusion within the medium on the basis of the short sonication time, which we take to be on the order of tens of seconds. Nyborg (1981, 1988) and Nyborg and Wu (1994) showed that the temperature rise can be computed using the following Green-function solution to the bioheat equation:

$$T(x,y,z,t) = \int F_g q(x',y',z')dV', \tag{8}$$

where

$$F_g = \frac{1}{8\pi KR}\mathrm{erfc}\left(\frac{R}{\sqrt{4\kappa t}}\right), \tag{9}$$

K is the thermal conductivity, κ is the thermal diffusivity, t is time, and $R=((x-x')^2+(y-y')^2+(z-z')^2)^{1/2}$ is the distance from the source point (x',y',z') to the observation point (x,y,z). The function q represents the heat source (energy per unit volume per unit time). We assume that the heat source can be adequately described within the plane-wave approximation for the beam (Nyborg, 1988)

$$q = 2\alpha I,$$

in the case of linear propagation, or

$$q = \sum_n 2\alpha_n I_n, \tag{10}$$

in the nonlinear case. Here the absorption α_n at the frequency nf_0 is given by Eq. (1). In this paper we will consider temperatures only along the beam axis, which we take to be the z-axis. For the moment, we also restrict our attention to the focus $z=d$ or $\zeta=1$. Upon using Eqs. (10) and (7) in Eq. (8) and using dimensionless cylindrical coordinates, we obtain the following Green-function integral for the transient temperature at the focus:

$$T(\rho=0,\zeta=1) = \frac{\alpha_0 a^2 d p_0^2}{2K\rho_0 c_0}\int_0^{\infty} d\rho'\rho' \int_0^{\infty} d\zeta' \frac{\mathrm{erfc}\left(\frac{R(\rho',\zeta')}{\sqrt{4\kappa t}}\right)}{R(\rho',\zeta')} S(\rho',\zeta'), \tag{11}$$

where

$$R(\rho,\zeta) = \sqrt{a^2\rho^2 + d^2(1-\zeta)^2} \tag{12}$$

and S is the dimensionless modal sum for the heat source

$$S(\rho,\zeta) = \sum_{n=1}^{\infty} n^{\eta}|a_n(\zeta)|^2 \exp\left(\frac{-2G^2 n\rho^2}{\zeta^2 + (1-\zeta)^2 G^2}\right). \tag{13}$$

Upon changing the order of integration we obtain

$$T(0,1)=\frac{\alpha_0 a^2 dp_0^2}{2K\rho_0 c_0}\sum_{n=1}^{\infty} n^{\eta}\int_0^{\infty} d\rho'\rho'\int_0^{\infty}|a_n(\zeta')|^2 \times\frac{\operatorname{erfc}\left(\frac{R}{\sqrt{4\kappa t}}\right)}{R}\exp\left(\frac{-2G^2 n\rho^2}{\zeta^2+(1-\zeta)^2G^2}\right)d\zeta'. \quad (14)$$

The integrals in Eq. (14) may be performed using the method of matched asymptotic expansions (Myers, 2006a), under the assumptions that the dimensionless frequency ka $(=\omega a/c_0)$ is large, and the heat-conduction length is small relative to the focal length for the exposure times of interest $(d\gg(\kappa t)^{1/2})$. The details of the analysis are provided in the Appendix. The resulting temperature rise at the focus is

$$T(0,1)=\frac{\alpha_0 a^2 I(0,0)}{4K}\sum_{n=1}^{\infty} n^{\eta}\frac{|a_n(1)|^2}{nG^2}\ln\left(1+\frac{8G^2 n\kappa t}{a^2}\right). \quad (15)$$

For linear propagation, Eq. (15) can be compared with the result of Bacon and Shaw (1993). Bacon and Shaw (1993) modeled the energy absorbed from the beam as a heated cylinder (no axial intensity variation) in the focal region. The radius of the cylinder is the beam radius and the focal temperature rise can be written as

$$T_f=\frac{\alpha_0 r_b^2 I_f}{2K}\ln\left(1+\frac{4\kappa t}{r_b^2}\right), \quad (16)$$

where I_f is the focal intensity and r_b is the beam radius in the focal plane in the linear theory. In the linear case the intensity for a Gaussian beam may be found analytically (Wu and Du, 1990). In dimensional coordinates (r,z) it is

$$I(r,z)=I(0,0)A(z)\exp(-2\alpha_0 z)\exp(-2A(z)(r/a)^2), \quad (17)$$

where

$$A(z)=\frac{1}{(z/(dG))^2+(1-z/d)^2}. \quad (18)$$

In the focal plane $(z=d)$, these results simplify to $A(d)=G^2$ and

$$I(r,d)=I(0,0)G^2\exp(-2\alpha_0 z)\exp(-2G^2(r/a)^2). \quad (19)$$

If r_b is interpreted as the radial location at which the intensity is $1/e$ times the on-axis value, then Eq. (19) yields

$$r_b=a/(\sqrt{2}G), \quad (20)$$

and Eq. (16) becomes

$$T_f=\frac{\alpha_0 I(0,0)a^2 e^{-2\alpha_0 d}}{4K}\ln\left(1+\frac{8G^2\kappa t}{a^2}\right). \quad (21)$$

This result agrees with Eq. (15), when we assume a linear dependence of absorption on frequency ($\eta=1$ for tissue) and linear acoustic propagation [$a_n=0$ for $n>1$ and $|a_1|^2=\exp(-2\alpha_0 d)G^2$, consistent with Eq. (19)].

The heated cylinder model may be extended pre-focally or post-focally, using the radial intensity profile at the given axial location. Defining the effective radius $r_b(z)$ of the heated cylinder by the $1/e$ intensity value for the radial profile, we obtain from Eq. (17)

$$r_b(z)=a/(2A(z))^{1/2}. \quad (22)$$

Using this general beam radius in Eq. (15), we find the resulting temperature rise at location z to be

$$T(z)=\frac{\alpha_0 r_b^2(z)I(z)}{2K}\ln\left(1+\frac{4\kappa t}{r_b^2(z)}\right). \quad (23)$$

The temperature modes of Eq. (15) for the nonlinear case may be extended pre-focally and post-focally in a similar manner. The effective radius of the nth mode may be derived from Eqs. (7) and (6). Again defining the width of the heated cylinder by the $1/e$ intensity value of the radial profile, we find that in dimensionless variables,

$$\rho_n(\zeta)=r_n/a=1/(2Gn\ \mathrm{Re}[b])^{1/2}=\left(\frac{\phi(\zeta)}{2n}\right)^{1/2}, \quad (24)$$

where

$$\phi(\zeta)=(1-\zeta)^2+\zeta^2/G^2. \quad (25)$$

Using this effective radius in place of the radius a in Eq. (15), and evaluating the amplitude function at the general location ζ, gives the temperature rise as a function of axial position

$$T(0,\zeta)=\frac{\alpha_0 a^2 I(0,0)\phi(\zeta)}{4K}\sum_{n=1}^{\infty} n^{\eta}\frac{|a_n(\zeta)|^2}{n}\ln\left(1+\frac{8n\kappa t}{a^2\phi(\zeta)}\right). \quad (26)$$

The accuracy of this expression will be examined in Sec. III.

C. Thermal dose

The thermal dose associated with the Gaussian temperature modes can be computed using the formula for cumulative equivalent minutes at 43° (Sapareto and Dewey, 1984) as follows:

$$\mathrm{CEM}_{43}=\int_0^{t_f} 2^{T_C(t)-43}dt, \quad (27)$$

where time is measured in minutes, t_f is the duration of the procedure of interest, and $T_C(t)$ is the actual temperature (not temperature rise) measured in degree Celsius. We assume that the temperature rise $T(t)$ derived in Sec. II B is measured from 37 °C, so that the exponent in Eq. (27) becomes $T(t)-6$. Using Eq. (15) for $T(t)$ and switching to base e provides

$$\mathrm{CEM}_{43}=2^{-6}\int_0^{t_f} e^{\Sigma C_n \ln(1+A_n t)}dt, \quad (28)$$

where

$$C_n=\frac{\ln(2)\alpha_0 I_0 a^2 n^{\eta-1}a_n^2(1)}{4KG^2} \quad (29)$$

and

$$A_n=\frac{2(ka)^2 n\kappa}{d^2}. \quad (30)$$

The summation in the exponent leads to an integrand containing products

$$\mathrm{CEM}_{43} = 2^{-6}\int_0^{t_f} dt(1+A_1 t)^{C_1}(1+A_2 t)^{C_2},\ldots. \tag{31}$$

The product $A_n t$ may be rewritten as

$$\left(\frac{2\sqrt{\kappa t}}{r_n}\right)^2, \tag{32}$$

where r_n is the modal width defined in Eq. (24). The quantity $A_n t$ is the square of a ratio representing the distance diffused by heat over a time t divided by the width of the nth mode. For most times of interest (especially the larger times dominating the thermal-dose integral), this ratio is large at the focus, and the approximation $A_n t \gg 1$ may be made. After making this simplification and performing the integration in Eq. (31), we obtain

$$\mathrm{CEM}_{43} = 2^{-6}[A_1^{C_1}A_2^{C_2}\cdots]\frac{t_f^{1+C_1+C_2+\cdots}}{1+C_1+C_2+\cdots}. \tag{33}$$

We consider next the $\eta=1$ approximation for soft tissue. In that case, the sum of the coefficients C_n may be written as [see Eq. (29)]

$$C_1 + C_2 + \cdots = T_{\mathrm{ref},C}\bar{I}(0,1), \tag{34}$$

where

$$T_{\mathrm{ref},C} = \frac{\ln(2)\alpha_0 a^2 I(0,0)}{4KG^2} \tag{35}$$

is a reference temperature and the C subscript denotes that the temperature must be expressed in the Celsius scale [as required by Eq. (27)]. The quantity

$$\bar{I}(0,1) = \sum_{n=1}^{\infty} |a_n(1)|^2 \tag{36}$$

is the focal ($\rho=0$, $\zeta=1$) intensity normalized by $I(0,0)$. By similarly defining $\bar{I}_n$ to be the normalized intensity of the nth mode, we can write the thermal dose at the focus as

$$\mathrm{CEM}_{43} = \frac{2^{-6}t_f}{T_{\mathrm{ref},C}\bar{I}(0,1)}\prod\left(\frac{4\kappa t_f}{r_n^2}\right)^{T_{\mathrm{ref},C}\bar{I}_n(0,1)}. \tag{37}$$

At pre-focal or post-focal locations on the beam axis, repeating the above steps using Eq. (26) for the transient temperature at location $(0,\zeta)$ gives

$$\mathrm{CEM}_{43}(0,\zeta) = \frac{2^{-6}t_f}{T_{\mathrm{ref},C}\bar{I}(0,\zeta)}\prod\left(\frac{4\kappa t_f}{r_n^2(\zeta)}\right)^{T_{\mathrm{ref},C}\bar{I}_n(0,\zeta)}. \tag{38}$$

Here $\bar{I}_n(0,\zeta)$ is the normalized intensity for the nth mode at location $(0,\zeta)$, and $\bar{I}(0,\zeta)$ is the total normalized intensity at that location. The beam radius r_n for the nth mode is defined in Eq. (24). Like Eq. (37), Eq. (38) requires $A_n t \gg 1$, and is consequently valid only where $(8n\kappa t)/(a^2\phi(\zeta)) \gg 1$ [see Eqs. (24) and (32)].

D. Application to non-Gaussian transducers

While the present theory applies to ultrasound transducers with Gaussian shading, it is possible to extend the results to non-Gaussian transducers by fitting the relevant intensity profiles to Gaussian shapes. Bacon and Shaw (1993) noted that the main lobe of many profiles may be closely approximated with a Gaussian function. One scenario in which this approximation may be made is to suppose that the pressure trace is measured experimentally both on axis and at an off-axis radius r_*. In practice, r_* could be on the order of the half-width of the beam, about 1 mm. A temporal Fourier transform at each location will yield a set of modal intensities $I_n(0,z)$ and $I_n(r_*,z)$. If each intensity mode is fitted to a function of the form

$$I_n(r,z) = I_n(0,z)\exp(-r/r_{n,\mathrm{eff}})^2, \tag{39}$$

where $r_{n,\mathrm{eff}}$ is the effective radius of the nth mode, then we obtain

$$r_{n,\mathrm{eff}} = r_*/[\ln(I_n(0,\zeta)/I_n(r_*,\zeta))]^{1/2}. \tag{40}$$

Using this effective radius in Eq. (16) and summing the temperature modes gives the equation for the focal temperature rise for a non-Gaussian transducer

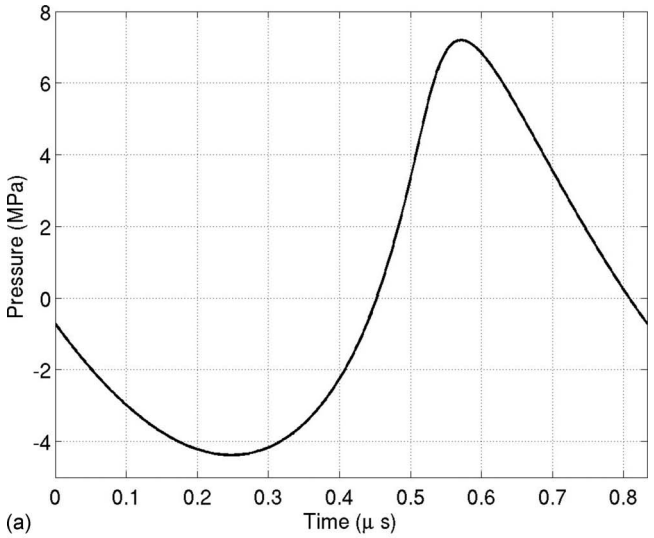
$$T(0,d) = \frac{\alpha_0}{2K}\sum_{n=1}^{\infty} n^{\eta} I_n(0,z) r_{n,\mathrm{eff}}^2 \ln\left(1+\frac{4\kappa t}{r_{n,\mathrm{eff}}^2}\right). \tag{41}$$

III. RESULTS

The accuracy of the analytic expressions of Sec. II was evaluated through a series of numerical computations. To obtain the Gaussian modal amplitudes a_n, the ordinary differential equations contained in Soneson and Myers, 2007 were solved. We refer the interested reader to Soneson and Myers, 2007 for details of the numerical approach. Once the a_n values were determined, the temperature and thermal dose were obtained analytically using Eqs. (26) and (37).

The temperature-mode model was compared against finite-difference solutions to the energy equation. The finite-difference formulation of the energy equation was implemented in the software HIFU_SIMULATOR (Soneson, 2009). The pressure field used to derive the intensity contained in the finite-difference energy equation was obtained from the Gaussian-mode approach when the transducer was Gaussian, and a finite-difference solution to the KZK equation (HIFU_SIMULATOR, Soneson, 2009) when the transducer was uniform.

A transducer of radius $a=4$ cm was used in the calculations. For the Gaussian transducer, $a=4$ cm is the radial location at which the pressure across the transducer face drops off to $1/e$ of its centerline value. The focal length d was 10 cm, and the operating frequency was 1.2 MHz. The linear pressure gain of the transducer was $G=40$. The thermal conductivity and diffusivity of the tissue were taken to be $K=0.006$ W/(cm K) and $\kappa=0.001\,44$ cm^2/s, and the density was $\rho_0=1.0$ g/cm^3. The speed of sound in the tissue was $c_0=1.5\times10^5$ cm/s, and the absorption α_0 at 1 MHz was 0.1 Np/cm. The nonlinearity parameter $\beta=1+B/2A$ was 4.5. The exposure time was 10 s unless otherwise noted. Transducer powers ranging from 12 to 300 W were considered.



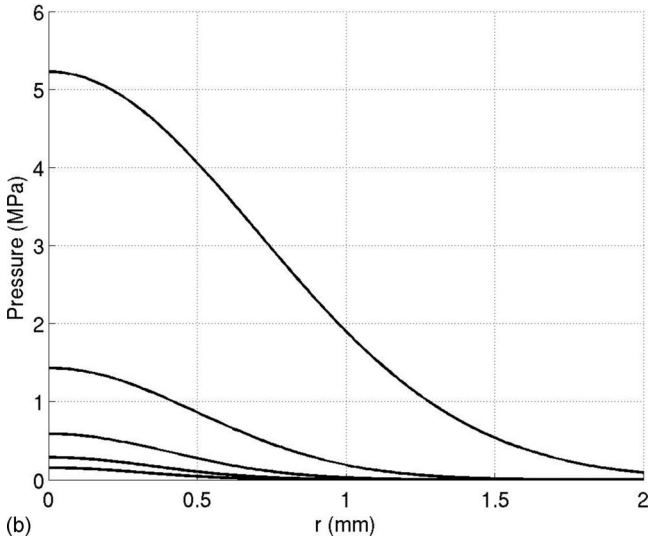


FIG. 1. (a) Pressure waveform at the focus of the Gaussian transducer considered in the computations. Power is 330 W. (b) Pressure distribution as a function of radial position for the harmonics comprising the waveform of (a).

Figure 1(a) shows the focal waveform [computed from Eq. (2)] for the case where the acoustic power is 165 W and the nonlinearity parameter $N=2\pi p_0\beta df/(\rho_0 c_0^3)$ was 0.45. The amplitudes of the harmonics in the focal plane [Eq. (5)] are shown in Fig. 1(b). The temperature modes arising from the absorption of the Gaussian pressure modes were computed using Eq. (26). The total (sum of all modes) temperature is plotted as a function of axial position in Fig. 2. Also plotted is the temperature computed from a finite-difference solution to the heat equation using the Gaussian intensity formulation (7) in the source term. The two curves are essentially overlapping.

The thermal dose at the focus for the conditions of Figs. 1 and 2 was computed using Eq. (37), as well as a numerical integration of Eq. (27) with the transient temperature obtained from a finite-difference calculation. The cumulative equivalent minutes are plotted for both approaches in Fig. 3, for acoustic powers of 12.6, 25.1, and 50.3 W. The analytical

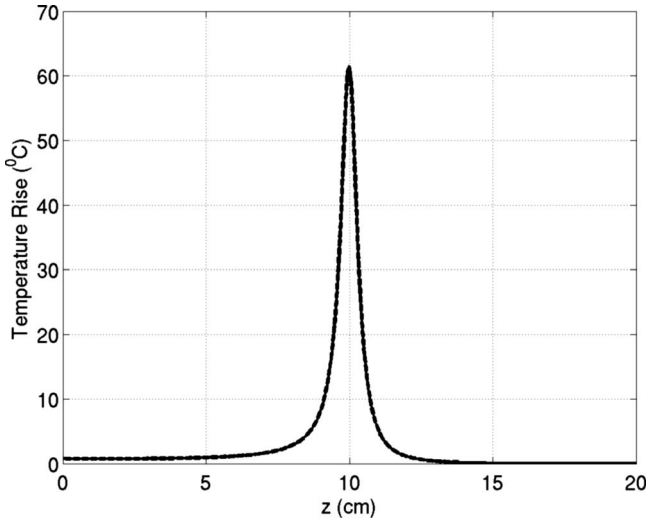


FIG. 2. Temperature as a function of position generated by the Gaussian temperature-mode approach (solid line) and a finite-difference solution to the energy equation (dashed line). For both approaches, intensity is computed using the Gaussian pressure-mode series. Exposure time is 10 s. Pressure wave at the focus is shown in Fig. 1(a).

and numerical curves are close except for very small times, where the condition $A_n t \gg 1$ required by Eq. (33) is violated.

To examine the applicability of the Gaussian temperature-mode approach to non-Gaussian transducers, the temperature rise from a transducer with a uniform pressure distribution across the face was computed. The properties of the transducer were those of the Gaussian transducer described at the beginning of this section: radius 4 cm, focal length 10 cm, and frequency 1.2 MHz. The power was 165 W. To compute the temperature rise for this transducer, the pressure field was first obtained using a finite-difference solution to the KZK equation (Soneson, 2009). The approach employed in the HIFU_SIMULATOR software automatically yielded frequency-domain information, including pressure amplitudes for the harmonics of the fundamental frequency. The

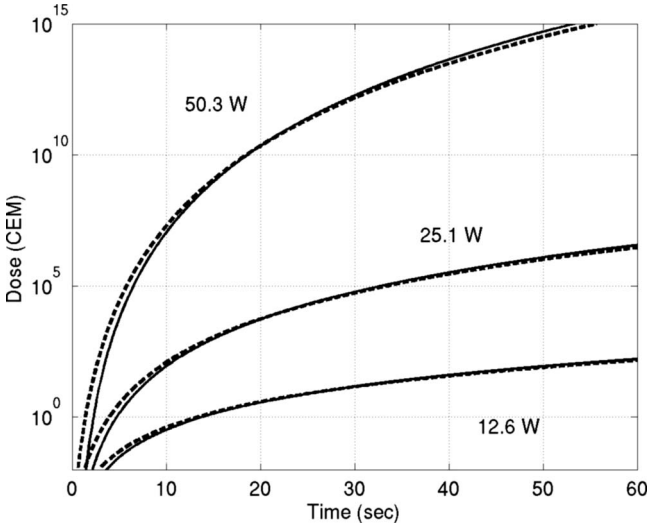


FIG. 3. Thermal dose computed for three different transducer powers. Cumulative equivalent minutes were computed from the analytical formula (37), as well as the thermal-dose integral utilizing the temperature trace obtained from finite-difference calculations.

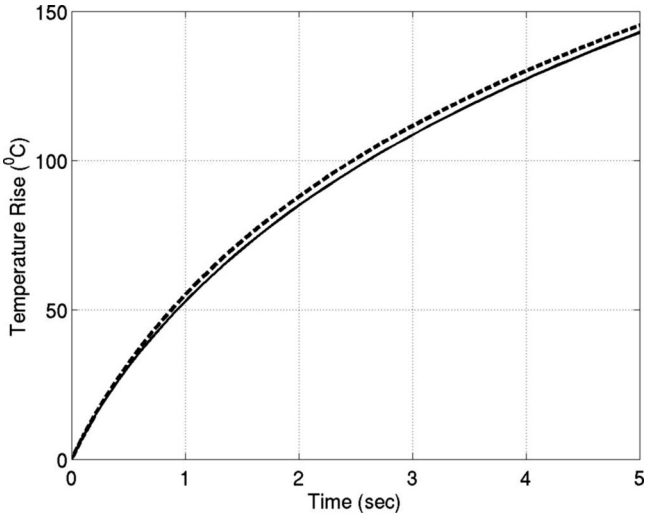


FIG. 4. Temperature rise for a transducer with uniform pressure distribution across the transducer face. Pressure field computed using finite-difference solution to KZK equation. Solid line—temperature-mode approach, with main lobe of finite-difference intensity profile fitted to a Gaussian function. Dashed line—finite-difference solution to energy equation.

intensities required in Eqs. (39) and (40) were derived by squaring the pressure amplitudes and dividing by $2\rho_0 c_0$. The axial location was the focal plane, and we chose the off-axis radial location r_* to be 0.7 mm. (This selection is discussed in Sec. IV.) The intensities $I_n(0,d)$ and $I_n(r_*,d)$ represented what would be obtained from a Fourier transform of experimental data measured at the chosen locations. The temperature distribution from the Gaussian temperature-mode approach could then be calculated from Eq. (41). A finite-difference solution to the energy equation, using the KZK-derived pressure field in the source term, was also computed. A comparison of the temperature fields derived from the two approaches is shown in Fig. 4. The Gaussian-mode approach underpredicts the finite-difference temperature throughout the procedure, with a maximum difference of about 5%.

IV. DISCUSSION

The range of validity of the Gaussian pressure-mode formulation was examined in Soneson and Myers, 2007. There the pressure waveforms based on the Gaussian-mode approach were compared with finite-difference solutions to the KZK equation for a Gaussian-shaded transducer. For a representative HIFU transducer (4 cm radius, 10 cm focal length, 1.18 MHz frequency, and linear pressure gain of 40), it was found that peak positive pressures at the focus as determined using the Gaussian-mode approach were within 10% of the finite-difference values in water for transducer powers up to 53 W. The corresponding N value for the 53-W power level was 0.2. In tissue, which is more attenuating, the agreement was within 10% for powers up to 156 W ($N=0.4$). The waveforms were distorted to the point where the peak positive pressure was roughly 2.5 times as large as the peak rarefactional pressure. As we discuss at the end of this section, it is possible that the range of validity of the temperature series may be broader than that of the pressure series, both in terms of power level and applicability to non-Gaussian transducers.

Expression (26) for the modal temperatures agreed well with finite-difference solutions to the energy equation at all axial locations (Fig. 2). As explained below Eq. (38), Eq. (38) is an accurate expression for the thermal dose (Fig. 3) only near the focus. Accurate calculations of thermal dose away from the focus can be obtained by inserting Eq. (26) into Eq. (27) and integrating numerically. The integrand [visible in Eq. (31)] is smooth and the integration can be performed via standard numerical methods.

The modal representation is useful for identifying the relative importance of nonlinear propagation on different quantities of interest, such as intensity, heat-rate, temperature, and thermal dose. From Eqs. (10) and (7), the heat-rate at the focus is

$$Q = \alpha_0 I(0,0) \sum_{n=1}^{\infty} n^{\eta} |a_n(1)|^2, \tag{42}$$

and hence the heat-rate associated with the nth mode is

$$Q_n = \alpha_0 I(0,0) n^{\eta} |a_n(1)|^2. \tag{43}$$

From Eq. (15), we see that

$$\frac{T_n}{T_1} = \frac{Q_n}{Q_1}\frac{1}{n}\left[\frac{\ln\left(1+\frac{8nG^2\kappa t}{a^2}\right)}{\ln\left(1+\frac{8G^2\kappa t}{a^2}\right)}\right]. \tag{44}$$

Ignoring the slow variation of the logarithmic terms, this result states that higher-order modes contribute less to temperature rise than to heat generation by a factor of $1/n$. Thus, while propagation may be highly nonlinear (many modes contributing) with respect to heat generation, it can be much less so (substantially fewer modes contributing) with respect to temperature rise. The stronger manifestation of nonlinearity in heat-rate compared with temperature is consistent with the results of Curra *et al.* (2000). If we form the ratio of the nth intensity mode to the fundamental, analogous to Eq. (44), we find from Eqs. (7) and (26) (with $\eta=1$) that both I_n/I_1 and T_n/T_1 have an n-dependence of $|a_n|^2$, if we ignore the logarithmic terms. In other words, nonlinear-propagation effects on intensity are comparable to those on temperature. Harmonics for both are weaker relative to heat-rate by a factor of $1/n$ (when $\eta=1$). For thermal dose, the importance of higher harmonics can be assessed by considering how close the nth factor in the infinite product of Eq. (38) is to 1. The nth factor is

$$\exp\left[T_{\mathrm{ref},C}\bar{I}_n(0,\zeta)\ln\left(\frac{4\kappa t_f}{r_n^2(\zeta)}\right)\right]. \tag{45}$$

As n gets larger the exponent grows smaller and the term may be approximated by

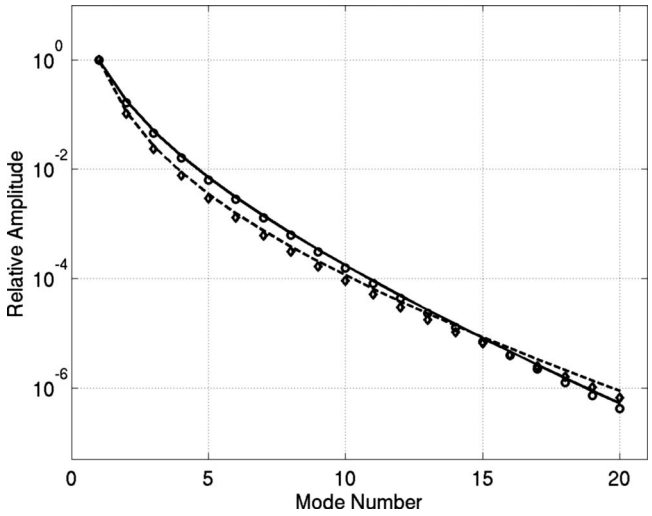


FIG. 5. Magnitude of the nth mode (normalized by magnitude of the first mode) as a function of mode number, for both heat-rate Q and temperature T. Solid line: Heat-rate, based on finite-difference solution to KZK equation for the pressure field of a uniform transducer. Dashed line: heat-rate, based on Gaussian-mode representation of pressure field for Gaussian transducer that best approximates the uniform transducer. Circles: temperature, based on finite-difference computations of both KZK and energy equations. Diamonds: temperature, based on Gaussian-mode approach for both pressure and temperature.

$$1 + T_{\mathrm{ref},C}\bar{I}_n(0,\zeta)\ln\left(\frac{4\kappa t_f}{r_n^2(\zeta)}\right). \tag{46}$$

Again ignoring the slow dependence of the logarithm on n, we see that the importance of the nth term decays with the n-dependence of $|a_n|^2(\bar{I}_n=|a_n|^2)$, in the same manner as the intensity and temperature.

Since the preceding results regarding mode-number dependence were derived assuming a Gaussian intensity profile, the question arises about the decay with harmonic number for more general intensity profiles. To identify the dependence of heat-rate and temperature on n for a uniform transducer, the modal intensity amplitudes I_n were taken from the finite-difference analysis leading to Fig. 4, and used [Eq. (10)] to compute the quantity $(1/n)Q_n/Q_1$. Additionally, for each mode of the uniform transducer, a finite-difference solution of the heat equation was performed to determine the temperature field associated with that mode. The ratio T_n/T_1 at the focus was computed for each mode number. For comparison with the results for the uniform transducer, the heat-rate and temperature vs n were computed for an "optimized" Gaussian transducer having amplitude and width chosen so that the focal pressure best matched that of the uniform transducer. The intensity at the center of the transducer was approximately double that for the uniform transducer, and the $1/e$ width was approximately half that of the uniform transducer. Figure 5 contains the ratios $(1/n)Q_n/Q_1$ and T_n/T_1. As expected from Eq. (44), the plots of $(1/n)Q_n/Q_1$ and T_n/T_1 essentially overlap for the Gaussian transducer. For the uniform transducer as well, the nth temperature mode is very nearly $1/n$ weaker relative to the fundamental than the nth heat-rate mode relative to the fundamental. Because the Gaussian parameters were chosen "optimally" (pressure match at focus), the two modal-dependence curves in Fig. 5 closely match one another. For other choices of parameters for either the Gaussian or uniform transducer the curves would diverge, but for both transducers the $(1/n)Q_n/Q_1$ and T_n/T_1 values would remain nearly equal.

The conclusions regarding the dependence on mode number n require that the time t of interest not be much smaller than the diffusion time, i.e., conduction of heat must be significant. Mathematically, for small enough time, the logarithm in Eq. (26) may be approximated by $8n\kappa t/(a^2\phi(\zeta))$, and the temperature and heat-rate acquire the same dependence on n. The diffusion time can be approximated by $r_n^2/4\kappa$ where r_n is the beam radius for the nth mode. The diffusion time is on the order of 1 s for the fundamental, and smaller for higher harmonics. Hence the conclusions regarding mode-number dependence require that the time on interest be greater than a few tenths of 1 s.

The high temperatures displayed in Fig. 4, while beyond the boiling point and hence unrealistic, demonstrate that the error associated with the Gaussian-mode approach is bounded over a large temperature range. (Relative error actually decreases for large times/temperatures.) Figure 4 demonstrates that fitting the main lobe of the intensity profiles to a Gaussian form is a feasible approach for treating non-Gaussian beams, when intensity information in tissue is available. The case of intensity measurement in water is discussed below. We remark that all of the modes are fitted to Gaussian functions by Eqs. (39) and (40). Using a choice of r_* that is too large will reduce the accuracy of the approach, as the intensity of higher modes at r_* will be so small as to make the calculation in Eq. (40) unstable. However, making r_* too small also results in instability, as $I_n(0,z)$ and $I_n(r_*,z)$ are difficult to distinguish [$\log(I_n(0,z)/I_n(r_*,z))$ in the denominator of Eq. (40) approaches zero]. The fitting approach will also degrade in accuracy for times sufficiently large that energy from side lobes becomes important. This time is on the order of r_l^2/κ, where r_l is the radial location of the side lobe and κ is the thermal diffusivity. For a side-lobe location of around 2 mm, this time would be on the order of 30 s. This diffusion time decreases with increasing mode number, as the beam width decreases with mode number [see Fig. 1(b) for the Gaussian case]. However, the higher modes contain less energy and contribute less to the temperature field. For highly nonlinear propagation (many modes contributing), it may be advisable to use smaller values of the off-axis measurement location r_*, in order to better resolve higher-order modes. The value $r_*=0.7$ mm used in our fitting procedure seemed large enough to provide sufficient discrimination between on-axis intensities and off-axis intensities, and small enough to resolve energy-containing modes and avoid side-lobe contribution. However, further optimization [involving finding a range of r_* values where r_{eff} in Eq. (40) is insensitive to the value of r_*] would be worthwhile, including accounting for different transducer and media properties.

The procedure just discussed requires intensity (or pressure) measurements or estimates in the desired tissue medium. Often pressure measurements are made in water, and a method for applying the measurements to the prediction of modal temperatures in tissue is desirable. One possibly useful method is the following.

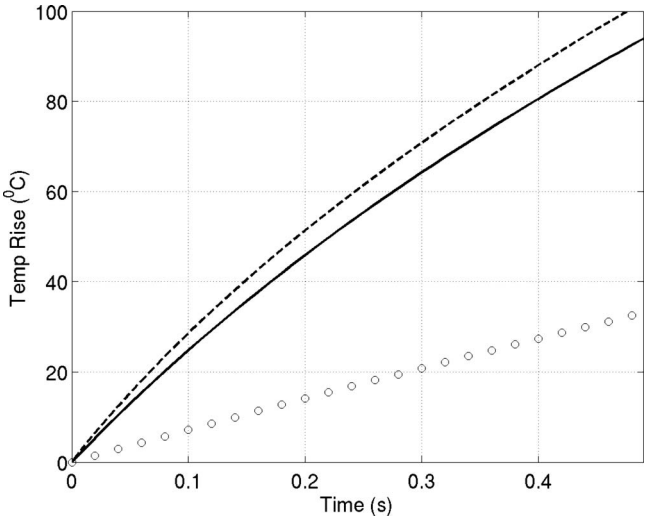


FIG. 6. Focal temperature as a function of sonication time for a 1 MHz uniform transducer, having diameter of 4 cm and focal length of 6 cm. Acoustic power is 300 W and the tissue attenuation is 0.5 dB/cm. Solid line: finite-difference simulations of acoustic propagation and energy equation. Dashed line: result from temperature-mode series using modal intensities computed in water and derated for tissue. Circles: finite-difference solution to heat equation with intensity derived from linear acoustics.

We suppose that the modal intensities in water are available at $(z=d,\ r=0)$ and $(z=d,\ r=r_*)$, e.g., through a temporal Fourier transform of the experimental pressure traces at those locations. We add all of the modal intensities at the focus to form the total focal intensity $I_w(0,d)$ in water. We then calculate a theoretical tissue total intensity $I_t(0,d)$ using a KZK equation or other nonlinear-propagation algorithm. This value is then used to derate the intensities in water, by multiplying each modal intensity in water by $I_t(0,d)/I_w(0,d)$. The effectiveness of this derating method was examined by applying it to a numerical example designed to illustrate applicability to highly nonlinear propagation. The transducer frequency was 1 MHz and its diameter was 4 cm. The focal length was 6 cm and the transducer radiated with 300 W of power. The pressure distribution across the transducer face was uniform. The modal intensities in water were computed using a finite-difference solution to the KZK equation (via HIFU_SIMULATOR). The derating scheme was applied and the transient temperature in a typical tissue medium of attenuation 0.50 dB/cm was computed using the fitting procedure ($r_*=0.7$ mm) and Eqs. (40) and (41). Figure 6 shows the temperature rise in tissue computed using this fitting and derating algorithm. The temperature trace is compared against a full (both propagation and energy equations) finite-difference simulation of the temperature rise. The temperature rise predicted by linear theory is provided as well for comparison. It is evident from viewing the linear and nonlinear traces that at this power level, substantial heat is produced through absorption of higher-order modes generated by nonlinear propagation. The derating/fitting scheme employed predicts the finite-difference result within about 10% accuracy.

The derating scheme just described is very simple, in that each mode in water is derated by the same factor based on the focal intensity derived from a KZK simulation in tissue. Other approaches incorporating mode-number dependent derating factors into Eq. (41) could very well produce more accurate results. The appropriate derating factor to use as a function of the operational parameters (gain, nonlinearity, and attenuation) is an important topic for further study. However, Fig. 6 demonstrates the potential of the temperature-mode series for making useful temperature predictions in tissue from measurements in water, even in high-power regimes beyond the range of validity of the Gaussian pressure-mode series.

V. CONCLUSION

This paper provides a simple yet accurate formula [Eq. (26)] for computing temperature rise in the case of nonlinear beam propagation, when the beams may be adequately modeled by a Gaussian function. Examples include Gaussian-shaded transducers, or uniform transducers with the main lobe modeled by a Gaussian profile. The pressure modal amplitudes required in the temperature-mode series can be found using a method previously reported (Soneson and Myers, 2007) for rapidly simulating nonlinear propagation of Gaussian beams. The analytic form of the temperature-mode series shows that the importance of higher-order modes for temperature is weaker than for heat-rate, by a factor of $\log(n)/n$, or roughly $1/n$, where n denotes the nth harmonic generated by nonlinear propagation. The temperature-mode approach yielded accurate estimates of temperature rise over the entire axial range of the beam, under conditions representing a typical HIFU procedure. An analytical formula for thermal dose, also containing the modal pressure amplitudes, was also presented [Eq. (38)]. This expression is accurate only near the focus for typical HIFU procedures.

APPENDIX

To evaluate the integral in Eq. (14), we introduce the local variables

$$\tilde{\zeta}=\zeta'-1,\quad \tilde{\rho}=\rho'\frac{\sqrt{2}G}{\sqrt{\zeta'^2+(1-\zeta')^2G^2}}. \tag{A1}$$

In terms of the local variables, the temperature rise (14) at the focus is

$$T=\frac{\alpha_0 I_0 a^2}{K}\sum n^{\eta}\int_0^{\infty}d\tilde{\rho}\tilde{\rho}e^{-n\tilde{\rho}^2}\int_{-1}^{\infty}d\tilde{\zeta}\,\mathrm{erfc}\left(\frac{\tilde{R}d}{\sqrt{4\kappa t}}\right)\frac{f(\tilde{\zeta})}{\tilde{R}}, \tag{A2}$$

where

$$\tilde{R}^2=\epsilon^2\delta^2\tilde{\rho}^2((1+\tilde{\zeta})^2+G^2\tilde{\zeta}^2)+\tilde{\zeta}^2,$$

$$\epsilon=1/G,$$

$$\delta=a/\sqrt{2}d,$$

and

$$f(\tilde{\zeta}) = \frac{|a_n(1+\tilde{\zeta})|^2(\tilde{\zeta}^2G^2 + (1+\tilde{\zeta})^2)}{2G^2}.$$

We take $\epsilon\delta = \sqrt{2}/ka \ll 1$. Under this high-frequency assumption, we can use the method of matched asymptotic expansions to estimate the integrals, as described in Myers, 2006a and Myers, 2006b. An asymptotic representation for the $\tilde{\zeta}$ integrand is

$$\frac{f(\tilde{\zeta})}{|\tilde{\zeta}|}\mathrm{erfc}\left(\frac{|\tilde{\zeta}|d}{\sqrt{4\kappa t}}\right) + \frac{f(0)}{\sqrt{\delta^2\epsilon^2\tilde{\rho}^2 + \tilde{\zeta}^2}} - \frac{f(0)}{|\tilde{\zeta}|}. \tag{A3}$$

The second and third terms in the integrand can be readily integrated, with a result that has a simple logarithmic dependence on $\tilde{\rho}$. The first term in Eq. (A3) has no $\tilde{\rho}$ dependence. Following the steps in Myers, 2006a to compute the $\tilde{\rho}$ integral, we obtain the following expression for the focal temperature:

$$T = \frac{\alpha_0 I_0 a^2}{K}\sum n^{\eta}\frac{|a_n(1)|^2}{4nG^2}[S_1 + \ln(2k^2a^2n) + C_e], \tag{A4}$$

where C_e is Euler's constant (≈ 0.577) and

$$S_1 = \ln M + \int_0^M d\tilde{\zeta}\frac{\mathrm{erfc}\left(\frac{\tilde{\zeta}d}{\sqrt{4\kappa t}}\right)^{\frac{|a_n(1+\tilde{\zeta})|^2(\tilde{\zeta}^2G^2+(1+\tilde{\zeta})^2)}{|a_n(1)|^2}} - 1}{|\tilde{\zeta}|}. \tag{A5}$$

The parameter M is a number large enough to adequately represent an infinite domain; in practice a value of 3 or 4 typically suffices (due to the rapid decay of the complementary error function with axial distance.)

The rapid decay of the complementary error function as the argument increases from zero is evidence of the dominance of localized heating (near $\tilde{\zeta}=0$) in HIFU procedures. Except for very large times, we can take advantage of the rapid decay of the complementary error function as $\tilde{\zeta}$ increases from zero and set $\tilde{\zeta}$ equal to zero in the terms multiplying $\mathrm{erfc}(\tilde{\zeta}d/\sqrt{4\kappa t})$. Upon making this approximation, integrating once by parts, and using

$$\int_0^{d/\sqrt{4\kappa t}} \ln(s\sqrt{4\kappa t}/d)e^{-s^2}ds \approx \frac{-\sqrt{\pi}}{4}(C_e + \ln 4) - \frac{\sqrt{\pi}\ln(d/\sqrt{4\kappa t})}{2} \tag{A6}$$

for $d/\sqrt{4\kappa t} \gg 1$, we obtain the following expression for the temperature rise at the focus:

$$T(0,1) = \frac{\alpha_0 a^2 I(0,0)}{4K}\sum_{n=1}^{\infty} n^{\eta}\frac{|a_n(1)|^2}{nG^2}\ln\left(\frac{8G^2n\kappa t}{a^2}\right). \tag{A7}$$

Here we have rewritten the high-frequency parameter ka in terms of the gain $G = ka^2/(2d)$. Equation (A7) does not apply for small times, since the approximation used in Eq. (A3),

$$R/\sqrt{4\kappa t} \approx \tilde{\zeta}d/\sqrt{4\kappa t} \tag{A8}$$

(source regions contributing to the temperature rise are primarily on axis, owing to the narrowness of the beam), does not hold when t is small. However, for small times, a direct integration of the energy equation with zero conduction shows that the temperature rise is directly proportional to the source term and time:

$$T \approx \frac{2\alpha_0 I(0,0)}{\rho_0 c_p}\left(\sum_{n=1}^{\infty} n^{\eta}|a_n(1)|^2\right)t \tag{A9}$$

This result can be recovered from Eq. (A7) in the limit as $t \to 0$ if a 1 is added to the argument of the logarithm in Eq. (A7). This modification does not affect the temperature-rise predictions of Eq. (A7) for times typical of HIFU procedures (tens of seconds), and hence a more uniformly valid expression for the temperature rise is

$$T(0,1) = \frac{\alpha_0 a^2 I(0,0)}{4K}\sum_{n=1}^{\infty} n^{\eta}\frac{|a_n(1)|^2}{nG^2}\ln\left(1 + \frac{8G^2n\kappa t}{a^2}\right). \tag{A10}$$

Bacon, D. R., and Shaw, A. (**1993**). "Experimental validation of predicted temperature rises in tissue-mimicking materials," Phys. Med. Biol. **38**, 1647–1659.

Curra, F. P., Mourad, P. D., Khokhlova, V. A., Cleveland, R. O., and Crum, L. A. (**2000**). "Numerical simulations of heating patterns and tissue temperature response due to high-intensity focused ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1077–1089.

Kuznetsov, V. P. (**1971**). "Equations of nonlinear acoustics," Sov. Phys. Acoust. **16**, 467–470.

Myers, M. R. (**2006a**). "Tissue deformation induced by radiation force from Gaussian transducers," J. Acoust. Soc. Am. **119**, 3147–3152.

Myers, M. R. (**2006b**). "Long-time temperature rise due to absorption of focused Gaussian beams in tissue," J. Acoust. Soc. Am. **120**, 4064–4070.

Nyborg, W. L. (**1981**). "Heat generation within a relaxing medium," J. Acoust. Soc. Am. **70**, 310–312.

Nyborg, W. L. (**1988**). "Solutions of the bio-heat transfer equation," Phys. Med. Biol. **33**, 785–792.

Nyborg, W. L., and Wu, J. (**1994**). "Solution of the linear bio-heat transfer equation," Phys. Med. Biol. **39**, 924–925.

Pennes, H. H. (**1948**). "Analysis of tissue and arterial blood temperature in the resting human forearm," J. Appl. Physiol. **1**, 93–122.

Saito, S., and Tanaka, H. (**1990**). "Harmonic components of finite-amplitude sound in a focused Gaussian beam," J. Acoust. Soc. Jpn. (E) **11**, 225–233.

Sapareto, S. A., and Dewey, W. C. (**1984**). "Thermal dose determination in cancer therapy," Int. J. Radiat. Oncol., Biol., Phys. **10**, 787–800.

Soneson, J. E. (**2009**). "A user-friendly software package for HIFU simulation," Proceedings of the Eighth International Symposium on Therapeutic Ultrasound, Minneapolis, MN, September.

Soneson, J. E., and Myers, M. R. (**2007**). "Gaussian representation of high-intensity focused ultrasound beams," J. Acoust. Soc. Am. **122**, 2526–2531.

Wu, J., and Du, G. (**1990**). "Temperature elevation generated by a focused Gaussian beam of ultrasound," Ultrasound Med. Biol. **16**, 489–498.

Zabolotskaya, E. A., and Khokhlov, R. V. (**1969**). "Quasi-plane waves in the nonlinear acoustics of confined beams," Sov. Phys. Acoust. **15**, 35–40.

Voice of the turtle: The underwater acoustic repertoire of the long-necked freshwater turtle, *Chelodina oblonga*

Jacqueline C. Giles[a] and Jenny A. Davis
School of Environmental Science, Murdoch University, South Street, Murdoch, Perth 6150, Western Australia

Robert D. McCauley
Centre for Marine Science and Technology, Curtin University, Kent Street, Bentley, Perth 6102, Western Australia

Gerald Kuchling
School of Animal Biology, The University of Western Australia, 35 Stirling Highway, Nedlands, Perth 6009, Western Australia

(Received 6 August 2008; revised 8 February 2009; accepted 12 May 2009)

Chelodina oblonga is a long-necked, freshwater turtle found predominantly in the wetlands on the Swan Coastal Plain of Western Australia. Turtles from three populations were recorded in artificial environments set up to simulate small wetlands. Recordings were undertaken from dawn to midnight. A vocal repertoire of 17 categories was described for these animals with calls consisting of both complex and percussive spectral structures. Vocalizations included clacks, clicks, squawks, hoots, short chirps, high short chirps, medium chirps, long chirps, high calls, cries or wails, hooos, grunts, growls, blow bursts, staccatos, a wild howl, and drum rolling. Also, a sustained vocalization was recorded during the breeding months, consisting of pulse sequences that finished rhythmically. This was hypothesized to function as an acoustic advertisement display. *Chelodina oblonga* often lives in environments where visibility is restricted due to habitat complexity or poor light transmission due to tannin-staining or turbidity. Thus the use of sound by turtles may be an important communication medium over distances beyond their visual range. This study reports the first records of an underwater acoustic repertoire in an aquatic chelonian.
 [DOI: 10.1121/1.3148209]

PACS number(s): 43.80.Ka [MCH] Pages: 434–443

I. INTRODUCTION

Little is known of the biological contributors to underwater sound in freshwater environments. Most underwater sound research has been conducted within the marine environment, with little focus of this directed at chelonians. Until recently, there was a paucity of research on the use of sound by chelonians, which perhaps is not surprising as they were once thought to be "the silent group" (Campbell and Evans, 1972), and sound perception reported in reptiles less important than the senses of vision and chemoreception (Bogert, 1960; Lehrer, 1990). However, more than 30 years ago, Campbell and Evans (1972) drew attention to the possibility that turtle vocalizations may play a functional role in chelonian social behavior.

Several studies have described sound production in terrestrial chelonians (Bogert, 1960; Campbell and Evans, 1967, 1972, Auffenberg, 1964; Jackson and Awbrey, 1972; McKeown *et al.*, 1990; Sacchi *et al.*, 2003; Galeotti *et al.*, 2005a, 2005b, 2005c). It appears that only a single marine turtle species has been recorded producing sound—although this was in-air (Mrosovsky, 1972). Most of the research has focused on sounds emitted during breeding activities. For example, roaring or bellowing is observed in the Galapagos tortoise (*Testudo elephantopus*) (Bogert, 1960; Jackson and Awbrey, 1972); with sounds produced by the red-footed tortoise (*Geochelone carbonaria*) during mating described as "a series of clucks"—similar to the calls produced by chickens (Campbell and Evans, 1967). The Malagasy radiated tortoises (*Geochelone radiata*) vocalize in synchrony with xiphiplastral blows (mounted male thrusting on the lower rear edge of the female's carapace by the xiphiplastron, with vocalizations thought to immobilize the female) with these vocalizations described as being similar to "… a nail being pulled from a board in two short jerks" (Auffenberg, 1978, p. 282). Galeotti *et al.* (2005b) described long sequences of whimpers and wailings emitted by Hermann's tortoises (*Testudo hermanii*) during mounting behavior.

There are also references to chelonian vocalizations outside of breeding activities. Carr (1952) described "short rasping calls" or a noise similar to that of a "mewing kitten" produced by the Gopher tortoise (*Gopherus polyphemus*). "Hissing" sounds have been described in agonistic encounters by the Wood turtle (*Clemmys insculpta*) (Kaufmann, 1992). The vocalizations of *Geochelone travancorica* appear

[a]Author to whom correspondence should be addressed. Electronic mail: turtle111@aapt.net.au

to be unique, as it is the only species reported to call in "chorus" where a number of individuals call together at regular intervals (Auffenberg in Campbell and Evans, 1972). Only the young of an aquatic Asian species, *Platysternon megacephalum*, has been observed to "squeal," particularly when disturbed. Loss of this ability to vocalize appears to be related to a certain level of maturation corresponding to a change in appearance, i.e., loss of bright colors, when the carapace length measures around 3 in. and when jaw development was such that they could bite (Campbell and Evans, 1972). Interestingly, despite this range of documented vocalizations, some of the earlier researchers attached little importance to them. For example, Carr (1952) noted that marine turtles produced calls when being hurt or killed—which he considered were incidental and due to the exhalation of breath. Mrosovsky (1972) noted that Leatherback turtles produced a variety of calls when nesting, but he also considered that sound production in turtles was probably of minor functional importance. Goode (1967) observed hissing noises from *C. oblonga* and other Australian chelids such as *Chelodina expansa* and *Elseya latisternum*, which he considered to be involuntary exhalations associated with aggressive behavior. However, recent investigations by Sacchi *et al.* (2003) and Galeotti *et al.* (2005a, 2005b) on *Testudo marginata* and *Testudo hermanni*, respectively, suggest sound is more than "involuntary exhalations," and plays an important role in breeding activities and mate selection.

Little is known of sound production in the group Pleurodira, and indeed, considered not to vocalize (Galeotti *et al.*, 2005c), with nothing known of underwater vocalizations by chelonians. This research investigated underwater sound production by the long-necked, freshwater turtle—the Oblong turtle (*Chelodina oblonga*) (Testudine; Pleurodira; Chelidae). This species was long considered to represent *Chelodina oblonga* Gray 1841, but it was shown recently that the correct name should be Chelodina *colliei* Gray 1856 (Thomson, 2006). Since at present (May 2008) a ruling of the International Commission of Zoological Nomenclature (ICZN) plenum in this case is still outstanding and not yet published, the authors maintain prevailing usage of names in this paper be referring to the south-western Australian longneck turtle species as *C. oblonga* as required by the ICZN rules. This turtle is considered relatively common, although it is under threat by the processes of urbanization (Guyot and Kuchling, 1998). It is only found in the south-west of Western Australia and is common around metropolitan Perth (Cann, 1998) where it inhabits seasonal and permanent wetland areas (Burbidge, 1967).

II. METHODS

Acoustic recordings of *C. oblonga* were made between 2002 and 2005 with more than 500 h of recordings undertaken. Calls with high signal-to-noise ratio were selected for analysis and presentation. Recordings of the adult turtles were undertaken in artificial ponds. These were aboveground, round plastic ponds (0.65 m depth $\times$1.80 m diameter$\sim$1500 L) and were assembled to simulate small wetlands with logs and emergent/floating aquatic plants and a sand base. Juvenile turtles were kept in belowground ponds approximately 0.20 m depth$\times$0.5 m length $\times$0.5 m width and were lined with plastic polythene sheeting 2 mm thick.

Although there are known difficulties associated with recording in laboratory aquariums (Parvulescu, 1966; Hawkins and Myrberg, 1983), artificial ponds were used to ensure any calls recorded could definitely be attributed to turtles rather than other species. Ambient sound recordings in wetlands have revealed diverse biologic acoustic activity (Giles *et al.*, 2005), and although invertebrates later colonized the tanks in small numbers, they were easily recognized. All recordings were accompanied by a human observer, which enabled all external extraneous noises to be accounted for (such as twigs or branches falling in the ponds, presence of avian fauna, or bees buzzing on the hydrophone cable). Due to logistical constraints of equipment, such as the need for battery recharge and downloading of recordings, continuous recordings could not be undertaken. Therefore recordings occurred in the time periods dawn (4am to 6am), midday (11am to 1pm), dusk (5pm to 7pm), and midnight (11pm to 1am). The hydrophones (described in B) were suspended in the center of each pond at a depth of 0.5 m.

A. Turtle populations

In total, 106 adult turtles were used in this research and five juveniles. Turtles were separated into males (47), females (59), and juveniles (5) and placed in separate ponds. The first group of turtles (19 males and 23 females) included sub-adults. A carapace length of 10 cm was used as the threshold length to delineate sub-adults from juveniles. This group was recorded in September and October 2003. The second group of turtles (20 males and 26 females) was recorded from May to November 2004 (juveniles were recorded during this period). This group was also used to determine spring temporal call patterns. The third group of turtles (10 males and 10 females) was recorded from January to February 2005 and was used in summer temporal call studies. In addition, four adult female turtles that had been seized by customs officers were also recorded while recuperating at the authors' residence for 2 days prior to release.

B. Recording equipment

Two hydrophones were used throughout this research: a Cetacean Research Technology (CRT) C53 hydrophone model (frequency response: 14 Hz to 60 kHz$\pm$4.5 dB; nominal sensitivity: $-$165 dB, re 1 V/μPa), and a Hightech HTI-96-MIN (frequency response: 20 Hz to 30 kHz$\pm$3 dB; sensitivity: $-$164 dB re 1 V/μPa).

Recordings were made with a TASCAM DAT recorder, using a sampling rate of 48 kHz (frequency response: 20 Hz–20 kHz). The input level was set at "8" and remained at this setting throughout the study. Equipment was checked to ensure that the DAT recorder gain was the same for left and right channels for all input settings. Frequency response of the tape deck was checked with white noise of known level and was found to be flat.

C. Data analysis

1. Acoustic recordings

Sounds recorded were digitized using a Sound Blaster Audigy DE 24-bit/96 kHz stereo sound card in an Intel Pentium 4 PC. The sound card recording levels remained fixed throughout all analysis. The software used in sound analysis was SpectraPLUS v.2 using a sampling rate of 48 kHz and a sampling precision of 16 bits. Spectrogram images displayed were prepared using MATLAB using a Fast Fourier Transforms (FFT) size of 1024 points with an averaging of 4 (using a 75% overlap), time resolution of 5.3 ms, and frequency resolution of 46.9 Hz. A Hanning smoothing window was used.

2. Classification and terminology of turtle calls

Classification of turtle calls was based on terminology used in insect (e.g., Broughton, 1963; Jansson, 1973), bird song (e.g., Shiovitz, 1975; Thompson *et al.*, 1994), and some of the aquatic mammalian groups such as cetaceans (e.g., Clark, 1982). Turtle calls were assigned to categories based on aural character, frequency, and duration. While most of the turtle calls could be separated into distinctive categories, there was some variation in the spectral nature within groupings. In particular, "chirp" vocalizations were generally polymorphic and it was difficult to separate these sounds into distinctive categories. While they were intuitively similar because of the way in which they sounded, use of the name chirp is essentially generic, as phonetically, these vocalizations consisted of a range of sounds such as "eeaow," "MmM," "M," "oi," "ow," or "ar." A format was used similar to that of Cosica *et al.* (1991) where chirp calls were identified and separated according to at least two acoustic parameters. For example, the "short chirp," the "high short chirp," the "medium chirp" and the "long chirp with the long frequency up/down-sweep elements" were a similar type of sound to listen to, but they differed in their duration and their spectral structure—in particular, the long up/down-sweep elements were distinctive and long chirps often contained three harmonics. Short chirps were brief calls containing one to two harmonics, but the high short chirps contained three harmonics and higher frequency elements compared to short chirps. Call types were further defined into acoustic units as being a pulse, note, syllable, bout, or burst (for definitions, see Table I). As calls were recorded in an artificial environment, it is understood that not all acoustic descriptions will reflect free-field recordings, in particular, duration.

TABLE I. Definitions used for acoustic units. These definitions have been adapted from existing literature to enable classification of turtle calls.

Acoustic unit	Definition
Pulse	Defines the shortest and simplest of turtle calls. Pulses had a duration of ~0.05 s.
Note	Defines the shortest of the complex calls, with a duration of ~0.075 s usually with first and second harmonics present with little to no frequency modulation.
Syllable	Defines a longer duration call (~0.30 s or longer) and was more complex. Harmonically-structured with different rates of frequency modulation throughout the call and often finishing with well-defined up-/down-sweep elements.
Bout	Defines a sequence of three or more pulses, notes, syllables, or bursts—called at intervals of ~1–25 s and could occur over a period of several minutes.
Burst	Defines noisy calls with a harsh or strident sound quality. These chaotic segments had a spectral structure either coherent or incoherent, but usually no harmonics or frequency modulation (aperiodic). Call length was of variable duration, but usually long syllables.

D. Temporal call patterns

To determine if differences in temporal calling patterns occurred between different size/sex classes and reproductive readiness, the second turtle group was separated into five groups based on carapace length and for females, follicle development (using follicle diameter of 12–19 mm) (Table II), and placed in separate ponds. 1 h recordings were made in each pond for the time periods dawn, midday, dusk, and midnight, once a week for 3 weeks in spring (20 September 2004 to 10 October 2004). In total 60 h of recordings were made (12 h per group or 3 h in total for each time period). The third turtle group was separated into males and females (Table II) and was recorded at similar times in summer (11–31 January 2005), with a total of 48 h of recordings (24 h per group or 6 h total for each time period). Summer was a time when the influence of breeding activities was not expected to be important. Temperatures were taken at all recording times.

TABLE II. Size and sex of turtles used in the spring and summer temporal call patterns, and number of chirp calls made in a 3-week period.

Turtle group	Season	Carapace length (cm)	Mean ± SD	Total No. of calls	Total No. of turtles	No. of calls per turtle
Large males (LMs)	Spring	20.6–22.6	21.2 ± 0.64	7	10	0.7
Small males (SMs)	Spring	16.5–18.4	17.7 ± 0.65	15	8	1.9
Large females with follicles (12–19 mm) (LFWFs)	Spring	24.2–28.2	25.9 ± 1.4	4	6	0.7
Females with follicles (12–19 mm) (FWFs)	Spring	19.9–23.05	21.9 ± 1.2	13	10	1.3
Females without follicles (FWOFs)	Spring	19.9–24.1	22.0 ± 1.4	1	10	0.1
Females	Summer	17.5–20.9	19.1 ± 1.0	98	10	9.8
Males	Summer	15.0–20.6	17.4 ± 1.5	104	10	10.4

TABLE III. Summary of the complex calls produced by *Chelodina oblonga*. Calls were categorized according to their aural character and spectral structure. Parameters measured were (1) frequency range in kilohertz (from the lowest to the highest measurable frequency), (2) mean duration (and SD) of the signal in seconds, and (3) dominant frequency in kilohertz (frequency of harmonic with the most energy). The first harmonic was taken as the fundamental or lowest frequency in the harmonically-structured calls.

Aural character and category	Spectral output	Mean duration (s)	Frequency range (kHz)	Periodic/ aperiodic	Dominant frequency (kHz)	Sex
Short chirp (single note) ($n=10$)	Harmonics (1–2)	0.07 ± 0.02	0.60–2.0	Periodic (complex)	0.90 and 1.80	M/F
Short chirp (juvenile) (single note) ($n=1$)	Short-upsweep	0.05	1.0–1.2	Periodic	1.1	Juvenile
High short chirp (syllable) ($n=7$)	Harmonics (3–5)	0.11 ± 0.02	0.40–2.8	Periodic (complex) FM	0.88	M
Hoots (syllable) ($n=12$)	Richly harmonic (10)	0.15 ± 0.08	0.12–2.3	Periodic (complex)	0.16–0.20	M/F
Squawks (syllable) ($n=5$)	Harmonics (=2)	0.21 ± 0.04	0.60–1.8	Periodic (complex)	1.5–1.8	M
Medium chirp (syllable) ($n=27$)	Sparsely harmonic (≥ 3)	0.29 ± 0.12	0.78–1.6	Periodic (complex) FM	0.60–0.80	M/F
SH long chirp (syllable) ($n=31$)	Sparsely harmonic (=3)	0.36 ± 0.14	0.65–1.6	Periodic (complex) FM	0.70–0.80	M/F
RH long chirp (syllable) ($n=34$)	Richly harmonic (=5–6)	0.34 ± 0.12	0.26–3.8	Periodic (complex) FM	0.46–0.65	M/F
High calls (syllable) ($n=6$)	Sparsely harmonic (=3)	0.46 ± 0.19	0.21–3.5	Periodic (complex) FM	0.95–1.8	F
Hooo's (syllable) ($n=5$)	Sparsely harmonic (≥ 3)	0.66 ± 0.70	0.13–2.5	Periodic (complex)	0.13–0.60	M/F
Wails (syllable) ($n=5$ long and 2 short)	Sparsely harmonic (≥ 4)	Long wails 1.32 ± 0.4; short wails 57 ± 0.02	0.19–1.5	Periodic (complex) FM	0.42	M/F
Duck Honks (syllable) ($n=11$)	Discordant coherent structure	0.11 ± 0.02	0.10–3.15	Periodic (complex)	0.18	Sub-adult male

III. RESULTS

A. Sound classification

1. Description of turtle calls

Turtle calls recorded in the artificial ponds were classified into 17 categories, with call characteristics summarized in Tables III and IV (juvenile calls are included in this table). Calls were also produced in-air at the waters' surface but these are not presented here. Vocalizations consisted of clacks, clicks, squawks, hoots, short chirps, high short chirps, medium chirps, long chirps, high calls, cries or wails,

TABLE IV. Summary of the percussive calls produced by *Chelodina oblonga*. Calls were categorized according to their aural character and spectral structure. Parameters measured were (1) frequency range in kilohertz (from the lowest to the highest measurable frequency), (2) mean duration (and SD) of the signal in seconds, and (3) dominant frequency in kilohertz.

Aural character and category	Spectral output	Mean duration (s)	Frequency range (kHz)	Periodic/ aperiodic	Dominant frequency (kHz)	Sex
Clacks (pulse) (numerous: using random $n=17$)	Continuous	0.05 ± 0.01	1.4–2.1	Periodic	1.5–2.0	M/F
Broadband Clicks echo-location pulses? (numerous: using random $n=18$)	Continuous	0.05 ± 0.01	$0.10 - > 20$	Aperiodic	8.0–16.0 and 0.85–1.70	M/F
Grunts (pulses) (numerous using $n=12$)	Noisy	0.08 ± 0.3	0.10–2.5	Aperiodic	≤ 0.36	M/F
Growls (bursts) (numerous)	Noisy unstructured varied	≤ 2.0	0.10–1.1	Aperiodic	≤ 0.20	M/F
Blow bursts (bursts) (numerous)	Noisy spectrally coherent	Varied ≤ 3.0	0.10–10	Aperiodic	≤ 0.30	M/F
Drum rolls (pulses) ($n=5$)	Coherent repetitive pulses	2.72 ± 1.23 (Males)	0.10–0.75	Aperiodic	≤ 0.21	M/F
Staccato (pulses)	Rapid pulse-series	Varied	0.103–1.0 0.10–10.0 (only adult M/F using the higher frequencies)	Aperiodic	<0.12 and <0.21	M/F and juveniles
Wild howl (syllable and pulses) ($n=1$)	Contains richly harmonic elements (=7) pulsed and noisy elements	10.263	0.10–3.0	Periodic/ aperiodic	Howl 0.21–0.39 0.57–0.70; growling rattle ≤ 0.17	F

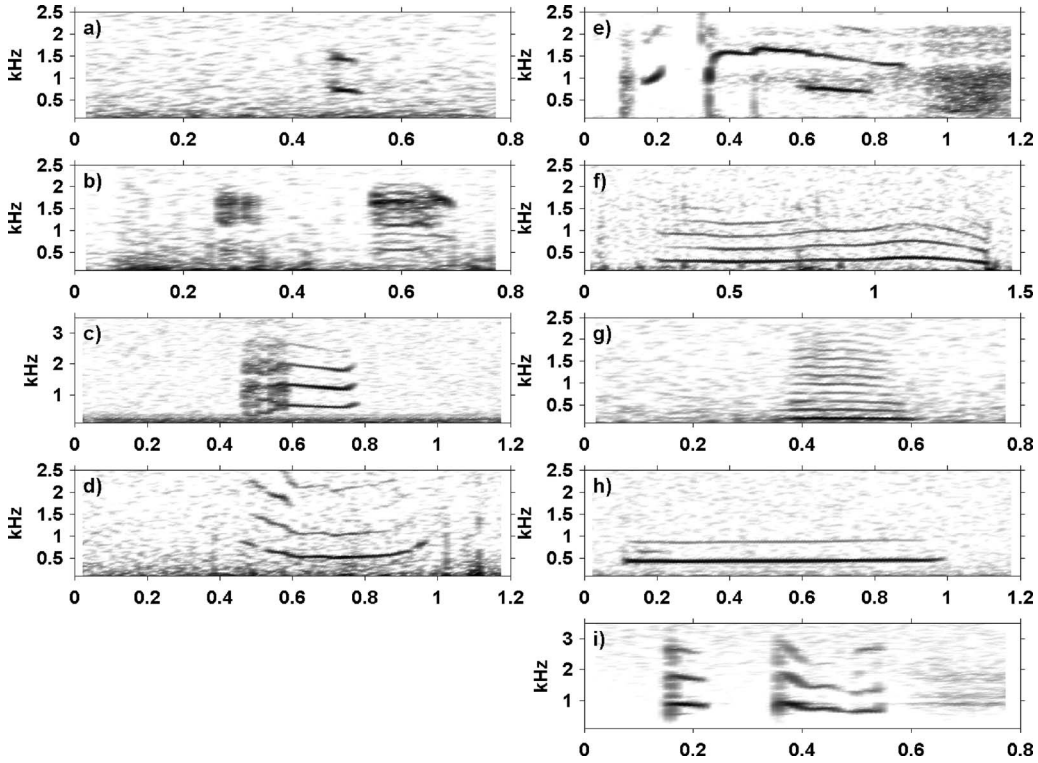


FIG. 1. Spectrograms of vocalizations produced by *Chelodina oblonga* described in Table III: (a) short chirps, (b) squawks, (c) hoots, (d) sparsely harmonic long chirp, (e) high call, (f) wail, (g) richly harmonic long chirp, (h) hooo, and (i) high short chirp with a medium chirp (juvenile calls are not shown here). Analysis frequency and time resolution used are 23.44 Hz and 0.043 s, respectively, with a 0.8 FFT overlap. The *x*-axis is time in seconds.

hooos, grunts, growls, blow bursts, staccatos, a wild howl, and drum rolling.

Spectrograms for each call type are shown in Figs. 1 and 2. Each call has three main parameters measured from the spectrograms: (1) frequency range in kilohertz (from the lowest to the highest measurable frequency), (2) average duration of the signal in seconds, and (3) the dominant frequency in kilohertz (frequency of that harmonic with the

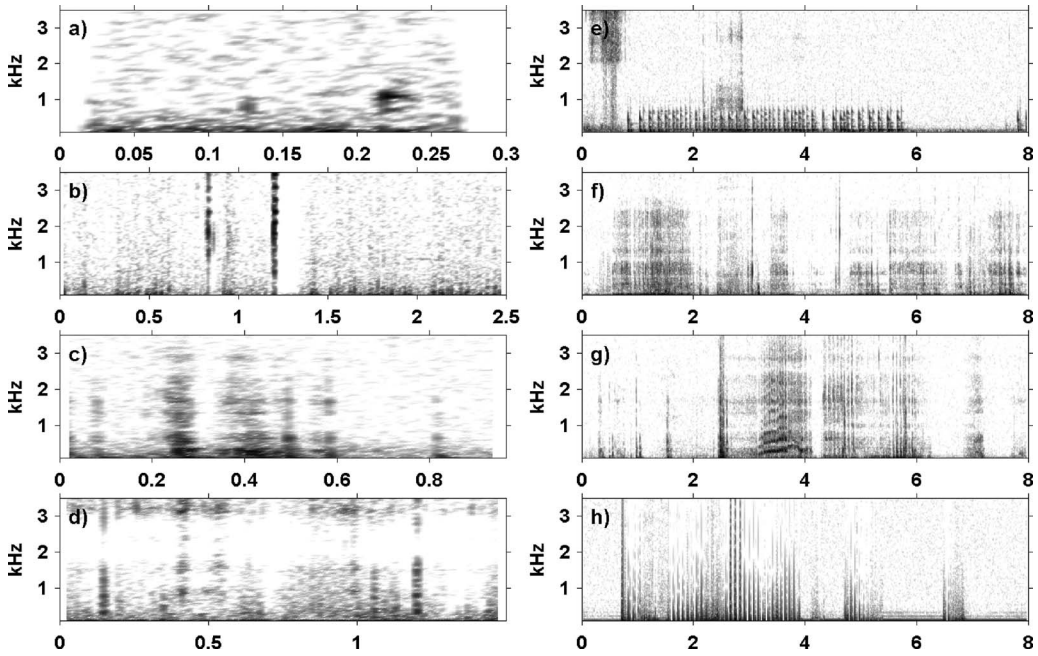


FIG. 2. Spectrograms of percussive call vocalizations produced by *Chelodina oblonga* described in Table III: (a) clacks, (b) double-clicks, (c) grunts, (d) blow bursts, (e) drum rolls, (f) growling, (g) wild howl, and (h) staccato. Analysis frequency and time resolution used are 23.44 Hz and 0.043 s, respectively, with a 0.8 FFT overlap. The *x*-axis is time in seconds.

TABLE V. Summary of the sustained vocalization recorded in the artificial pond produced by a male long-necked, freshwater turtle. The call was described according to its aural character and spectral structure. Parameters measured were (1) frequency range in kilohertz (from the lowest to the highest measurable frequency), (2) duration of the sustained call (minutes), and (3) the dominant frequency in kilohertz (frequency of that harmonic with the greatest energy).

Vocalization: aural character	Spectral output	Duration of display (min)	Frequency range (kHz)	Dominant frequency (kHz)	Sex
Bongo-drums ($n=1$ pulse-bout)	Pulsed-spectral peaks	9.5	0.10–1.8	0.26–3.0	Male

greatest amplitude). Also noted for each call was the presence of frequency modulation and harmonics. The first harmonic was taken as the lowest frequency in the harmonically-structured calls.

Some similar vocalization types occurred in a bout, such as "chirps," "hoots," "drum rolls," and "wails." Wails occurred in a bout, which lasted ~79 s. Wails consisted of sparse, harmonically-related elements, in a pattern of a long wail followed by a short wail, with the long wails becoming progressively longer and finishing on a short "FM wail." Drum rolls consisted of coherent and repetitive "rolling-pulse-like" sounds in a doublet pattern, occurring as a bout of five separate rolls, with each roll lasting on average 2.7 s (SD=1.3, $n=5$). One drum roll bout lasted around 1 min while the other lasted around 5 min. In addition, different vocal categories occurred together in a bout. For example, "grunts" and "staccatos" were heard in the "wild howl."

"Broadband clicks" were sudden onset sounds and occurred singly, or as distinctive "double-clicks," and occasionally, in a rapid click series. Energy in the clicks often overloaded the recording gear with a frequency range above the recording system limit (i.e., >20 kHz). The double-click sequence had an intra-click interval (time between the two audible clicks on the spectrogram measured from the end of one pulse to the beginning of the next pulse) of 0.36 s (SD =0.15, $n=9$). Growls and blow bursts were continuous chaotic (noisy), broadband calls with low frequency concentration of energy that overloaded the recording equipment at the settings used for all the complex, periodic calls. Staccatos consisted of a series of rapid, "thump-like" pulses, considered more mechanical than biological.

Juveniles (<10 cm) rarely vocalized using complex calls. Only a single short chirp of brief duration (0.05 s) was recorded by juveniles and consisted of a rapid up-sweep and no harmonics. Staccatos were, however, a prevalent call and occurred in single pulses or short series, and had a similar structure to the adult staccato.

2. Sustained vocalizations

One display of a sustained vocalization consisting of a bout of pulses was recorded in the artificial ponds (October 2003). Call characteristics have been summarized in Table V and acoustic units are presented in Table VI. The "pulse-bout" has been divided into "first phase" and "second phase" (Fig. 3). The two phases have then been further separated into four sections and these are listed below.

The first phase has the following.

(1) The introductory stage. This was a single pulse/slow tempo stage.
(2) The second stage had a fast tempo with minimal silent intervals.
(3) The third stage had well-defined shorter pulse-series with silent intervals more often.

The second phase has the following.

(4) The fourth stage or *vibrato* was the rhythmic part of the pulse-bout.

The pulse-bout recorded in the artificial pond was the longest vocal display by *C. oblonga*. It was performed over 9.5 min, consisting of 817 individual pulses with a low dominant frequency of around 280 Hz, with lesser peaks at 800 and 1900 Hz. The tempo in the second phase was more rapid than the first phase. The vibrato occurred after almost 7.5 min of the first phase and lasted for 2 min. It consisted mainly of doublets, triplets, and/or quadruplets, with higher frequency elements in each pulse "rolling down" from approximately 800 to 620 Hz, and similarly, the lower frequency elements in each pulse rolling down from around 350 to 155 Hz. The sound of this vibrato is described as bongo-drums."

B. Temporal call patterns

Chirp calls were the most prevalent of all turtle calls. In spring, complex turtle vocalizations consisted of the short duration chirps: (1) short chirps and (2) medium chirps—with most calls occurring at midday. Although dusk was the

TABLE VI. Definitions used for acoustic units related to the sustained vocalizations. These definitions have been adapted from those in literature in order to relate them to the sustained pulse-bouts produced by turtles.

Acoustic unit	Definition
Pulse	Defines the individual component of the sustained turtle vocalizations. Pulses ranged from 200 Hz to around 1.8–2 kHz often revealing three or more spectral peaks.
Inter-pulse interval (IPI)	The interval between the end of one pulse and the beginning of the next pulse in a pulse-series, or between single pulses that had an interval of less than 1 s.
Pulse-series	Consists of a number of pulses ranging from 2 to 65, heard as sequences of pulses separated by short intervals of silence. The IPI in a pulse-series were usually irregularly spaced.
Silent interval	This refers to the brief periods of silence separating single pulses or a pulse-series from the next sequence of pulse(s). The silent interval usually ranged from 1 to 8 s, but occasionally up to 35 s.
Vibrato	Consists of a rapid series of pulses heard as a rhythmic percussive display appearing after the first phase (after the single pulse and irregular pulse-series). The vibrato phrase denotes the second phase.
Pulse-bout	The entire sequence of the sustained turtle vocalization.

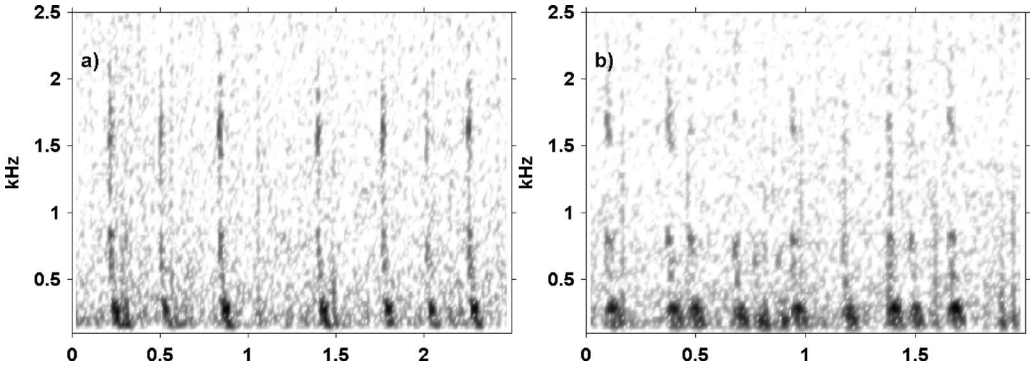


FIG. 3. Spectrograms of the sustained vocalization called by a male *Chelodina oblonga* turtle in the artificial pond: (a) a short segment of pulses in the first phase and (b) a short segment of the vibrato of the second phase. Analysis frequency and time resolution used are 23.44 Hz and 0.043 s, respectively, with a 0.8 FFT overlap. The *x*-axis is time in seconds.

warmest time period, and nearly all groups, except the large males (LMs), utilized the dusk period for calling, fewer calls were made in this period compared to midday. Only females with follicles (FWFs) and the small males (SMs) utilized the dawn period and only males vocalized at midnight. Overall, it was the small males (total of 15 calls in 12 h of recordings) and the FWFs (total of 13 calls in 12 h of recordings) that were most vocally active, and appeared to follow similar calling trends; the exception being at midnight, when only males were vocally active. Females without follicles (FWOFs) called the least with only one call being recorded. Large females with follicles (LFWFs) called approximately the same amount as the LMs (Table II).

Turtles called more in summer than spring despite the doubled recording effort (Table II). In summer, females called more often at dawn and midday than males (28 dawn calls and 39 midday calls by females; cf. 20 dawn calls and 26 midday calls by males), while males called more often at dusk and midnight than females (33 dusk calls and 25 midnight calls by males; cf. 26 dusk calls and 5 midnight calls by females) [Figs. 4(a) and 4(b)]. However, despite these differences in number of calls, no evidence was found between the sexes and the number of calls in each time period [two sample t-test (two-tailed): dawn: t-stat=0.339, $df=4$, and p-value=0.752; midday: t-stat=0, $df=3$, and p-value =1.0; dusk: t-stat=−0.717, $df=3$, and p-value=0.525; and midnight: t-stat=−1.031, $df=3$, and p-value=0.378]. The most utilized chirp vocalization by both Blue Gum female and male turtles was the short chirp, followed by the medium chirp with males using these vocalizations slightly more often than females (73 short chirps/20 medium chirps for females and 78 short chirps/22 medium chirps for males).

IV. DISCUSSION

A. Sound classification

Chelodina oblonga is not a vocal specialist but the 17 categories described in the vocal repertoire and their variety and use in bouts of multiple call components are suggestive of complex social roles. The number of categories is likely to be greater, as other calls were heard, but due to difficulties with equipment and recordings were unable to be described (e.g., cat whines and moans). While the present study provides a preliminary categorization of turtle vocalizations, future research may elicit information on the significance of the variation within each call and the function of calls.

The anatomical site at which sound production occurred in *C. oblonga* or the mechanism to produce sound is not known. However, larynx morphology in three species of tor-

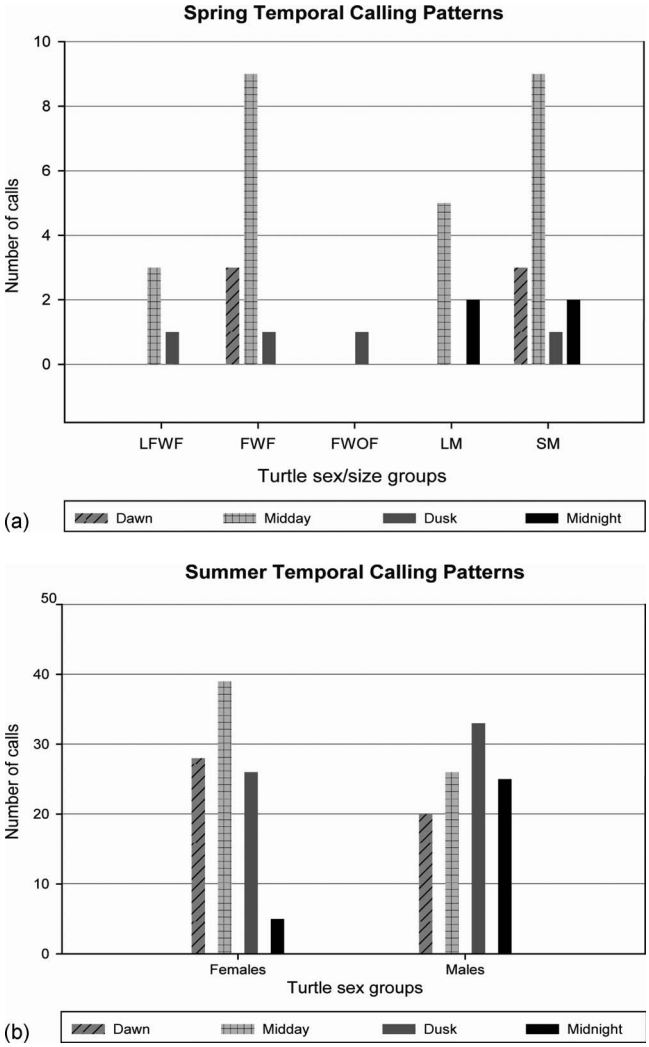


FIG. 4. (a) Spring temporal calling patterns for the LMs, SMs, LFWFs, FWFs, and FWOFs in a 3-week period. (b) Summer temporal calling patterns between male and female turtles in a 3-week period.

toises has recently been described by Sacchi *et al.* (2004). They found the larynx to be a simple structure consisting of three cartilages: the cricoid and two arytenoids, which form the frame of the larynx, with opening and closing of the glottis controlled by two pairs of muscles. While Sacchi *et al.* (2004) found no true vocal cords, two bands of elastic fibers were found in the anterior wall of the larynx. As the fibers were capable of vibrating, Sacchi *et al.* (2004) suggested that they might be the equivalent of vocal cords.

Given the structural complexity seen in some of *C. oblonga's* calls, vocalization will involve the movement of air across specific vibrating structures. However, the authors propose that the aquatic chelonian is using a closed system with circulating air movement. In particular, air movement was distinctly heard at the end of the "high call," but was not accompanied by air bubbles at the surface (or heard in recordings), and neither did there appear to be repeated need for surfacing.

Although three populations of turtles were recorded in this research, this was not considered to be enough to indicate whether the repertoire presented reflects the complete extent of *C. oblonga's* vocal ability. However, the large number of categories suggests that a considerable proportion of the vocal repertoire has been described. While inter-individual variation in vocalizations has been noted in some tortoises, e.g., *Geochelone radiata* (Auffenberg, 1978) and *Testudo marginata* (Sacchi *et al.*, 2003), this was difficult to test for in *C. oblonga*. When recording a group of animals underwater, it is difficult to know which animal has called (see Tyack, 2001). While hydrophone arrays are used to give bearing information in field conditions, arrays were not used in this research. In an attempt to test for inter-individuality, 7 h of recordings were made of a lone female; however, no vocalizations were recorded.

Similarly, for the reasons above, and tannin-staining in the water, behavioral observations were difficult to correspond with a call type. Although some anecdotal observations were made throughout the study and some methods for observation were trailed, it was considered inappropriate to disturb the turtles, so as to record calling behavior that might reflect their natural vocal activity. From field recordings made in the wetlands, standing at the waters' edge was noted to cause turtles to swim away quickly in what was considered a "flee response."

Frequency use in turtle vocalizations started from around 100 Hz in some of the percussive displays, ranging to as high as 3.5 kHz in some of the complex calls such as the high calls, with clicks extending beyond the upper 20 kHz limit of the recording equipment. However, most turtle vocalizations had dominant frequencies below 1 kHz. Based on the frequency ranges recorded for the complex vocalizations of *C. oblonga*, it might be reasonable to suggest that their aquatic hearing frequency range extends to these higher frequencies. The auditory sensitivity of turtle hearing is considered to fall away above 1–2 kHz (Wever, 1978; Legler, 1993). However, given the work by Wever (1978) on the effects of temperature on auditory sensitivity, and the presence of higher frequencies in some harmonically-structured calls; it suggests that turtles may well have greater sensitivity at higher frequencies than previously thought. This may be important during summer months when propagation characteristics of a shallower environment favor higher frequencies (see Forrest *et al.*, 1993).

B. Sustained vocalizations

In addition to the short single calls or call bouts produced by *C. oblonga*, this research has also revealed that males are capable of producing sustained vocalizations—lasting up to around 9 min. On one occasion, a sustained vocalization was also recorded in an urban wetland in September 2002, which was performed over almost 4 min (3.8 min) and consisted of 210 individual pulses utilizing a dominant frequency of around 886 Hz with lesser peaks at around 460 and 1645 Hz. The vibrato occurred after only 3 min of the first phase and lasted for less than 1 min (59 s). Pulses in the field vibrato generally appeared as doublets and were heard as a series of rocking pulses with alternate pulses having different dominant frequencies, i.e., the odd pulses had dominant frequencies extending from around 550 to 1030 Hz, while the even pulses had dominant frequencies extending from 760 to 1185 Hz. While it cannot be confirmed, it is thought likely to have been produced by a turtle as no other vertebrates were observed in the water and a turtle was nearby to the hydrophone on this occasion.

Females were not recorded producing sustained vocalizations, although their ability to do so cannot be discounted. Reproductive displays are common in many animal groups—terrestrial and aquatic, but acoustic advertisement displays related to reproductive cycles in reptiles are rarely reported in literature—with the exception of the Tokay gecko (*Gekko gecko*) (Tang *et al.*, 2001). The acoustic advertisement displays in the Tokay gecko are seasonal and are coincident with a rise in androgen levels and gonadal masses (Tang *et al.*, 2001) and from this would also be called reproductive advertisement displays. Acoustic advertisement displays have not been reported in chelonians, with most vocalizations associated with reproductive activity, only heard during mounting behavior in the terrestrial chelonians (e.g., Bogert, 1960; Auffenberg, 1964; Campbell and Evans, 1967; Jackson and Awbrey, 1972; Sacchi *et al.*, 2003; and Galeotti *et al.*, 2005a), with a single report of some pre-courtship vocalizations (McKeown *et al.*, 1990). The pulse-bouts of *C. oblonga* appear to share a similar pattern to that reported for the Tokay gecko. Tang *et al.* (2001) described a two-phase arrangement in the advertisement display for this species, which began with two to three simple multipulse sequences heard as rattles, with the second phase more complex containing a series of doublets (4–11) (or referred to as "binotes"). It is proposed that the two-phase approach to the acoustic display may perform two main functions: (1) the first phase may act as an "announcement phase" used to gain the attention of as many conspecifics (particularly females) as possible with (2) the second phase being an "identifying phase" where the male is able to "showcase" himself as a desirable mate.

Breeding aggregations have not been reported in literature for the Oblong turtle, and if females are widely dis-

persed throughout the wetland or are difficult to see, it seems reasonable to hypothesize that aquatic chelonians would have an acoustic advertisement display (both turtle pulse-bouts were recorded within the breeding months). The first phase of the turtle pulse-bout consisted of simple repetitions of stereotypical pulses, which enhances long-range communication as simple repetitions have the effect of making the call stand out from the background noise (Wiley and Richards, 1978). Both pulse-bouts exhibited a broad spectrum of high-energy elements. This wide frequency band would enable a range of propagation distances for frequency limited environments. This could be a strategy to call females at a range of distances. In addition, it would be important for the receiving female to then locate the male if she was interested. Pulse-bouts used repetition, a broad range of frequencies, and a broken structure, which are all characteristics to enable a receiving animal to binaurally locate the sender more easily (Marler, 1967, 1977).

In Hermann's tortoise (*Testudo hermanni*), Galeotti *et al.* (2005a) noted that successful breeding is likely to be a factor of how well a male can attract a female if they are widely dispersed (or unable to be seen as would be the case underwater), rather than aggressive encounters with other males, e.g., as seen in the Desert tortoise *Gopherus agassizii* (Niblick *et al.*, 1994). While there were some aggressive encounters observed throughout the study, these were intermittent and brief, consisting of biting and short chases (lasting less than a few seconds), there was no fighting observed. In the second phase of the turtle pulse-bout, which contained the vibrato, this was considered to be the stage of the pulse-bout where the male showcases himself acoustically. While there have only been two examples of pulse-bouts recorded to date for these animals, both vibratos were different in terms of their structure and aural character. There is some suggestion that vocal behavior may be costly to males and that their vocalizations would vary, containing information for females on the quality or desirable attributes of the male (Sacchi *et al.*, 2003). In Marginated tortoises (*Testudo marginata*), Sacchi *et al.* (2003) found that call features differed significantly between males and certain call characteristics (call rate and duration) were well correlated with condition of a male and successful mating—indicating that aspects of vocalizations may be revealing information to the female for choice of a mate. Female Hermann's tortoises were shown to prefer fast-rate and high-pitched calls (Galeotti *et al.*, 2005a). In this study, *C. oblonga's* vibrato consisted of fast-rate and complex sequences of doublets and triplets, with the field recording consisting of rocking pulses. Whether *C. oblonga* females exhibit a preference for a particular call type is not known. Galeotti *et al.* (2005b) found courtship displays in Hermann tortoises, which include vocal activity, consistently revealing condition of the male. Condition of the male was based on morphological characteristics (body mass, carapace length, and head size) and their relationship to blood parameters, courtship intensity, and mounting success.

C. Temporal call patterns

Ambient temperature can influence behavior in reptiles, including anti-predator behavior (Crowley and Pietruszka, 1983). Temperature also appears to influence vocal activity in *C. oblonga*. More calls were recorded in summer than spring. No complex calls were recorded below 10 °C, although, occasional clicks, grunts/growls, and blow bursts were recorded. Given the prevalence of chirp calls, they are considered to be their main social calls. The lack of complex calls at <10 °C would appear to correspond to the inactivity reported in some freshwater turtle species at this temperature, e.g., *Chrysemys picta* (Ernst, 1971).

The paucity of vocal activity by the FWOFs was a consistent trend exhibited throughout the winter/spring months. It is proposed that this may be related to the lack of follicle development and therefore reproductive non-receptiveness. The lack of activity—vocal or movement—in non-sexually active females would be an advantage for males seeking mates, as it would be energetically expensive for males to seek out reproductively non-receptive females. However, this explanation does not account for the paucity of vocal activity by LMs with follicles or by LMs in the spring temporal calling study, as both these groups were vocally active prior to the temporal call study and following it. Longer recording windows need to be undertaken in future to account for the type of weekly variation that a short window of recording cannot account for. Despite the short recording window, it appears that a pattern may be emerging that males have a preference for vocalizing later at night. However, no explanation can be offered at this point as to why this might be.

D. Future research

Future research is needed to determine the importance of sound to *Chelodina oblonga* by (1) examining the behavioral significance of all turtle vocalizations, (2) the biological function of the turtle calls, and (3) their sound-producing mechanism.

This study has presented the first records of an underwater acoustic repertoire in a freshwater turtle. The question remains as to how common and widespread underwater sound communication is in aquatic chelonians—fresh and marine. Some recent preliminary recordings made of the Murray River turtle (*Emydura macquarii*) suggest it is likely to be more widespread than currently realized, at least among freshwater chelonians.

ACKNOWLEDGMENTS

This research was funded under a Ph.D. scholarship provided by Murdoch University with additional financial support kindly donated by the Australian artist William Boissevain. Sincere thanks to Dan Hewitt, Alec Duncan, Joe Olson, and David Matthews for their assistance in underwater acoustics and technical matters. This research was undertaken on approval by Murdoch University Animal Ethics Committee Grant No. 931W/02 and performed under WA Conservation and Land Management License Nos. SF003994, SF004286, SF004311, and SF004702. The authors wish to state that there are no conflicting interests.

Auffenberg, W. (**1964**). "Notes on the courtship of the land tortoise *Geochelone travancorica* (Boulenger)," J. Bombay Natural History Society **61**, 247–253.
Auffenberg, W. (**1978**). "Courtship and breeding behavior in G*eochelone radiata* (testudines: Testudinidae)," Herpetologica **34**, 277–287.
Bogert, C. M. (**1960**). "Influence of sound on amphibians and reptiles," in *Animal Sounds and Communication*, edited by W. E. Lanyon and W. N. Tavolga (American Institute of Biological Sciences, Washington, DC), pp. 137–320.
Broughton, W. B. (**1963**). "Method in bioacoustics terminology," in *Acoustic Behavior of Animals*, edited by R.-G. Busnel (Elsevier, Amsterdam), pp. 3–24.
Burbidge, A. A. (**1967**). "The biology of south-western Australian tortoises," Ph.D. thesis, University of Western Australia, Australia.
Campbell, H. W., and Evans, W. E. (**1967**). "Sound production in two species of tortoises," Herpetologica **23**, 204–209.
Campbell, H. W., and Evans, W. E. (**1972**). "Observations on the vocal behavior of Chelonians," Herpetologica **28**, 277–280.
Cann, J. (**1998**). *Australian Freshwater Turtles* (John Cann and Beaumont, Queensland), pp. 75–80.
Carr, A. (**1952**). *Handbook of Turtles: The Turtles of the United States, Canada and Baja California* (Constable and Company, London), pp. 5–41.
Clark, C. W. (**1982**). "The acoustic repertoire of the Southern Right whale, a quantitative analysis," Anim. Behav. **30**, 1060–1071.
Cosica, E. M., Phillips, D. P., and Fentress, J. C. (**1991**). "Spectral analysis of neonatal wolf *Canis lupus* vocalizations," Bioacoustics **3**, 275–293.
Crowley, S. R., and Pietruszka, R. D. (**1983**). "Aggressiveness and vocalization in the Leopard lizard (*Gambella wislizennii*): The influence of temperature," Anim. Behav. **31**, 1055–1060.
Ernst, C. H. (**1971**). "Population dynamics and activity cycles of *Chrysemys picta* in southeastern Pennsylvania," J. Herpetol. **5**, 151–160.
Forrest, T. G., Miller, G. L., and Zagar, J. R. (**1993**). "Sound propagation in shallow water: Implications for acoustic communication by aquatic animals," Bioacoustics **4**, 259–270.
Galeotti, P., Sacchi, R., Rosa, D. P., and Fasola, M. (**2005a**). "Female preference for fast-rate, high-pitched calls in Hermann's tortoises (*Testudo hermanni*)," Behav. Ecol. **16**, 301–308.
Galeotti, P., Sacchi, R., Fasola, M., Pellitteri, D., Rosa, D. P., Marchesi, M., and Ballasina, D. (**2005b**). "Courtship displays and mounting calls are honest condition-dependent signals that influence mounting success in Hermann's tortoises," Can. J. Zool. **83**, 1306–1313.
Galeotti, P., Sacchi, R., Fasola, M., and Ballasina, D. (**2005c**). "Do mounting vocalizations in tortoises have a communication function? A comparative analysis," Herpetological Journal **15**, 61–71.
Giles, J. C., Davis, J. A., McCauley, R. D., and Kuchling, G. (**2005**). "The ambient sound field in three freshwater environments," in *Acoustics in a Changing Environment: Proceedings of the Annual Conference of the Australian Acoustical Society* (Australian Acoustical Society, Castlemaine), pp. 383–389.
Goode, J. (**1967**). *Freshwater Tortoises of Australia and New Guinea (in the Family Chelidae)* (Lansdowne, Melbourne), pp. 119–122.
Guyot, G., and Kuchling, G. (**1998**). "Some ecological aspects of populations of Oblong Turtles (*Chelodina oblonga*) in the suburbs of Perth (Western Australia)," in *Le Bourget Du Lac*, edited by C. Miaud and R. Guyetant (Societas Europaea Herpetologica, France), pp. 173–181.
Hawkins, A. D., and Myrberg, A. A., Jr. (**1983**). "Hearing and sound communication underwater," in *Bioacoustics: A Comparative Approach*, edited by B. Lewis (Academic, London), pp. 347–405.
Jackson, C. G., and Awbrey, F. T. (**1972**). "Mating bellows of the Galapagos Tortoise, *Geochelone elephantopus*," Herpetologica **34**, 134–136.
Jansson, A. (**1973**). "Stridulation and its significance in the genus Cenocorixa (Hemiptera: Corixidae)," Behavior **46**, 1–36.
Kaufmann, J. H. (**1992**). "The social behavior of Wood Turtles, *Clemmys insculpta*, in Central Pennsylvania," Herpetological Monographs **6**, 1–25.
Legler, J. M. (**1993**). "Morphology and physiology of the Chelonia," in *Fauna of Australia*, edited by C. J. Glasby, G. J. B. Ross, and P. L. Beesley (Australian Government Publishing Service, Canberra), Vol. **2A**, pp. 108–119.
Lehrer, J. (**1990**). *Turtles and Tortoises* (Headline, London), pp. 9–39.
Marler, P. R. (**1967**). "Animal communication signals," Science **157**, 769–774.
Marler, P. R. (**1977**). "The structure of animal communication sounds," in *Recognition of Complex Acoustic Signals*, edited by T. H. Bullock (Abakon Verlagsgellschaft, Dahlem Konferenzen, Berlin), pp. 17–36.
McKeown, S., Meier, D. E., and Juvik, J. O. (**1990**). "The management and breeding of the Asian Forest Tortoise (*Manouria emys*) in captivity," in *Proceedings of the First International Symposium on Turtles and Tortoises: Conservation and Captive Husbandy*, edited by K. R. Beaman, F. Caporaso, S. McKeown, and M. D. Graff (California Turtle and Tortoise Club, Van Nuys), pp. 138–159.
Mrosovsky, N. (**1972**). "Spectrographs of the sounds of Leatherback Turtles," Herpetologica **28**, 256–258.
Niblick, H. A., Rostal, D. C., and Classen, T. (**1994**). "Role of male-male interactions and female choice in the mating system of the Desert tortoise, *Gopherus agissizii*," Herpetological Monographs **8**, 124–132.
Parvulescu, A. (**1966**). "The acoustics of small tanks," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. **2**, pp. 7–13.
Sacchi, R., Galeotti, P., and Fasola, M. (**2003**). "Vocalizations and courtship intensity correlate with mounting success in marginated tortoises *Testudo marginata*," Behav. Ecol. Sociobiol. **55**, 95–102.
Sacchi, R., Galeotti, P., Fasola, M., and Gerzeli, G. (**2004**). "Larynx morphology and sound production in three species of Testudinae," J. Morphol. **261**, 175–183.
Shiovitz, K. A. (**1975**). "The process of species-specific song recognition by the Indigo Bunting (Passerina cyanea)," Behavior **55**, 128–179.
Tang, Y.-Z., Zhuang, L.-Z., and Wang, Z.-W. (**2001**). "Advertisement calls and their relation to reproductive cycles in Gekko gecko (Reptilia; lacertilia)," Copeia **2001**(1), 248–253.
Thompson, N. S., LeDoux, K., and Moody, K. (**1994**). "A system for describing bird song units," Bioacoustics **5**, 267–279.
Thomson, S. A. (**2006**). "*Chelodina rugosa* Ogilby, 1890 (currently *Macrochelodina rugosa*; Reptilia, Testudines): Proposed precedence over *Chelodina oblonga* Gray, 1841," Bulletin of Zoological Nomenclature **63**, 187–193.
Tyack, P. L. (**2001**). "Bioacoustics," in *Enclyclopaedia of Ocean Sciences* (Woods Hole Oceanographic Institution, Woods Hole, MA), pp. 295–302.
Wever, E. G. (**1978**). *The Reptile Ear: Its Structure and Function* (Princeton University Press, Princeton, NJ), pp. 832–922.
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.

Analysis of the temporal structure of fish echoes using the dolphin broadband sonar signal

Ikuo Matsuo[a)]
Department of Information Science, Tohoku Gakuin University, Tenjinzawa 2-1-1, Sendai 9813193, Japan

Tomohito Imaizumi and Tomonari Akamatsu
National Research Institute of Fisheries Engineering, Fisheries Research Agency, Hasaki 7620-7, Kamisu 3140408, Japan

Masahiko Furusawa
Tokyo University of Marine Science and Technology, Konan 4-5-7, Minato-ku, Tokyo 1088477, Japan

Yasushi Nishimori
Furuno Electric Co., Ltd., Ashihara-cho 9-52, Nishinomiya 6628580, Japan

(Received 1 December 2008; revised 12 April 2009; accepted 11 May 2009)

Behavioral experiments indicate that dolphins detect and discriminate prey targets through echolocating broadband sonar signals. The fish echo contains components from multiple reflections, including those from the swim bladder and other organs, and can be used for the identification of fish species and the estimation of fish abundance. In this paper, temporal structures were extracted from fish echoes using the cross-correlation function and the lowpass filter. First, the echo was measured from an anesthetized fish in a water tank. The number, reflector intensity, and echo duration were shown to be dependent on the species, individual, and orientation of the fish. In particular, the echo duration provided useful information on the fish body height and for species identification. Second, the echo was measured from the live fish suspended by nylon monofilament lines in the open sea. It was shown that this duration could be estimated regardless of whether or not the fish were moving. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3147505]

PACS number(s): 43.80.Ka, 43.30.Sf, 43.30.Gv, 43.20.Fn [WA] Pages: 444–450

I. INTRODUCTION

Acoustic surveys of fisheries are carried out worldwide using scientific echo sounders (MacLennan, 1990; Simmonds *et al.*, 1992; Simmonds and MacLennan, 2005), and often rely on the target strength (TS) to identify the species or estimate the abundance of fishery resources (Nakken and Olsen, 1977; Foote, 1987; Miyanohana *et al.*, 1990). However, this can be difficult if only the acoustics are analyzed, and it might be necessary to extract detailed information on specific body characteristics from the echo.

Dolphins assess similar information using broadband sonar signals to detect and pursue their prey. Behavioral experiments have found that bottlenose dolphins (*Tursiops truncates*) can determine the target size, material and shape (Hammer and Au, 1980; Au, 1993), and object characteristics (Harley *et al.*, 2003) directly from echoes. Therefore, it is possible that such a dolphin sonar capability is applicable to the improvement of the artificial sounders that are used in acoustic surveys.

An examination of fish echoes by Zakharia *et al.* (1996) using a 40–60 kHz broadband chirp signal and neural network allowed the classification of sardine (*Sardina pilchardus*), anchovy (*Engraulis encrasicolus*), and horse mackerel (*Trachurus trachurus*) at sea. Furthermore, Simmonds *et al.* (1996) identified fish species for caged aggregations with a success rate of approximately 95% using neural networks to analyze the spectra of echoes. However, in order to assess selective fisheries, it is necessary to obtain information on individual fish as well as on fish schools.

The echo from an individual fish contains components from multiple reflections, including those from the swim bladder and other organs (Foote, 1980; Reeder *et al.*, 2004). Previous studies have shown that the temporal structure of the echo from anesthetized fish is highly variable, and relies on both the distance and the amplitude of the reflecting points (Barr, 2001). The temporal and spectral structures were useful cues for discrimination of fish species by using the dolphin sonar signal (Au and Benoit-Bird, 2003; Benoit-Bird *et al.*, 2003a, 2003b; Au and Benoit-Bird, 2008). Also, it was clarified that the echo duration was changed dependent on the fish orientation, i.e., the time between first and last arrivals of the processed echo (Reeder *et al.*, 2004; Stanton *et al.*, 2003). Therefore, it is possible that the duration of the echo from the fish at the dorsal aspect is an important cue for species identification as it is size dependent. However, fish movement affects both the echo temporal structure and the duration, and must therefore be taken into account.

The current paper analyzed echoes from fish in seas close to Japan using the sonar signal of the bottlenose dolphin. First, anesthetized fish in a water tank were used to determine whether the temporal structure and echo duration were dependent on the fish species and their orientation. Sec-

[a)]Author to whom correspondence should be addressed. Electronic mail: matsuo@cs.tohoku-gakuin.ac.jp

 0001-4966/2009/126(1)/444/7/$25.00

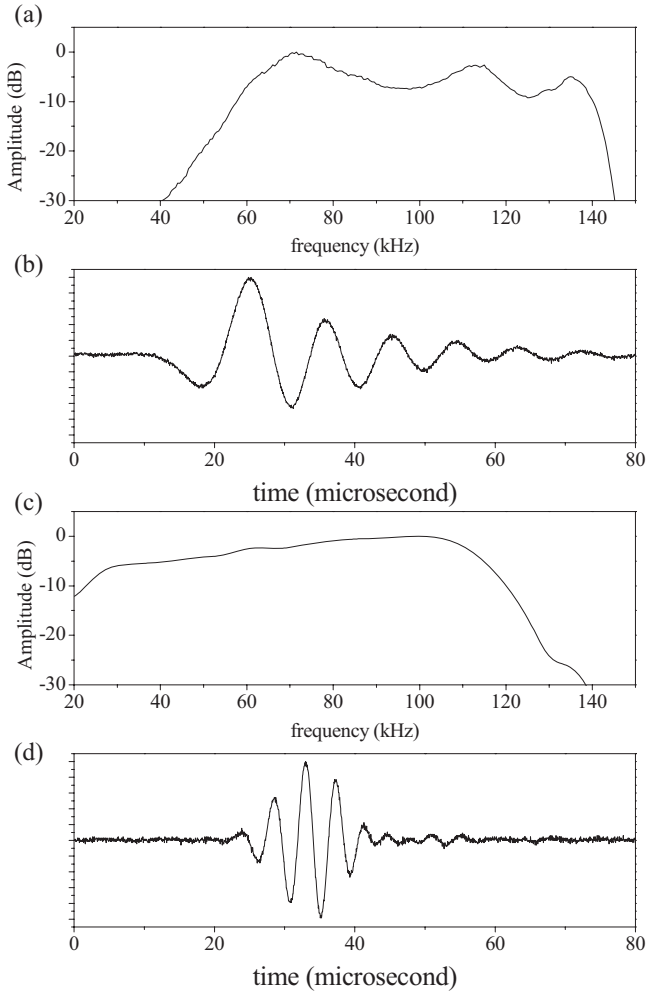


FIG. 1. System response and dolphin sound. (a) Frequency characteristics of both the transmit and receive transducers. (b) Original waveforms of the dolphin sonar sounds. (c) Amplitude spectra of the dolphin sound. (d) Waveform of incident signal, obtained by setting the transducers dace to face.

ond, it was examined whether these echo characteristics were dependent on fish movement, as measured in fish suspended by nylon monofilament lines in the open sea.

II. METHODS

The acoustic data analyzed in this paper were collected by Imaizumi *et al.* (2008), who estimated the TS spectra of fish using broadband signals. In the transmitting system, the signal from a signal generator (NF Co., WF1946) was amplified by a power amplifier (Accuphase, PRO-100) and sent to the custom-made transmit transducer (Furuno). In the receiving system, the reflected wave was sensed by the custom-made receive transducer (Furuno), and the signal was amplified by the amplifier. The output signals were observed, measured, and transformed into digital data by an oscilloscope (Tektronix, TDS3000). The transmit and receive transducers were similar but were separated. This enabled an easy measurement of the overall characteristics of the transmitting and receiving systems; the transducers were simply placed face to face. As shown in Fig. 1(a), the combined sensitivity of the transmitting and receiving system had broadband characteristics from 58 to 140 kHz at a level of −10 dB, and the

TABLE I. Fork length, body height, duration, and maximum value for fish species.

	Average fork length (mm)	Average body height (mm)	Average duration (μs)	Average maximum value (dB)
Horse mackerel	207.6(9.1)	50.4(2.8)	120.3(8.1)	−4.3(2.6)
Sea bream	294.3(9.5)	120.2(11.7)	168.4(31.2)	−4.3(3.7)
Mackerel	256.0(10.1)	64.8(3.4)	127.6(18.3)	−2.5(2.6)

combined beam width was 19.5° at 50 kHz and 7.1° at 130 kHz (Imaizumi *et al.*, 2006, 2008). For the calibration of this system, the echo was also measured from a tungsten carbide (TC) sphere and a copper sphere, which are normally used for the calibration of scientific echo sounders. It was examined that the measured TS almost corresponded almost to the theoretical TS, as shown in Fig. 5 of Imaizumi *et al.* (2008).

In the water tank, echoes were measured using anesthetized fish from the following three species, all of which had swim bladders (Imaizumi *et al.*, 2008): horse mackerel (n =10), sea bream (*Pagrus major*; n=6), and mackerel (*Scomber japonicus*; n=5). Table I shows the mean and standard deviation (SD) of the fish body sizes. The SDs were small, as the fish from each species were similar in size. In addition, the body height of the fish was dependent on the fork length and fish species.

Figure 2(a) depicts the measuring system. The experimental fish were anesthetized and suspended by nylon monofilaments in a freshwater tank (length × width × depth =4.8 × 3.8 × 2.8 m^3). Their orientation was adjusted so that the transmitted wave was incident on their dorsal regions. The echo from sea bream was also measured at different tilt angles [Fig. 2(b)], ranging from −90° to 90° at 10° increments. The definition of the tilt angle was 0 when the sound wave was perpendicular to the body axis of the fish, with θ =+90° for the fish head-on, and θ=−90° for the fish tail-on.

Within the sea, echoes from three similarly sized tethered live fish, which comprised one horse mackerel and two chicken grunt (*Parapristipoma trilineatum*), were measured from the training vessel *Hiyodori* of the Tokyo University of Marine Science and Technology at an anchor depth of 30 m. Table II shows the body sizes of these fish. Figure 2(c) de-

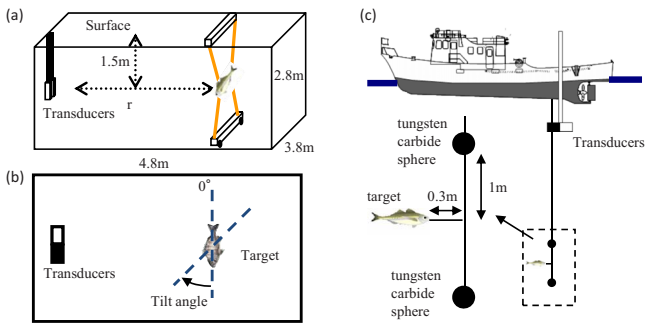


FIG. 2. (Color online) Diagram of measuring system. (a) Measurement of echo from dorsal aspect of fish. (b) Rotation of dorsal tilt angle. (c) Measurement of echo from tethered live fish.

TABLE II. Fork length, body height, and duration for three fish.

	Fork length (mm)	Body height (mm)	Average duration (μs)
Horse mackerel	175	43	105.7(10.2)
Chicken grunt	145	45	96.4(5.5)
Chicken grunt	152	51	100.9(6.4)

picts the experimental setup required to suspend a fish with two 38.1-mm-diameter spheres. Before the experiment, the fish were kept alive in a small tank onboard the deck of the vessel. The fish were then tethered to a nylon monofilament and returned to the sea as soon as possible. At a depth of 17.3 m of the upper sphere, the beam spreads were 4.7 m at 50 kHz and 1.9 m at 130 kHz. The sphere and the fish were near to the center of the beam, as they were suspended almost vertically beneath the transducers. The angles from the transducers to the two spheres and the fish were similar, as the distance from the main suspension line to the center of the fish body was about 0.4 m. The suspended fish moved within the beam. The dolphin-like sounds were transmitted with a pulse-repetition period of 0.1 s.

Dolphin sonar signals are composed of trains of clicks, the time interval (click interval) of which is generally dependent on the target range (Au, 1993). In the current paper, we used the actual sound produced by the bottlenose dolphin (Nakamura and Akamatsu, 1998). As shown in Fig. 1(b), this sound had a broadband frequency bandwidth and was composed of a small number of cycles. As shown in Fig. 1(c), the sonar sound spectrum had a peak near 100 kHz, and a −3 dB bandwidth that was 60 kHz wide (55–115 kHz). Figure 1(d) shows the incident waveforms obtained by setting the transducers face to face with 2 m separation range.

A. Data analysis

The dotted curve in Fig. 3(a) shows the echo waveform of a horse mackerel. To estimate its temporal structure, noise reduction was carried out and the echo-envelope pattern was extracted. As shown by the solid curve of Fig. 3(a), this envelope pattern was calculated by the cross-correlation function between the incident wave and echo waveform and the Hilbert transform (Au and Benoit-Bird 2003, 2008; Oppenheim and Schafer, 1975; Stanton and Chu, 2008). In the case of the water-tank experiment, the delay time (which was equivalent to the time, t) of the envelope pattern was represented by setting the maximum value as $t=0$ in order to estimate the temporal features that were dependent on the fish species. In the case of the sea experiment, the delay time of the envelope pattern was represented by setting the maximum value of the echo from a shallower TC sphere as $t=0$.

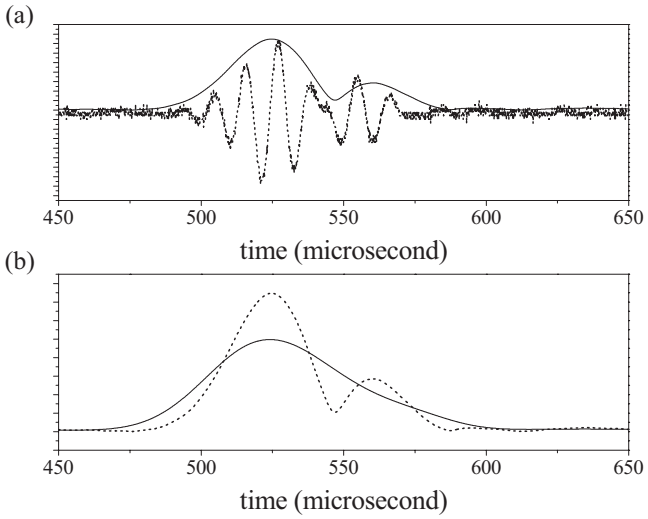


FIG. 3. Echo-analysis method. (a) Echo waveform (dotted curve) and envelope pattern (solid curve). (b) Low-passed pattern (solid curve) and envelope pattern (broken curve).

The temporal highlight structure was computed by extracting the local peak from the envelope pattern. In addition, the duration was estimated using this envelope pattern, and by defining the onset and offset of the echo from one fish. Because of the difficulty in automatically doing this, resulting from multiple temporal highlights, the envelope pattern was transformed into a low-passed pattern, including only one local peak, by convolution of the envelope pattern using the Hanning window, which has a width of 80 μs. In Fig. 3(b), the solid curve represents the low-passed pattern and the broken curve represents the envelope pattern. The duration of the echo waveform was computed by the 20 dB width of the low-passed pattern, that is, using the threshold that is 0.1 times the maximum value of this pattern.

III. RESULTS

A. Anesthetized fish in the water tank

Echoes were measured from the dorsal aspect of the fish of three species. Figure 4 shows the echo waveforms and envelope patterns computed using the cross-correlation function and the Hilbert transform. The envelope-pattern shapes varied according to both the species and the individual, and included one or more highlights. To estimate the echo duration, the low-passed pattern was computed and was almost shown to include only one local peak (Fig. 5). Therefore, the duration of each echo could be estimated within 20 dB of the maximum value. Figure 6 shows the relationship between the fish body height and the number of highlights, the maximum value of the envelope pattern, and the duration of the echo. The number of highlights and the echo duration varied with the body height, whereas the maximum value of the envelope pattern did not (Table I). Significant differences in duration were observed between the horse mackerel and sea bream, and between the mackerel and sea bream by using student's t-test ($p<0.05$). The duration could therefore be used to discriminate between fish species with differing body heights when the signal level of the echo from the fish was 20 dB higher than noise level.

Figure 7 shows the echo waveform and envelope patterns measured at different tilt aspects from the sea bream. At a 0° tilt angle, the longitudinal axis of the fish was perpendicular to the direction of the sonar signal. These time structures varied with the tilt angle, such that the lower the maximum value of the envelope pattern, the longer the duration of the pattern. As shown in Fig. 8, the maximum, echo duration,

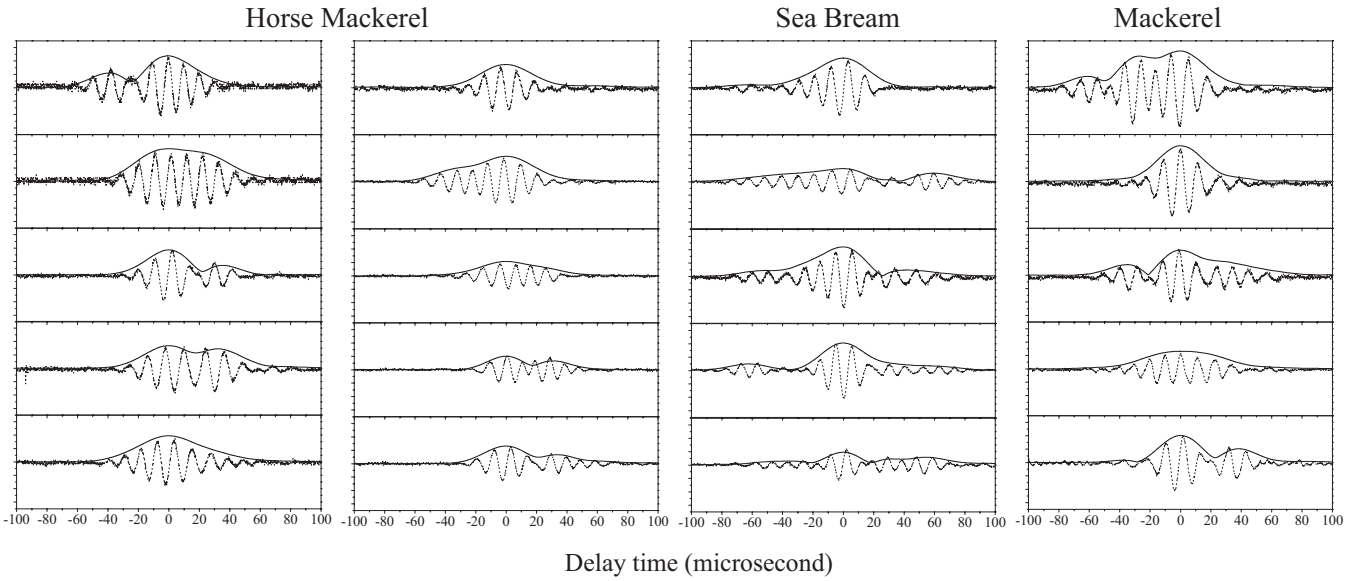


FIG. 4. Echo waveforms and envelope patterns. Echoes were measured from the dorsal aspect. Envelope patterns were computed by the cross-correlation function between the incident waveform and echo waveform and the Hilbert transform.

and the number of highlights varied with the tilt angle; however, the number of highlights showed only minor variation within the tilt angle range of −30° (tail-on) to 30° (head-on). The highest envelope pattern and the shortest duration occurred at a −40° tilt angle. In addition, the dotted line in Fig. 8(c) shows the TS at 100 kHz, which was shown in Fig. 9 of Imaizumi *et al.* (2008). The results for the maximum values and the TS at 100 kHz showed similar trends concerning the fish tilt.

B. Moving fish in the sea

Echoes from the tethered fish were continuously measured, and the envelope pattern and the low-passed pattern were analyzed from these waveforms using the same procedure. The duration was computed from these echoes which have a high enough signal-to-noise ratio (SNR). Figure 9(a) shows the plot of the ping number versus the delay time of the envelope-pattern maximum value. The delay time corresponded to the distance. The radius was dependent on the echo magnitude, which corresponded to the envelope-pattern maximum value. Figure 9(b) shows the plot of the ping number versus the duration, which could be only estimated from echoes with the high SNR. It was shown that the estimated duration was changed within the range of 87–119 μs by the fish movement. Table II shows the average and SD of the duration. It was examined that this average duration was almost similar to each another in the case of measuring the echoes from similar size fish.

IV. DISCUSSION AND CONCLUSIONS

This study estimated the temporal highlight structures and the echo duration for the identification of fish species.

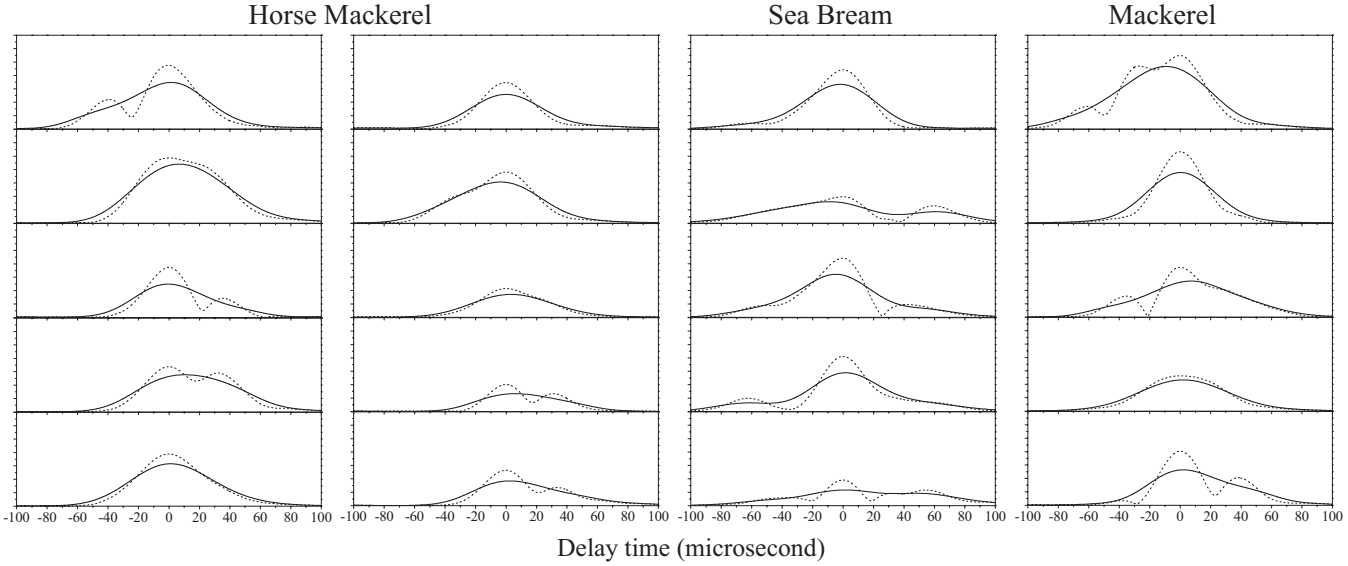


FIG. 5. Envelope patterns and low-passed patterns. Low-passed patterns were computed by convolving the envelope patterns with the 80 μs Hanning window.

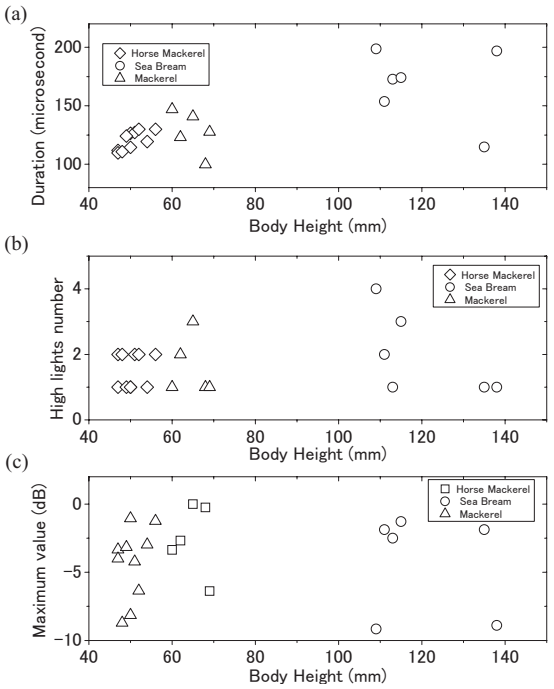


FIG. 6. Characteristics of echo temporal structure. (a) Duration with 20 dB width of low-passed pattern. (b) Numbers of envelope-pattern highlights. (c) Maximum values of envelope pattern. The normalized maximum amplitude is shown.

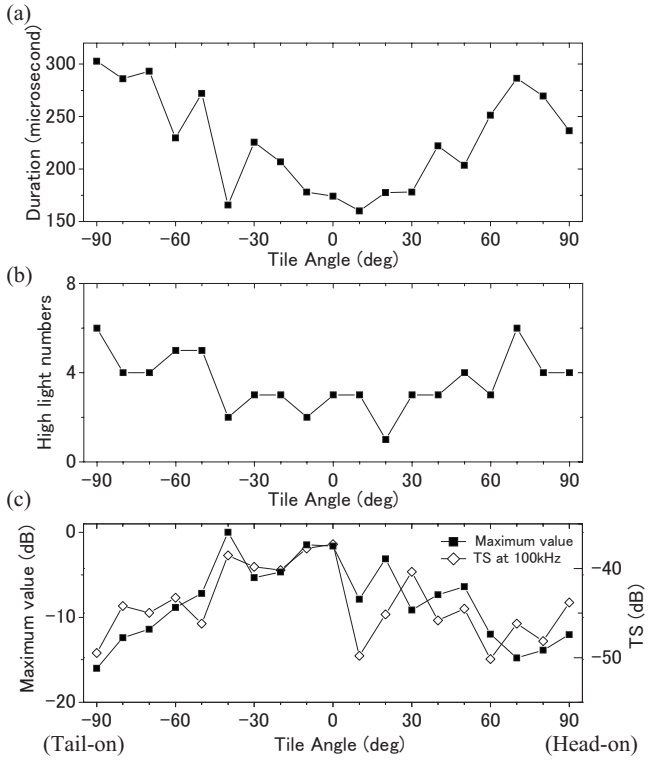


FIG. 8. Characteristics of echo temporal structure measured from anesthetized fish at different tilt angles. (a) Duration. (b) Numbers of envelope-pattern highlights. (c) Maximum values of the envelope pattern and the TS at 100 kHz. The normalized maximum amplitude at each tilt angle is shown.

Fish size was also of importance in discriminating between species and affected the TS (Foote, 1980), but was not related to the maximum value of the envelope pattern (Fig. 6). We found that echo duration could be derived using the low-passed filter, and that it was possible to estimate the echo duration of individual fish from the school using this method. As shown in Fig. 5 and Table I, this duration was dependent on the body heights and the fish species, as the longitudinal axis of the fish was perpendicular to the direction of the sonar signal. At the dorsal aspect, the echo was mainly influenced by the swim bladder (Foote, 1980).

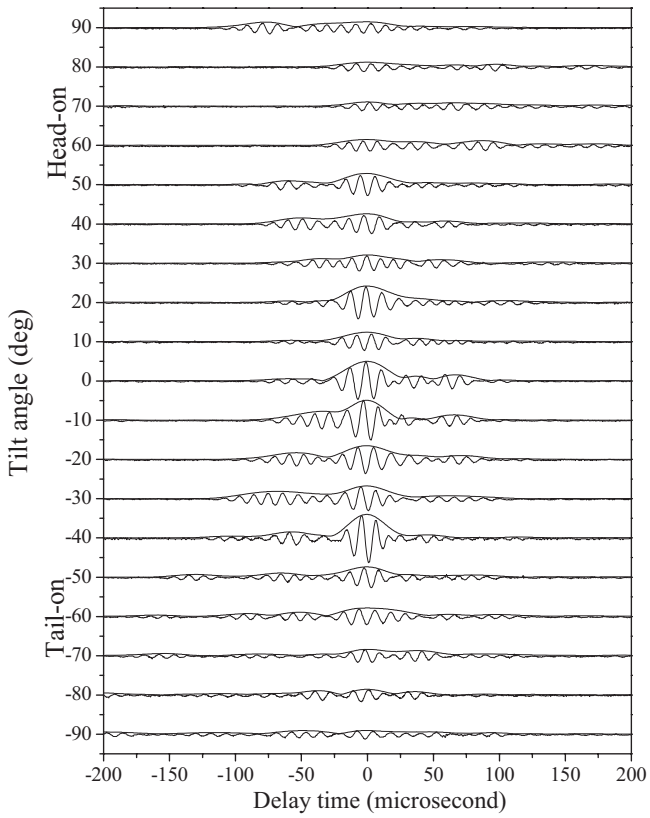


FIG. 7. Echo waveforms and envelope patterns at different tilt angles.

Figure 10 shows the changes of the estimated echo duration that were dependent on different thresholds. When the threshold was above 0.1 (20 dB), the echo duration was not found to be dependent on the fish species. Thus, when the threshold was more than 0.1 times the maximum value, the duration could be estimated based on the swim bladder. The identification of fish species required the estimation of the duration within a 20 dB (0.1 times) width. The accurate estimation of the duration requires the improvement of the SNR and therefore the cross-correlation function was used in this paper. However, it was difficult to determine the duration in the sea experiment because of not enough SNR. The chirp signal was useful for measurements at high SNRs (Reeder *et al.*, 2004; Stanton *et al.*, 2003). Therefore, it would be necessary to compare the ability and performance of various objects and conditions by using both the dolphin sonar signal and chirp signal.

It has previously been verified that the strongest echo occurs when the incident signal is perpendicular to the long axis of the swim bladder, that the number of echo highlights and the duration increases as fish tilt from this orientation (Au and Benoit-Bird, 2003; Reeder *et al.*, 2004; Stanton *et al.*, 2003), and that the longitudinal axis of the fish differs

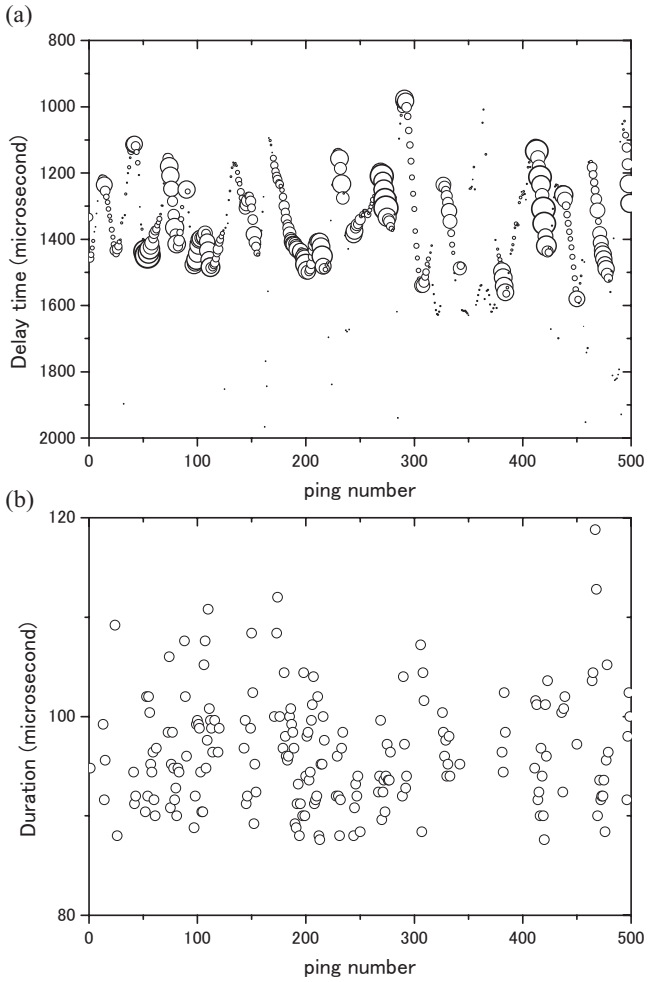


FIG. 9. Output of moving fish, chicken grunt. (a) Highlight structures plot. Amplitude of each highlight represented by circle radius. (b) Duration plot. This repetition period is 0.1 s.

from the axis of the swim bladder (Benoit-Bird *et al.*, 2003a; Reeder *et al.*, 2004; Stanton *et al.*, 2003). It is therefore likely that the envelope pattern has the highest amplitude and shortest duration at a −40° tilt angle. Further study will be necessary to compare the temporal highlight structures with the location and shape of the swim bladder and other organs.

It was examined that the echo duration was dependent on the body height at the dorsal aspect (Fig. 6). In the sea experiment, we analyzed echoes reflected from moving fish. The echo duration with a 20 dB width largely corresponded to the duration measured from the dorsal aspect in the water-tank experiment, as shown in Fig. 6 and Table II. These findings verified that the echo duration provided useful information for identifying the species with differing body heights regardless of whether or not the fish were moving.

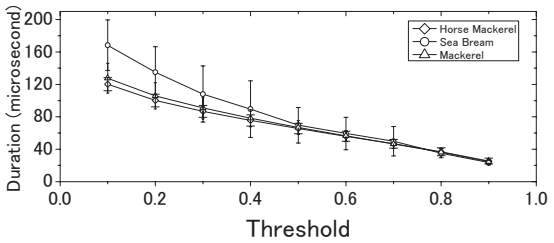


FIG. 10. Duration versus threshold value. The durations were calculated using values of 0.1–0.9 times the maximum of the low-passed pattern at each echo. The average duration for each threshold and species was calculated by averaging these values.

In order to estimate fish abundance, it is necessary to identify the species and to calculate the number of fish based on the echoes produced by schools. The proposed method can estimate temporal structures including individual echoes by analyzing these echoes. In future studies, we will explore the possibility of improving the proposed method by applying a split beam approach (see, for example, Simmonds and MacLennan, 2005), which might allow both the three-dimensional location and the species to be estimated from the echoes produced by schools of fish.

ACKNOWLEDGMENTS

This work was supported by the Research and Development Program for New Bio-industry Initiatives of the Bio-oriented Technology Research Advancement Institution, Japan. We are grateful to Dr. Whitlow W. L. Au and two anonymous referees for their useful comments and advice.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).
Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Acoustic backscattering by Hawaiian lutjanid snappers. II. Broadband temporal and spectral structure," J. Acoust. Soc. Am. **114**, 2767–2774.
Au, W. W. L., and Benoit-Bird, K. J. (**2008**). "Broadband backscatter from individual Hawaiian mesopelagic boundary community with implications for spinner dolphin foraging," J. Acoust. Soc. Am. **123**, 2884–2894.
Barr, R. (**2001**). "A design study of an acoustic system suitable for differentiating between orange roughy and other New Zealand deep-water species," J. Acoust. Soc. Am. **109**, 164–178.
Benoit-Bird, K. J., Au, W. W. L., and Kelley, C. D. (**2003a**). "Acoustic backscattering by Hawaiian lutjanid snappers. I. Target strength and swimbladder characteristics," J. Acoust. Soc. Am. **114**, 2757–2766.
Benoit-Bird, K. J., Au, W. W. L., Kelley, C. D., and Taylor, C. (**2003b**). "Acoustic backscattering by deepwater fish measured in situ from a manned submersible," Deep-Sea Res., Part I **50**, 221–229.
Foote, K. G. (**1980**). "Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackerel target strength," J. Acoust. Soc. Am. **67**, 2084–2089.
Foote, K. G. (**1987**). "Fish target strengths for use in echo integrator surveys," J. Acoust. Soc. Am. **82**, 981–987.
Hammer, Jr. C. E., and Au, W. W. L. (**1980**). "Porpoise echo-recognition: An analysis of controlling target characteristics," J. Acoust. Soc. Am. **68**, 1285–1293.
Harley, H. E., Putman, E. A., and Roltblat, H. L. (**2003**). "Bottlenose dolphins perceive object features through echolocation," Nature (London) **424**, 667–669.
Imaizumi, T., Furusawa, M., and Akamatsu, T. (**2006**). "Measurement of the frequency characteristics of the scattering amplitude using dolphin's sonar signal," J. Marine Acoust. Soc. Jpn. **33**, 143–150.
Imaizumi, T., Furusawa, M., Akamatsu, T., and Nishimori, Y. (**2008**). "Measuring the target strength spectrum of fish using sonar signals of dolphins," J. Acoust. Soc. Am. **124**, 3440–3449.
MacLennan, D. N. (**1990**). "Acoustical measurement of fish abundance," J. Acoust. Soc. Am. **87**, 1–15.
Miyanohana, Y., Ishii, K., and Furusawa, M. (**1990**). "Measurements and analyses of dorsal-aspect target strength of six species of fish at four frequencies," Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **189**, 317–324.
Nakamura, K., and Akamatsu, T. (**1998**). "Comparison of echolocation signal among dolphins and porpoises," Trans. Tech. Comm. Psychol. Physiol. Acoust. Soc. Jpn. **H-98**, 106 (in Japanese).
Nakken, O., and Olsen, K. (**1977**). "Target strength measurements of fish," Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **170**, 52–69.
Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Reeder, D. B., Jech, J. M., and Stanton, T. K. (**2004**). "Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling," J. Acoust. Soc. Am. **116**, 747–761.
Simmonds, J., and MacLennan, D. (**2005**). *Fisheries Acoustics* (Blackwell, Oxford).
Simmonds, E. J., Armstrong, F., and Copland, P. J. (**1996**). "Species identification using broadband backscatter with neural network and discriminant analysis," ICES J. Mar. Sci. **53**, 189–195.
Simmonds, E. J., Williamson, N. J., Gelotto, F., and Aglen, A. (**1992**). "Acoustic survey design and analysis procedure: A comprehensive review of current practice," ICES Cooperative Research Report No. 187 (International Council for the Exploration of the Sea, Denmark, 1992).
Stanton, T. K., and Chu, D. (**2008**). "Calibration of broadband active acoustic systems using a single standard spherical target," J. Acoust. Soc. Am. **124**, 128–136.
Stanton, T. K., Reeder, D. B., and Jech, J. M. (**2003**). "Inferring fish orientation from broadband acoustic echoes," ICES J. Mar. Sci. **60**, 524–531.
Zakharia, M. E., Magand, F., Hetroit, F., and Dener, N. (**1996**). "Broadband sounder for fish species identification at sea," ICES J. Mar. Sci. **53**, 203–208.

A versatile pitch tracking algorithm: From human speech to killer whale vocalizations

Ari Daniel Shapiro[a)]
Department of Biology, Woods Hole Oceanographic Institution, MS 50, Woods Hole, Massachusetts 02543

Chao Wang[b)]
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, Massachusetts 02139

(Received 1 November 2008; revised 6 April 2009; accepted 17 April 2009)

In this article, a pitch tracking algorithm [named discrete logarithmic Fourier transformation-pitch detection algorithm (DLFT-PDA)], originally designed for human telephone speech, was modified for killer whale vocalizations. The multiple frequency components of some of these vocalizations demand a spectral (rather than temporal) approach to pitch tracking. The DLFT-PDA algorithm derives reliable estimations of pitch and the temporal change of pitch from the harmonic structure of the vocal signal. Scores from both estimations are combined in a dynamic programming search to find a smooth pitch track. The algorithm is capable of tracking killer whale calls that contain simultaneous low and high frequency components and compares favorably across most signal to noise ratio ranges to the peak-picking and sidewinder algorithms that have been used for tracking killer whale vocalizations previously.
 [DOI: 10.1121/1.3132525]

PACS number(s): 43.80.Ka, 43.72.Ar [WA] Pages: 451–459

I. INTRODUCTION

Robust pitch detection is a crucial first step in the analysis and modeling of human speech. The fundamental frequency (f_0) plays an important role in modeling linguistic attributes including lexical stress, tone, and intonation, as well as paralinguistic attributes such as emotion. However, it is difficult to build reliable statistical models involving f_0 because of pitch estimation errors and the discontinuity of the f_0 contour. Specifically, inaccurate voiced pitch hypotheses and erroneous voiced/unvoiced (V/UV) decisions can lead to noisy and undependable feature measurements. This is especially true for telephone speech due to inferior pitch detection performance caused by the noisy and band-limited telephone channel.

Previously, a pitch detection algorithm (PDA) was developed utilizing the discrete logarithmic Fourier transformation (DLFT) of the speech signal (Wang and Seneff, 2000b). This algorithm, which will be referred to as DLFT-PDA, is based on a robust pitch estimation method known as harmonic matching (Hess, 1983). Reliable estimates of both pitch and the temporal change of pitch are derived based on harmonic matching principles, which are then combined in a dynamic programming (DP) search to find a globally optimal solution. The DP search tracks pitch continuously, avoiding the propagation of V/UV decision errors to voiced pitch hypotheses. Evaluation results have demonstrated that the algorithm is particularly suitable for telephone speech and prosodic modeling applications (Wang and Seneff, 2000a, 2000b; Wang, 2001; Wang and Seneff, 2001a, 2001b).

Pitch tracking is also important for analyzing and quantifying features of the vocalizations of marine mammals. Several different manual and automatic approaches have been implemented previously. A labor-intensive but fairly reliable method is to trace the f_0 on the spectrogram by hand using a digital interface (Watwood *et al.*, 2004, 2005; Shapiro, 2006). Because the number and placement of (f_0, *time*) points will vary between contours, a subsequent interpolation is used to hold this number constant and represent all contours with a uniform number of evenly spaced points. Another commonly used automated method is to select the peak frequency value from a sliding power spectrum of the signal (i.e., peak-picking), followed by subsequent manual correction to remove pitch doubling and halving errors (Buck and Tyack, 1993; Janik *et al.*, 1994; McCowan, 1995). Often the signal is band-pass filtered over an appropriate frequency range before the frequency associated with the peak spectral energy is selected. Another automated technique, the sidewinder algorithm, is similar to the spectral autocorrelation method for tracking human speech (Lahat *et al.*, 1987). The algorithm computes an autocovariance sequence for each spectral slice [in contrast with human speech, no spectral flattening is necessary for killer whale (*Orcinus orca*) vocalizations] whose peaks occur at multiples of the spacing of the frequency bands (Deecke *et al.*, 1999). Two pitch extraction methods have been implemented for this technique: one searches for the second highest peak directly in the autocovariance sequence itself, while the other computes the real cepstrum of the autocovariance sequence to locate the highest peak. The fast Fourier transform FFT based comb-filter method described in Brown *et al.* (2006) applies the

[a)]Author to whom correspondence should be addressed. Electronic mail: ashapiro@whoi.edu
[b)]Present address: Vlingo Corporation, 17 Dunster Street, Cambridge, MA 02138.

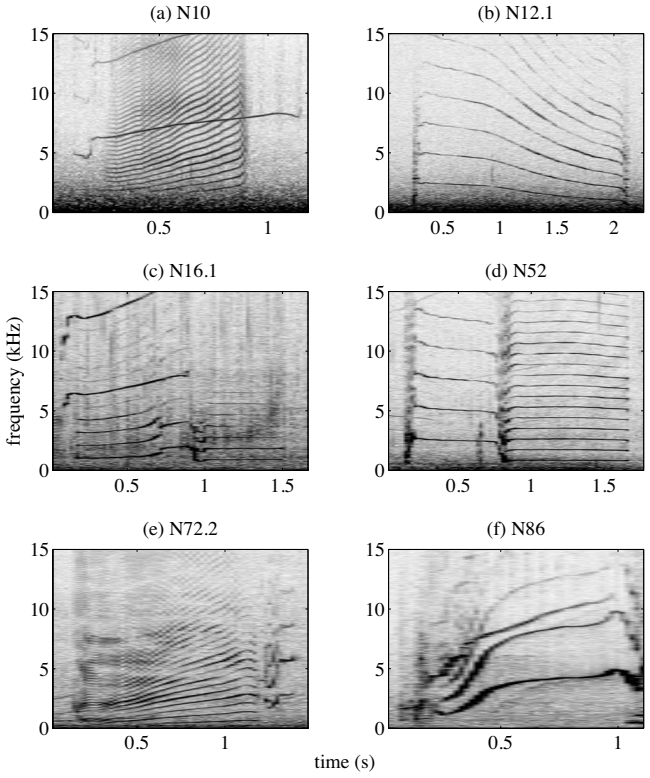


FIG. 1. Spectrograms of various killer whale calls, labeled by call type, and generated using a 2048-point FFT with 50% overlap and a Hamming window.

spectral comb method to tracking killer whale calls. It is the most similar in spirit to the DLFT-PDA algorithm.

Killer whale pods produce a set of stereotyped, harmonically-structured calls that often consist of multiple temporal and spectral components (Ford, 1987, 1989, 1991; Fig. 1). Individual killer whales tend to match the call types produced by other group members (Miller *et al.*, 2004). This kind of communication may facilitate group cohesion (Miller, 2002) and/or allow individuals to discriminate between one another (Miller *et al.*, 2007). Killer whales produce these vocalizations by varying the pulse repetition rate, which corresponds to the relative spacing between different frequency bands spectrographically (Watkins, 1967).

Each killer whale call type generally consists of a modulated low frequency component (LFC) with an f_0 typically ranging between 80 and 2400 Hz (Ford, 1987). Some call types also contain a high frequency component (HFC) [often with an f_0 between 2 and 12 kHz (Hoelzel and Osborne, 1986)] that is synchronously produced with the LFC but separately modulated to produce two sets of unique harmonics. The challenge of tracking the pitch of both of these components simultaneously is similar to that encountered by multi-pitch estimation of mixtures of music or speech signals (see Klapuri, 2008; Klapuri and Virtanen, 2008). In these scenarios, it is necessary to isolate and then track each constituent component of the additive signal. Automatic pitch tracking of killer whale calls would greatly facilitate the study and characterization of their stereotypy, cultural transmission (see Deecke *et al.*, 1999), and individual variability (see Miller and Bain, 2000; Nousek *et al.*, 2006).

Because multiple spectral components can be embedded within a single killer whale vocalization, a time domain representation of the signal would lead to a loss of the harmonic structure. For accurate pitch tracking, a frequency solution is required. This article examines the application of the DLFT-PDA to determine the fundamental frequencies of both the LFC and HFC of Norwegian killer whale stereotyped calls (see Nousek *et al.*, 2006, Miller *et al.*, 2007, for an application of this method). The DLFT-PDA has been designed especially for telephone speech, where the f_0 is often weak or missing and the signal to noise ratio (SNR) is usually low compared to microphone-recorded speech. Coincidentally, the recordings of marine mammal vocalizations are often characterized by the same features because of boat noise and substantial distances between the vocalizing animals and recording equipment. Several characteristics of our algorithm render it especially well suited for tracking the pitch of killer whale calls. First, the algorithm relies on the harmonic structure (i.e., spectral peaks at multiples of the f_0) to estimate pitch and deliberately ignores low-frequency spectrum, which makes it robust to interference from low-frequency boat noise. Second, the DLFT can be tuned to a sub-band in the spectrum, allowing the algorithm to track calls with simultaneous LFC and HFC that are somewhat separated in the frequency space.

In Sec. II, an overview and then the specifics of our algorithm are given (see Wang and Seneff, 2000b; Wang, 2001 for the full details), highlighting features that make it suitable for telephone speech and killer whale recordings. Adaptations of the algorithm are then described for killer whale calls. Finally, evaluation results of our algorithm are presented.

II. METHODS

A. Overview of algorithm

The DLFT-PDA is based on the observation that harmonic peaks will be spaced by a constant distance on a logarithmic frequency scale regardless of f_0. More formally, if a signal has harmonic peaks spaced by f_0, then on a logarithmic scale the peaks will occur at $\log f_0, \log f_0 + \log 2$, $\log f_0 + \log 3, \ldots$, etc. The fundamental frequency determines the position of the first peak and the subsequent harmonic peaks are at fixed distances from the first peak. Thus, harmonic spectra with different fundamental frequencies can be aligned by simple linear shifting. By correlating a spectrum sampled on the logarithmic frequency scale with a harmonic template (a logarithmic spectrum of an impulse train), a robust estimation of the $\log f_0$ of the signal can be obtained. The correlation of two logarithmic spectra from adjacent frames of a vocal signal leads to a very reliable estimation of the change in $\log f_0$ ($\Delta \log f_0$).

Instead of determining an f_0 value for each frame by picking the correlation maximum, a DP search is used to combine the $\log f_0$ and $\Delta \log f_0$ estimations to find an optimal solution overall. All values (quantized in the search space) are considered as possible f_0 candidates with different qualities. The quality of a pitch candidate P is indicated by the correlation between the spectrum and the template (Fig.

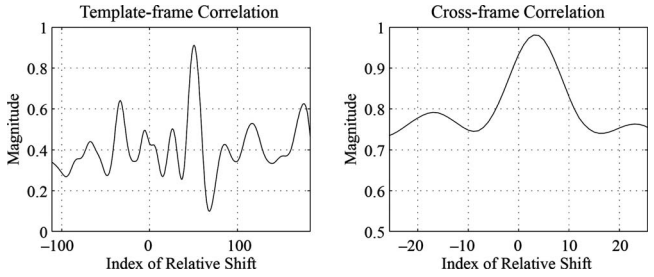


FIG. 2. Examples of "template-frame" and "cross-frame" correlations for DLFT spectrum.

2). The "consistency" of two consecutive pitch candidates is indicated by the correlation of the spectra of the adjacent frames at the position corresponding to the difference between the pitch candidates. These two constraints are used to define a score function for the DP search. The DP search algorithm solves the optimization problem iteratively (i.e., finding the optimal score at time t is achieved by finding the optimal score at time $t-1$). The path in the quantized (*frequency*, *time*) space with the highest score yields the optimum pitch track.

The algorithm requires defining a small set of parameters: the window size, the frequency range $[f_s,f_e]$ for the DLFT, and the f_0 search range and resolution. The pseudo-code of our pitch tracking algorithm is shown in Fig. 3 and the implementation details are discussed in Wang and Seneff, 2000b and Wang, 2001.

B. Details of algorithm

1. Signal representation

To obtain a logarithmically spaced spectrum for the frequency region $[f_s,f_e]$, the discrete-time Fourier transform is directly sampled at linear intervals on the logarithmic frequency scale. This representation is defined as a DLFT. Assuming $x_t(n)$ is a Hamming-windowed audio signal centered at time t ($n=0,1,\ldots,N-1$, where N is the window size), the DLFT is computed as follows:

N: the total number of frames in an input waveform
M: the total number of quantized pitch candidates
P_i: the quantized pitch candidates ($i = 0, \ldots, M-1$)
T: the harmonic template
X_t: the logarithmic-frequency spectrum at the tth frame of the input
$S(t, i)$: the path score for the i^{th} pitch candidate at the t^{th} frame
begin
  compute T
  compute X_0
  compute the correlation of X_0 and T
  initialize $S(0, i)$ for all P_i ($i = 0, \ldots, M-1$)
  for $t = 1, \ldots, N-1$
    compute X_t
    compute the correlation between X_t and $X_t - 1$
    compute the correlation between X_t and T
    update the partial path score $S(t, i)$ and
    save the back trace pointer for all P_i ($i = 0, \ldots, M-1$)
  end
  back trace to find the best pitch contour $P(t)$ ($t = 0, \ldots, N-1$)
end

FIG. 3. Pseudo-code of the DLFT-PDA.

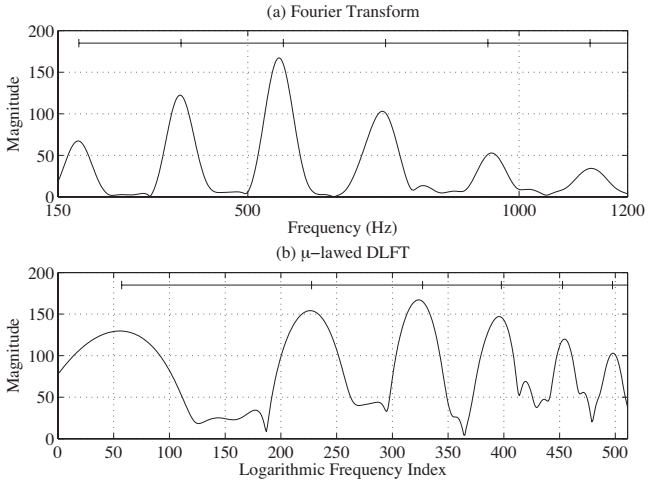


FIG. 4. (a) Fourier transform and (b) μ-law-compressed DLFT for a speech signal. Positions of harmonic peaks are indicated by ruler tick marks in the plots.

$$X_t(i) = \frac{1}{N}\sum_{n=0}^{N-1} x_t(n)e^{-j\omega_i n} \quad (i = 0,1,\ldots,N-1), \tag{1}$$

$$\omega_i = 2\pi e^{(\log f_s + i\cdot d\,\log f)} \cdot T_s, \tag{2}$$

$$d\log f = (\log f_e - \log f_s)/(N-1), \tag{3}$$

where T_s is the sampling period of the waveform. The term $d\log f$ can be viewed as the frequency resolution in the logarithmic domain.

The spectrum is normalized by a μ-law conversion to reduce the dynamic range of harmonic peak height due to formant influences:

$$X_t(i) = M_t \cdot \log(1+\mu\cdot X_t(i)/M_t)/\log(1+\mu)$$

$$(i = 0,1,\ldots,N-1), \tag{4}$$

where M_t is the maximum energy of the DLFT spectrum at the tth frame:

$$M_t = \max_i X_t(i). \tag{5}$$

The value $\mu=50$ was chosen in our implementation. The conversion holds the maximum value unchanged while promoting smaller values. Figure 4 shows the discrete Fourier transform (DFT) and μ-law-compressed DLFT for a Hamming-windowed voiced speech signal. Notice the dynamic range of the harmonic peaks in the DLFT spectrum is compressed due to the μ-law conversion.

2. Harmonic template

The harmonic template is constructed from an ideal periodic signal. The pulse train is first Hamming windowed, after which the DLFT spectrum is computed for the frequency range $[f'_s,f'_e]$. The parameters f'_s and f'_e can differ from those used for computing the DLFT of the signal. However, the equality $f'_e/f'_s=f_e/f_s$ must be ensured, so that the frequency resolution $d\log f$ [see Eq. (3)] of the signal and template's DLFT spectra match for the correlation operation.

TABLE I. Parameter settings for tracking killer whale calls. M is the number of pitch hypotheses in the search space. See text for details.

	Window	$[f_s, f_e]$	$[P_{\text{low}}, P_{\text{high}}]$	M
I	25 ms	[250,1750] Hz	[100,600] Hz	200
II	15 ms	[800,5600] Hz	[400,4000] Hz	500
III	2.5 ms	[6k,24k] kHz	[4,12] kHz	500

In our implementation, the harmonic template includes five complete harmonic lobes. The parameters of the template had been tuned using development data to achieve good performance for killer whale vocalizations (see Table I).

The harmonic lobes are wider for the low-frequency region on the logarithmic scale, which leads to a bias in reaching a correlation maximum when the first lobe in the template is included in the calculation. This is problematic when the harmonic component at f_0 is absent (very weak or out of range) in the signal's DLFT spectrum. Ideally, the template is also expected to match the signal maximally without the first component (i.e., $2P_T$ matches $2f_0$, $3P_T$ matches $3f_0$, and so on). However, due to the strong first lobe in the template, the correlation is likely to reach a maximum by matching the P_T, $2P_T$, and $3P_T$ components of the template with the $2f_0$, $4f_0$, and $6f_0$ components of the signal spectrum. This will result in pitch doubling errors.

To suppress this tendency, the energy of each harmonic lobe in the template is normalized, similar to the measure taken in Hermes, 1988. This is done by integrating over each lobe to find its area, followed by a scaling by the reciprocal of the area, subject to an exponential decay to tune the effect. The decay factor was determined empirically from development data to be 0.85.

A second measure taken to discourage pitch doubling errors is to add negative lobes between the positive lobes in the template. If the P_T, $2P_T$, and $3P_T$ components of the template match with the $2f_0$, $4f_0$, and $6f_0$ components of the signal spectrum, then the negative lobes would match the $3f_0$ and $5f_0$ components and would reduce the magnitude of the correlation value. The negative lobes are obtained by computing the DLFT spectrum of the same pulse train with a frequency shift equivalent to half of its fundamental P_T:

$$\omega_i = 2\pi e^{(\log f_s' + i \cdot d \log f - \log P_T/2)} \cdot T_s \tag{6}$$

where $d \log f$ and T_s are the same as in Eq. (2). The shift $\log P_T/2$ causes the harmonic peaks in the new spectrum to fall precisely between those in the original one. The final harmonic template is constructed by combining the DLFT spectrum with a negatively weighted shifted DLFT spectrum. The weight for the negative lobes was determined empirically to be 0.35.

3. Two correlation functions

The normalized "template-frame" correlation function provides an estimation for $\log f_0$ by correlating the speech DLFT spectrum with the harmonic template, as shown in:

$$R_{TX_t}(n) = \frac{\Sigma_i T(i) X_t(i-n)}{\sqrt{\Sigma_i X_t(i)^2}} \quad (N_L < n < N_H). \tag{7}$$

The template, $T(i)$, is normalized to have unit energy in advance so the correlation is normalized by the signal energy only. The f_0 search range $[F_{\min}, F_{\max}]$ determines the bounds for the correlation, $[N_L, N_H]$.

The mapping between pitch candidate P and the corresponding index in the template-frame correlation function can be derived from Eq. (2). Assuming the index of the trial pitch P in the signal DLFT spectrum is i_P, according to Eq. (2):

$$\omega_{i_P} = 2\pi \cdot P \cdot T_s = 2\pi e^{(\log f_s + i_P \cdot d \log f)} \cdot T_s, \tag{8}$$

where f_s is the low-frequency bound for the signal DLFT spectrum and $d \log f$ is the logarithmic frequency resolution. The relationship of P and i_P can be further simplified as:

$$\log P = \log f_s + i_P \cdot d \log f, \tag{9}$$

$$i_P = (\log P - \log f_s)/d \log f. \tag{10}$$

Similarly, the index of the fundamental frequency (P_T) in the template, i_{P_T}, can be determined as:

$$i_{P_T} = (\log P_T - \log f_s')/d \log f \tag{11}$$

where f_s' is the low-frequency bound for the template.

The relative shift in the template-frame correlation to align the two harmonic structures is simply the difference of these two indices:

$$I_P = i_{P_T} - i_P = (\log P_T - \log f_s' - \log P + \log f_s)/d \log f. \tag{12}$$

Conversely, P can also be determined from the correlation lag I_P by:

$$P = \frac{P_T \cdot f_s}{f_s' \cdot e^{I_P \cdot d \log f}}. \tag{13}$$

By substituting P into Eq. (12) with the pitch range $[F_{\min}, F_{\max}]$, the bounds for template-frame correlation are obtained as:

$$N_L = (\log P_T - \log f_s' - \log F_{\max} + \log f_s)/d \log f, \tag{14}$$

$$N_H = (\log P_T - \log f_s' - \log F_{\min} + \log f_s)/d \log f. \tag{15}$$

By aligning two adjacent frames of the signal DLFT spectra, the normalized "cross-frame" correlation function provides constraints for $\Delta \log f_0$, as shown in:

$$R_{X_t X_{t-1}}(n) = \frac{\Sigma_i X_t(i) X_{t-1}(i-n)}{\sqrt{\Sigma_i X_t(i)^2}\sqrt{\Sigma_i X_{t-1}(i)^2}} \quad (|n| < N). \tag{16}$$

The correlation is normalized by the energy of both signal frames. Since f_0 should not change dramatically across two frames, the correlation bound N is set to be around 10% of the number of samples in the DLFT spectrum. A robust estimation of the $\log f_0$ difference across two voiced frames is given by the maximum of the correlation. See Fig. 2 for

examples of the template-frame and cross-frame correlation functions of a speech signal.

4. DP search

The advantage of using DP in pitch tracking is to incorporate continuity constraints across adjacent frames to reduce pitch doubling and halving errors (Secrest and Doddington, 1983; Talkin, 1995; Geoffrois, 1996; Droppo and Acero, 1998). This is typically achieved by incorporating a transition cost in the DP score function to penalize large changes in neighboring f_0 hypotheses. In our implementation, the transition cost is defined by the cross-frame correlation function [Eq. (16)]. It goes beyond enforcing continuity: it provides an estimation of the actual change in $\log f_0$. Given the score functions of $\log f_0$ and $\Delta \log f_0$, the target function S of our DP search is defined in an iterative manner as:

$$S(t,i) = \begin{cases} R_{TX_0}(i) & (t=0), \\ \max_j\{S(t-1,j)\cdot R_{X_tX_{t-1}}(i-j)\} + R_{TX_t}(i) & (t>0), \end{cases} \quad (17)$$

where i and j are the indices in the template-frame correlation function. The pitch value P_i can be converted from the index i by Eq. (13). We compute a score for each pitch candidate at $t=0$ according to Eq. (7). For each subsequent time point, we compute the scores iteratively using the scores from the previous frame. The pointer to the best past node is saved for backtracking upon reaching the last frame. Due to the logarithmic sampling of the DLFT, the search space for pitch values is naturally quantized logarithmically with constant $\Delta f_0/f_0$. Despite the first harmonic of the spectrum being fairly weak, the DP search is able to track f_0 whenever there is clear harmonic structure.

C. Adaptations of algorithm for killer whale vocalizations

This paper focuses on tracking the pulsed calls recorded from several Norwegian killer whale groups, each of which produces 3–16 call types (Strager, 1993, 1995; Van Opzeeland *et al.*, 2005). The LFCs of these calls can be characterized by a variety of f_0 patterns, including gradual or abrupt upsweeps and downsweeps, relatively constant frequencies, and abrupt transitions between these constant frequencies. The HFC is typically characterized by a gradual upsweep, though downsweeps and constant frequencies are also observed. In general, the frequency modulation of the LFC is more variable than that of the HFC. Figure 1 displays the spectrograms of some examples of these call type patterns, including simultaneous LFC and HFC [Figs. 1(a), 1(c), and 1(f)], abrupt frequency transitions [Figs. 1(a), 1(c), and 1(d)], and very dynamic f_0 range [as low as 230 Hz in Fig. 1(e) and as high as 11.25 kHz in Fig. 1(f)]. Notice that the low-frequency spectral energy of the killer whale calls is often masked by ambient boat noise.

Given that killer whale calls have distinctive f_0 dynamics for different call types, it was unrealistic to expect that a single set of parameters would work well for all call types. More importantly, some call types contain both LFCs and HFCs, which clearly could not be tracked with one set of parameter settings. To solve this multi-pitch problem, three sets of parameters aimed at tracking f_0 in three frequency ranges for different types of killer whale calls were identified. The first setting is used to track LFCs that have an f_0 below 600 Hz [e.g., Fig. 1(e)]. The second setting aims to track LFCs between 400 and 4000 Hz. These include calls that have a rising or falling f_0 in that frequency range [e.g., Figs. 1(a) and 1(b)], as well as calls with a relatively flat f_0 contour that can contain abrupt changes [e.g., the LFC in Figs. 1(c) and 1(d)]. The third setting is used to track the HFCs, which typically range between 4 and 12 kHz [e.g., Figs. 1(a), 1(c), and 1(f)].

For killer whale recordings, the optimal set of parameters can depend on the dynamics of the f_0 contour (e.g., slow vs abrupt changes). Parameterization issues can be alleviated if the harmonic matching principle is incorporated and brute force approach to estimating the pitch is adopted. Recall that the frame-based pitch estimation is obtained by shifting the harmonic template linearly to find the correlation maximum with the signal's DLFT spectrum. Correlation of finite-length sequences tends to taper off as the relative shift between these sequences increases. With normalized cross-correlation (i.e., the correlation is normalized by the energy in the overlapped region), the harmonic template parameters are important in balancing the bias between shifting left and shifting right. The problem would be resolved if, instead of shifting the harmonic template, the f_0 of the pulse train is changed and its DLFT is recomputed to obtain a new harmonic template. In this way, the correlation is always computed on the same (full) length of the signal and pulse train's DLFT spectra. The drawback of this approach is that M DLFT spectra must be computed and stored as harmonic templates, where M is the number of pitch hypotheses in the search space and could be large to achieve a refined resolution. Given that pitch tracking for killer whale recordings is typically not done in real time, currently the added computation requirement is not likely to be a serious issue. Again, the three sets of parameters used here are summarized in Table I. The parameters were selected to optimize performance of the algorithm on a training set.

D. Data collection

Free-ranging killer whales were tagged with digital archival tags containing acoustic and movement sensors (Johnson and Tyack, 2003) in Tysfjord, Norway in November 2005. These tags sampled sound at 96 kHz and were recovered upon their scheduled release for data offload. Eight orcas were tagged in all, but only the vocalizations recorded by six of these tags were evaluated for performance here. These six tags recorded for 20.6 h in total. Manual auditing documented the times and durations of calls that were clearly audible. Clearly audible calls were excerpted from the recordings, down-sampled to 48 kHz to match the human data on which the algorithm had been trained, and saved as separate files. If the call type had been observed before, it was labeled according to its earlier designation

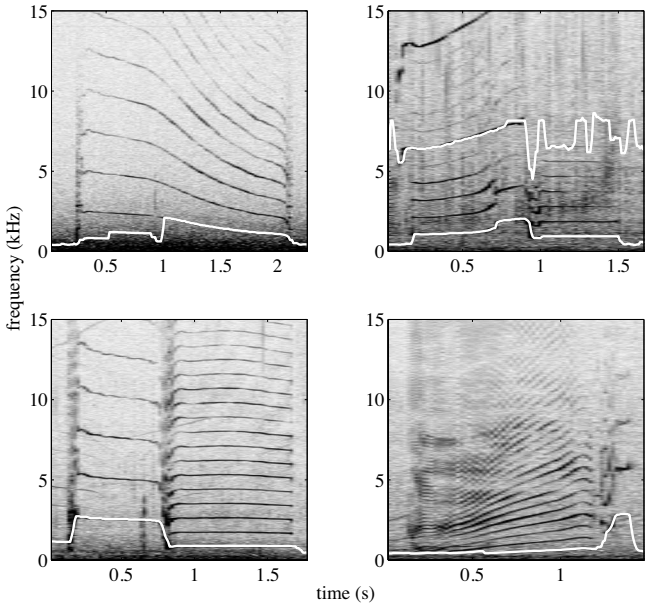


FIG. 5. Spectrograms of killer whale calls overlaid with pitch tracks computed by the DLFT-PDA. All calls contain a LFC but only the call in the upper right contains a HFC.

(Strager, 1993, 1995). Otherwise, new assignments were made (Shapiro, 2008). The DLFT-PDA was run on each of these files with the three parameter settings described previously, after which the traces were manually corrected to provide references for subsequent evaluation (see Fig. 5 for sample pre-corrected traces). During this post-processing stage, each call was also marked with beginning and ending times and labeled by its type. Each call took between 1 and 60 s to correct depending on the accuracy of the automatic trace.

E. Evaluation of performance

The algorithms for tracking killer whale calls (see below) were compared using gross error rate (GER) and fine error (FE) metrics. Specifically, GER is the percentage of f_0 hypotheses that deviate from the reference value (i.e., the manually-determined value) by more than 20%. The FE is characterized by the mean absolute value of the percentage deviation from the reference, excluding frames deviating from the reference by more than 20%. The f_0 of killer whale calls spans a large dynamic range, and so the pitch hypotheses were normalized by the reference values in assessing the FE. Because there are two outputs for the LFC (due to two parameter settings), the more accurate one (i.e., the contour with smaller GER) was chosen for evaluation, similar to the strategy a human labeler would use for post-editing.

To generate a profile of performance on different SNRs, the SNR for all non-overlapping calls was calculated. The SNR was calculated using a segment of background noise of the same duration (without vocalizing or surfacing sounds) occurring as close to the signal as possible within about 30 s. Before the SNR was computed, the recording was band-pass filtered with a two-pole Butterworth filter using the cutoff ranges of $[400,3000]$ Hz and $[4,12]$ kHz for the LFCs and HFCs, respectively. The SNR was measured in terms of energy flux density (Madsen, 2005).

The performance of DLFT-PDA was compared with the peak-picking and sidewinder algorithms (Deecke *et al.*, 1999) for tracking killer whale calls. The peak-picking method employed here selects the maximum frequency value of each power spectrum slice of the spectrogram followed by a smoothing step to remove any outliers. The sidewinder algorithm was implemented in both manners mentioned previously (i.e., the peak of the real cepstrum of the autocovariance sequence, and the second highest peak of the autocovariance sequence itself), and the better output of the two methods was selected for each call in the evaluation. In both the peak-picking and sidewinder algorithms, a restricted frequency range was considered for the LFC ($[400,3000]$ Hz) and the HFC ($[4,12]$ kHz). Before being implemented, these two alternative methods were checked against several call type exemplars to ensure their accuracy. In addition, the code used to deploy the sidewinder algorithm by Deecke *et al.* (1999) was kindly furnished by Deecke for this analysis.

TABLE II. Performance of DLFT-PDA, peak-picking, and sidewinder algorithms on the LFCs and HFCs of killer whale calls with different SNRs. The average SNR and the number of contours evaluated in each condition are also provided in the table. GER is the 20% gross error rate. FE is the mean of the normalized absolute fine error. See text for details.

			DLFT-PDA		Peak-picking		Sidewinder	
SNR range (dB)	N	Mean SNR (dB)	GER (%)	FE (%)	GER (%)	FE (%)	GER (%)	FE (%)
			LFC					
0–10	1180	4.7	29.0	1.5	64.3	4.9	49.8	6.7
10–20	647	14.4	16.8	0.9	51.7	3.6	37.8	4.3
>20	268	25.2	12.8	0.8	34.8	3.3	21.1	3.0
All	2095	10.3	22.5	1.2	55.9	4.2	41.7	5.3
			HFC					
0–10	182	6.6	24.3	5.6	27.1	2.0	25.1	9.5
10–20	346	15.2	13.3	2.7	28.2	1.9	22.9	8.4
>20	318	26.3	7.6	2.4	38.2	2.5	28.9	9.3
All	846	17.5	13.6	3.2	31.7	2.2	25.5	9.0

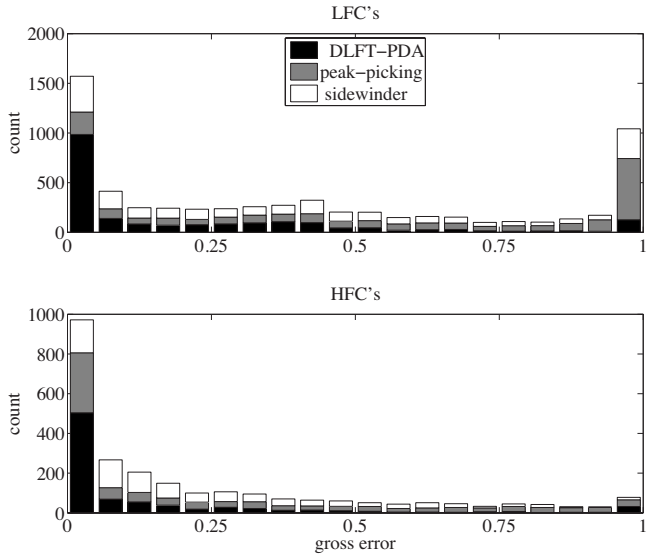


FIG. 6. Histogram of GERs for LFCs (top) and HFCs (bottom).

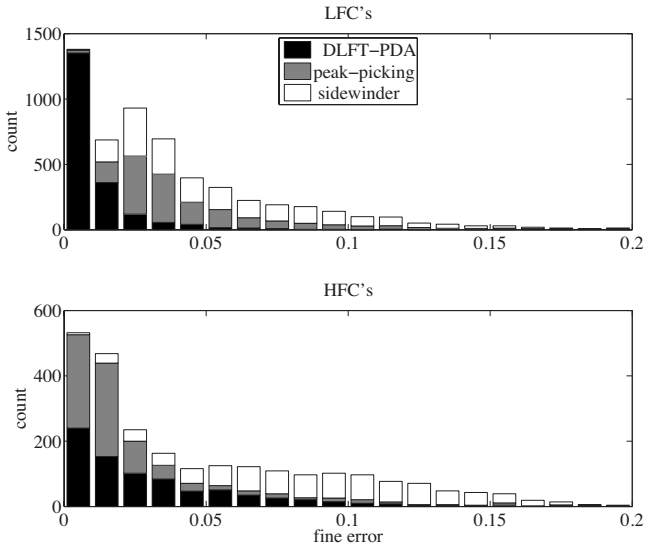


FIG. 7. Histogram of FE for LFCs (top) and HFCs (bottom).

III. RESULTS

The results from this comparison are reported in Table II for the LFCs and HFCs for three different SNR ranges: below 10 dB, between 10 and 20 dB, and above 20 dB. The DLFT-PDA performed better than the peak-picking and sidewinder algorithms on the LFCs in terms of both gross and FEs for all SNR ranges. The time required for manual post-editing will therefore be substantially reduced using the DLFT-PDA for initial contour tracing. The DLFT-PDA also outperformed the other two algorithms on the HFCs for all SNR ranges, except that the peak-picking algorithm had better FE results. The peak-picking algorithm performed much worse on the LFCs than on the HFCs. This is likely because the spectrogram of the HFCs normally has only one harmonic peak in the pitch search range for the HFC, in contrast with the spectrogram of LFCs that is characterized by multiple harmonic peaks in the search range (see Fig. 1). Consequently, it was easier for the peak-picking algorithm to locate f_0 in the HFC, except for the case when the LFCs had stronger harmonic peaks than the HFC in the search range. A potential refinement to our method is to use peak-picking to fine-tune the contour traced by DLFT-PDA prior to manual correction.

To provide a detailed view of how well the DLFT-PDA performs on individual calls, a histrogram of the GER for calls is plotted in Fig. 6. The algorithm is able to trace many of the contours accurately: 46.9% and 59.6% of all extracted LFCs and HFCs, respectively, were characterized by a GER of 5% or less. The percentage of LFCs that the DLFT-PDA seriously mistraced was small. For example, 9.1% of LFCs had a GER of over 70%, in contrast to 45.2% for the peak-picking algorithm and 23.9% for the sidewinder algorithm. A histogram of the FE for calls is plotted in Fig. 7.

IV. DISCUSSION AND CONCLUSIONS

In this article, a pitch tracking algorithm originally designed for human telephone speech was modified for killer whale vocalizations. The algorithm derives reliable estimations of pitch and the temporal change of pitch from the entire harmonic structure. The correlation of the DLFT spectrum with a carefully constructed harmonic template provides a robust estimation of the f_0, especially for low frequencies. The correlation of two DLFT spectra from adjacent frames gives a very reliable estimation of the f_0 change. The estimations of both f_0 and the temporal change in f_0 are then combined in a DP search to find a smooth pitch track. Evaluation results have demonstrated that the DLFT-PDA is capable of tracking killer whale calls that contain simultaneous LFC and HFC, and compares favorably to several other algorithms available for tracking killer whale vocalizations.

The challenges to DLFT-PDA plague all current pitch tracking algorithms. For example, it performs more poorly at low SNRs. In addition, it is unable to track multiple calls produced by multiple animals simultaneously. Currently, manual extraction and tracing of each call is necessary. To perform this task automatically, triangulation of the position of the callers using multiple recorders would be required at a minimum.

For the peak-picking and sidewinder algorithms, some GER and FE measurements that were associated with the HFC actually increased with a higher SNR (Table II). This might be explained by the relative energies of the LFC and HFC. Ambient boat noise occupied the lower frequency range. For calls with higher SNRs, more LFC energy would be visible to the tracker, making the HFC less obvious. In this situation, the peak-picking and sidewinder algorithms might have locked onto an upper harmonic of the LFC instead of selecting the fundamental frequency of the HFC, which would account for the poorer performance associated with increasing SNRs. This highlights one of the difficulties of tracking vocalizations with multiple simultaneous frequency components.

In addition to serving human telephone speech, the DLFT-PDA provides a reliable and unbiased approach toward determining the fundamental frequency of the pulsed calls of killer whales (see Nousek *et al.*, 2006; Miller *et al.*, 2007 for its successful application). A strongly performing

pitch tracker can lead to more objective results, which can then be used to characterize and classify vocalizations dependably. Future work should apply such a tracker to the calls of other species to automate, accelerate, and standardize the process. More generally, this manuscript highlights the benefit of introducing techniques developed for analyzing human speech into the realm of marine mammal vocalizations. The field of speech recognition has negotiated numerous challenges in signal processing. Rather than replicating these efforts, research on marine mammal acoustics will be well served by incorporating such advances in human speech processing.

ACKNOWLEDGMENTS

The authors would like to thank Stephanie Seneff and Peter L. Tyack for their support and advice, Ghinwa F. Choueiter for helpful comments on multiple drafts of the manuscript, Volker B. Deecke for providing implementations of the peak-picking and sidewinder algorithms, and two anonymous reviewers who improved the flow and transparency of this document. Special thanks to Sara Kim, Gary Matthias, Rebecca McGowan, Maitagorri Schade, and Ivan Dimitrov for their assistance with the post-processing corrections of the contours. We are grateful to Tiu Similä, Mads Christoffersen, Geoff Magee, Patrick Miller, Petter Helgevold-Kvadsheim, Sanna Kuningas, Sari Oksanen and Filipa Samarra for assistance in the field collecting the killer whale data. Financial support for collection of the Norwegian killer whale vocal data was provided by the Ocean Life Institute of the Woods Hole Oceanographic Institution and the National Geographic Society. The MIT Undergraduate Research Opportunities Program (UROP) provided support for some of the post-editing work. C.W. was supported by DARPA under Contract No. N66001-96-C-8526, monitored through Naval Command, Control, and Ocean Surveillance Center and by the National Science Foundation under Grant No. IRI-9618731. A.D.S. was supported by a National Defense Science and Engineering Graduate Fellowship.

Brown, J. C., Hodgins-Davis, A., and Miller, P. J. O. (**2006**). "Classification of vocalizations of killer whales using dynamic time warping," J. Acoust. Soc. Am. **119**, EL34–EL40.

Buck, J. R., and Tyack, P. L. (**1993**). "A quantitative measure of similarity for *Tursiops truncatus* signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506.

Deecke, V. B., Ford, J. K. B., and Spong, P. (**1999**). "Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects," J. Acoust. Soc. Am. **105**, 2499–2507.

Droppo, J., and Acero, A. (**1998**). "Maximum *a posteriori* pitch tracking," in Proceedings of ICSLP, Sydney, Australia, pp. 943–946.

Ford, J. K. B. (**1987**). "A catalogue of underwater calls produced by killer whales (*Orcinus orca*) in British Columbia," Canadian Data Report of Fisheries and Aquatic Sciences **633**, 1–165.

Ford, J. K. B. (**1989**). "Acoustic behavior of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia," Can. J. Zool. **67**, 727–745.

Ford, J. K. B. (**1991**). "Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia," Can. J. Zool. **69**, 1454–1483.

Geoffrois, E. (**1996**). "The multi-lag-window method for robust extended-range f_0 determination," in Proceedings of ICSLP, Philadelphia, PA, pp. 2399–2402.

Hermes, D. J. (**1988**). "Measurement of pitch by subharmonic summation," J. Acoust. Soc. Am. **83**, 257–264.

Hess, W. (**1983**). *Pitch Determination of Speech Signals* (Springer-Verlag, Berlin, Germany).

Hoelzel, A. R., and Osborne, R. W. (**1986**). "Killer whale call characteristics: Implications for cooperative foraging strategies," in *Behavioral Biology of Killer Whales*, edited by B. C. Kirkevold, and J. S. Lockard (Alan R. Liss, Inc., New York), pp. 373–403.

Janik, V. M., Dehnhardt, G., and Todt, D. (**1994**). "Signature whistle variations in a bottlenosed dolphin, *Tursiops truncatus*," Behav. Ecol. Sociobiol. **35**, 243–248.

Johnson, M. P., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," Inf. Sci. (N.Y.) **28**, 3–12.

Klapuri, A. (**2008**). "Multipitch analysis of polyphonic music and speech signals using an auditory model," IEEE Trans. Audio, Speech, Lang. Process. **16**, 255–266.

Klapuri, A., and Virtanen, T. (**2008**). "Progress towards automatic music transcription," in *Handbook of Signal Processing in Acoustics*, edited by D. Havelock, S. Kuwano, and M. Vorlander (Springer-Verlag, Berlin).

Lahat, M., Niederjohn, R. J., and Krubsack, D. A. (**1987**). "A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech," IEEE Trans. Acoust., Speech, Signal Process. **35**, 741–750.

Madsen, P. T. (**2005**). "Marine mammals and noise: Problems with root mean square sound pressure levels for transients," J. Acoust. Soc. Am. **117**, 3952–3957.

McCowan, B. (**1995**). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*)," Ethology **100**, 177–193.

Miller, P. J. O. (**2002**). "Mixed-directionality of killer whale stereotyped calls: A direction of movement cue?," Behav. Ecol. Sociobiol. **52**, 262–270.

Miller, P. J. O., and Bain, D. E. (**2000**). "Within-pod variation in the sound production of a pod of killer whales, *Orcinus orca*," Anim. Behav. **60**, 617–628.

Miller, P. J. O., Samarra, F. I. P., and Perthuison, A. D. (**2007**). "Caller sex and orientation influence spectral characteristics of "two-voice" stereotyped calls produced by free-ranging killer whales," J. Acoust. Soc. Am. **121**, 3932–3937.

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (**2004**). "Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*," Anim. Behav. **67**, 1099–1107.

Nousek, A. E., Slater, P. J. B., Wang, C., and Miller, P. J. O. (**2006**). "The influence of social affiliation on individual vocal signatures of northern resident killer whales (*Orcinus orca*)," Biol. Lett. **2**, 481–484.

Secrest, B. G., and Doddington, G. R. (**1983**). "An integrated pitch tracking algorithm for speech synthesis," in Proceedings of ICASSP, Boston, MA, pp. 1352–1355.

Shapiro, A. D. (**2006**). "Preliminary evidence for signature vocalizations among free-ranging narwhals (*Monodon monoceros*)," J. Acoust. Soc. Am. **120**, 1695–1705.

Shapiro, A. D. (**2008**). "Orchestration: The movement and vocal behavior of free-ranging Norwegian killer whales (*Orcinus orca*)," in *Biological Oceanography*, Ph.D. thesis (MIT/WHOI).

Strager, H. (**1993**). "Catalogue of underwater calls from killer whales (*Orcinus orca*) in northern Norway," (University of Århus, Århus, Denmark).

Strager, H. (**1995**). "Pod-specific call repertoires and compound calls of killer whales, *Orcinus orca* Linnaeus, 1758, in the waters of northern Norway," Can. J. Zool. **73**, 1037–1047.

Talkin, D. (**1995**). "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, edited by W. B. Kleijn and K. K. Paliwal (Elsevier, New York), pp. 495–518.

van Opzeeland, I. C., Corkeron, P. J., Leyssen, T., Similä, T., and Van Parijs, S. M. (**2005**). "Acoustic behaviour of Norwegian killer whales, *Orcinus orca*, during carousel and seiner foraging on spring-spawning herring," Aquat. Mamm. **31**, 110–119.

Wang, C. (**2001**). "Prosodic modeling for improved speech recognition and understanding," Ph.D. thesis (MIT, Cambridge, MA).

Wang, C., and Seneff, S. (**2000a**). "Improved tone recognition by normalizing for coarticulation and intonation effects," in Proceedings of the International Conference on Spoken Language Processing, Beijing, China.

Wang, C., and Seneff, S. (**2000b**). "Robust pitch tracking for prosodic modeling in telephone speech," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, pp. 1143–1146.

Wang, C., and Seneff, S. (**2001a**). "Lexical stress modeling for improved speech recognition of spontaneous telephone speech in the JUPITER domain," in Proceedings of EUROSPEECH, Aalborg, Denmark.
Wang, C., and Seneff, S. (**2001b**). "Prosodic scoring of recognition outputs in the JUPITER domain," in Proceedings of International Speech Communication Association Workshop on Prosody in Speech Recognition and Understanding, Red Bank, NJ.
Watkins, W. A. (**1967**). "The harmonic interval: Fact or artifact in spectral analysis of pulse trains," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 15–43.
Watwood, S. L., Tyack, P. L., and Wells, R. S. (**2004**). "Whistle sharing in paired male bottlenose dolphins, Tursiops truncatus," Behav. Ecol. Sociobiol. **55**, 531–543.
Watwood, S. L., Owen, E. C. G., Tyack, P. L., and Wells, R. S. (**2005**). "Signature whistle use by temporarily restrained and free-swimming bottlenose dolphins, *Tursiops truncatus*," Anim. Behav. **69**, 1373–1386.

Acoustic basis for fish prey discrimination by echolocating dolphins and porpoises

Whitlow W. L. Au
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Brian K. Branstetter
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 71510, 53560 Hull Street, San Diego, California 92152

Kelly J. Benoit-Bird
College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331

Ronald A. Kastelein
SEAMARCO, Julianalaan 46, 3843 CC Harderwijk, The Netherlands

(Received 24 December 2008; revised 1 April 2009; accepted 9 May 2009)

The biosonar system of dolphins and porpoises has been studied for about 5 decades and much has been learned [Au, W. W. L. (1993). *The Sonar of Dolphins* (Springer, New York)]. Most experiments have involved human-made targets; little is known about odontocetes' echolocation of prey. To address this issue, acoustic backscatter from Atlantic cod (*Gadus morhua*), gray mullet (*Chelon labrosus*), pollack, (*Pollachius pollachius*), and sea bass (*Dicentrarchus labrax*) was measured using simulated biosonar signals of the Atlantic bottlenose dolphin and harbor porpoise. The fish specimens were rotated so that the effects of the fish orientation on the echoes could be determined. Echoes had the highest amplitude and simplest structure when the incident angle was perpendicular to the longitudinal axis of the fish. The complexity of the echoes increased as the aspect angle of the fish moved away from the normal aspect. The echoes in both the time and frequency domains were easily distinguishable among the four species of fish and were generally consistent within species. A cochlear model consisting of a bank of band-passed filters was also used to analyze the echoes. The overall results suggest that there are sufficient acoustic cues available to discriminate between the four species of fish based on the echoes received, independent of aspect angle.
 [DOI: 10.1121/1.3147497]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Jz [JAS] Pages: 460–467

I. INTRODUCTION

Echolocation experiments with captive dolphins began about 5 decades ago with Scheville and Lawrence (1956) and Kellogg (1958) attempting to obtain evidence that bottlenose dolphins (*Tursiops truncates*) echolocated. Norris *et al.* (1961) provided unequivocal evidence to demonstrate echolocation in dolphins by using suction cup blindfold to cover a dolphin's eyes while the animal was required to swim and avoid obstacles and retrieve fish rewards that were thrown into the animal's tank. Busnel and Dziedzic (1967) also trained a blindfolded harbor porpoise (*Phocoena phocoena*) to swim through a maze of vertically hanging wire. Following these initial studies, various types of echolocation experiments have been performed to study the biosonar process and determine the capabilities of odontocetes to detect, discriminate, localize, and recognize targets.

The biosonar capabilities of dolphins to perform complex target discrimination tasks have been conducted mainly with objects that are foreign to these animals but familiar to humans. Review articles on the target discrimination experiments have been written by Nachtigall (1980), Au (1993), and Au and Hastings (2008). Some of these experiments included material and wall thickness discrimination of metallic plates (Evans and Powell, 1967), material composition of cylinders at arbitrary aspects (Au and Turl, 1991), material composition discrimination of spheres (Aubaurer *et al.*, 2000), shape discrimination of planar targets (Barta, 1969), shape discrimination between spheres and cylinders (Au *et al.*, 1980), shape matching of polyvinyl chloride (PVC) objects across vision and echolocation (Pack *et al.*, 2002), and wall thickness of metallic cylinders (Au and Pawloski, 1992). These and other experiments have clearly shown that dolphins possess a very sophisticated biosonar system that has certain capabilities beyond the most modern and sophisticated technological sonar. From these experiments, we have gained much knowledge about the target discrimination and recognition capabilities of the dolphin biosonar system, yet these experiments provide little insight on the issues involving dolphins and porpoises foraging for prey in the wild.

The question of how far an echolocating dolphin can detect fish prey has only been addressed recently by Au *et al.* (2004) who calculated the biosonar detection ranges of killer whales foraging for Chinook salmon, by Madsen *et al.* (2004) who estimated the detection ranges of false killer whales and Risso's dolphin foraging for unspecified species of fish, and by Au *et al.* (2007) who calculated the biosonar

 0001-4966/2009/126(1)/460/8/$25.00

detection ranges of bottlenose dolphins and harbor porpoises foraging for Atlantic cod, mullet, sea bass, and pollack. However, we still need to address the issue of acoustic cues from echoes that would allow a dolphin to discriminate and recognize different species of fish allowing for selective foraging. The focus of this paper is to determine what acoustic cues are present in the echoes of fish prey would allow dolphins and porpoises to discriminate and recognize different species of fish. While it is extremely difficult to address the issue of selective foraging by echolocating dolphins and porpoise because of the difficulties in making good, regular, and consistent observations of underwater foraging behavior in the wild, a clear case of selective foraging exists for fish eating killer whales in the waters of British Columbia (Ford and Ellis, 2006). Even in months when Chinook salmon may constitute less than 15% of the salmon population, the whales still forage mainly on Chinook salmon (Ford and Ellis, 2006). Visual observations of foraging killer whales strongly suggest that they depend on echolocation to detect and recognize their prey. Whales would often be observed swimming near the surface along nearly straight line tracks for minutes and then suddenly submerge and resurface several tens of meters away with a salmon in their mouths. Collection of scales after the whales bring the prey to the surface has allowed for the identification of the salmon species. Unfortunately, such selective foraging by other species of odontocetes has not been reported. The specific cues that odontocetes may use to discriminate and recognize different species of fish will not be addressed in this study; rather, the focus will be on determining if acoustic cues that can be used for species discrimination are indeed present in the echoes of fish, an important component of selective foraging.

II. PROCEDURE

A. Experimental geometry

This study is an extension of the work that was reported by Au *et al.* (2007) on modeling the biosonar detection range for four species of fish and the description of the procedure will be brief with only important aspects repeated. The backscatter measurements were conducted in an outdoor tank belonging to the Sea Mammal Research Company (Seamarco) at the field station of the Netherland's National Institute for Coastal and Marine Management (RIKZ) in Jacobahaven, Zeeland, The Netherlands. The surface dimension of the tank was 7×4 m^2 with a water depth of 2 m. Anesthetized fish subjects were constrained in a monofilament bag that was in turn attached to a monofilament net which was attached to a rotor, as shown in Fig. 1(a). The orientation system that will be used in this study is shown in Fig. 1(b) where the arrows indicate the direction of the incident acoustic signal. The fish were rotated as simulated biosonar signals of *Tursiops truncatus* and *Phocoena phocoena* were projected and the echoes collected. A monostatic system with the same transducer projecting the signals and receiving the echoes was used. Both signals are shown in Fig. 2, with the dolphin-like signal having a peak frequency of 130 kHz and the porpoise-like signal having a peak frequency of 138 kHz. The duration of the

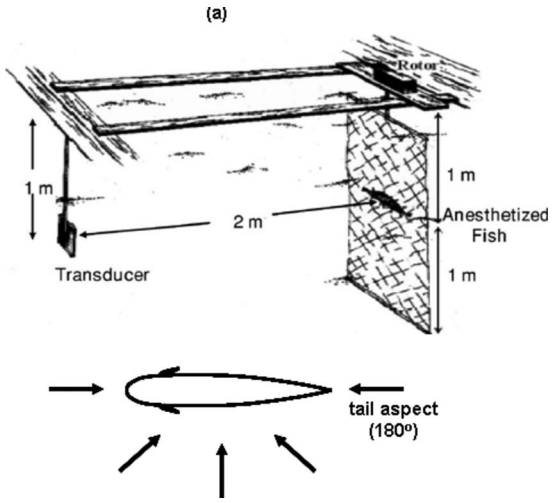


FIG. 1. (a) Experimental geometry showing a monofilament net curtain attached to a rotor with a fish subject to a monofilament net bag attached to the curtain. A monostatic echo ranging system was used in which the same transducer projected the signal and received the echoes. (b) The orientation system used in this study showing the direction of the incident signal with respect to the fish body.

dolphin-like signal was approximately 70 μs while the porpoise-like signal was approximately 270 μs in duration.

B. Fish subjects

The species of fish used were Atlantic cod (*Gadus morhua*), gray mullet (*Chelon labrosus*), pollack (*Pollachius pollachius*), and sea bass (*Dicentrarchus labras*). Three fish of each species except for the pollack were examined acoustically. The lengths of the subjects were cod (29–30 cm), mullet (15–17 cm), sea bass (14–17 cm), and pollack (21 cm). These fish were on loan from "The Arsenaal Aquarium," Vlissingen, The Netherlands. They were fed to satiation each day after the measurement sessions on a diet of raw fish and in compliance with The Animal Welfare Commission of The Netherlands. After the measurements they were returned to the aquarium. Since the fish were borrowed,

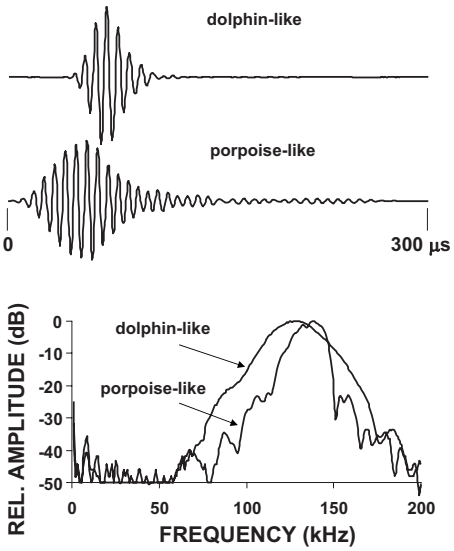


FIG. 2. The dolphin-like and porpoise-like signal waveforms and spectra.

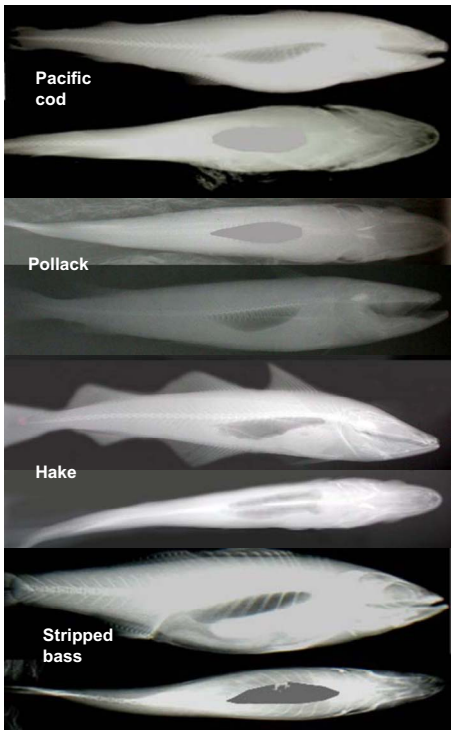


FIG. 3. (Color online) Radiograph images of four species of fish that are closely related to the Atlantic cod, mullet, and sea bass used in this study (courtesy of Dr. John Horne, U. Washington).

we did not attempt to x-ray them and risk potential injury. However, we were able to obtain radiograph images from closely related species and these are shown in Fig. 3. The images in the dorsal aspect for three of the species were digitally enhanced using PHOTOSHOP. The purpose of showing this figure is to convey to those unfamiliar with swimbladder geometry that the shape, orientation, and volume of swimbladders vary between species (Simmonds and MacLennan, 2005). Since the swimbladder is the most prominent structure affecting backscatter of acoustic signals (Foote, 1980; Foote and Ona, 1985) we expected the echoes from different fish species to have different temporal and spectral structures that could be resolved with dolphin and porpoise biosignals.

C. Analysis with a pheripheral auditory filter model

The auditory filter model of the bottlenose dolphin developed by Branstetter *et al.* (2007) was one of the tools used to examine the time-frequency characteristics of the fish echoes. The model consisted of a bank of gammatone filters, each followed by a half-wave rectifier and a low-pass filter. The output of this model resembles a spectrogram. However, unlike a spectrogram which applies the same arbitrary window lengths and shapes across frequencies (e.g., 512 point Hanning with 50% overlap), the auditory filter model incorporates the spectral and temporal resolution of bottlenose dolphin's auditory periphery. The resulting output provides a closer approximation to what the dolphin actually hears.

The auditory filter shapes of a bottlenose dolphin were measured by Lemonds (1999) at 60, 90, and 120 kHz. The shape of the auditory filters closely resembled those of a

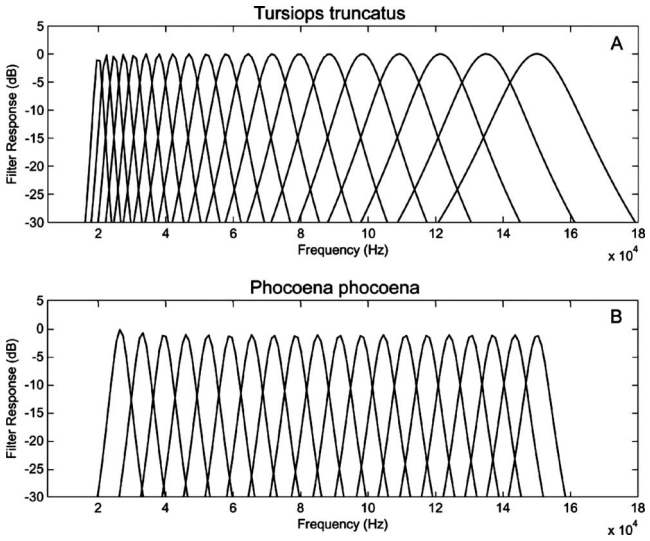


FIG. 4. Gammatone filter bank (a) modeling the auditory filterbank of a dolphin with constant Q filters using the results of Branstetter *et al.* (2007) and (b) modeling the auditory filterbank of a harbor porpoise using the results of Popov *et al.* (2006).

gammatone filter described by Patterson (1994). The impulse response of a gammatone filter is given by the equation (Patterson, 1994)

$$g_t(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c t - \phi), \quad (1)$$

where f_c is the center frequency of the filter, ϕ is the starting phase, and a, b, and n are the parameters determining the ramping and duration of the impulse function and consequently the width and shape of a filter (Slaney, 1993). The parameter b is related to the equivalent rectangular bandwidth (ERB), both defined by the equations

$$b = k \times \mathrm{ERB}(f_c), \quad (2)$$

$$\mathrm{ERB}(f_c) = \frac{f_c}{Q} + \min BW, \quad (3)$$

where k is a constant and is equal to 1.019 (Patterson *et al.*, 1992), Q is the ratio of center frequency over bandwidth, and min BW is the minimum bandwidth for the low frequency channels. Since the critical ratio in dolphins approaches that of humans (Johnson, 1968) the estimated value of 24.7 for humans (Glasberg and Moore, 1990) was used in this study. Branstetter *et al.* (2007) found that a gammatone filterbank with a Q of 11.3 will produce an excellent fit to the two roex filter derived for the dolphin by Lemonds (1999) and will be used here. The gammatone filter bank that will be used to analyze the fish echoes produced by the simulated dolphin biosonar signal is shown in Fig. 4(a). In reality, the filterbank consisted of 94 frequency channels spaced between approximately 80 and 160 kHz since the spectrum of the incident signal was lower than −30 dB below a frequency of approximately 70 kHz.

Popov *et al.* (2006) used a tone-tone masking paradigm with the envelope-following response evoked potential technique to measure the auditory filter shapes of *Phocoena phocoena* and *Neophocaena phocaenoides*. They described their results in terms of the roex function of Patterson *et al.* (1992)

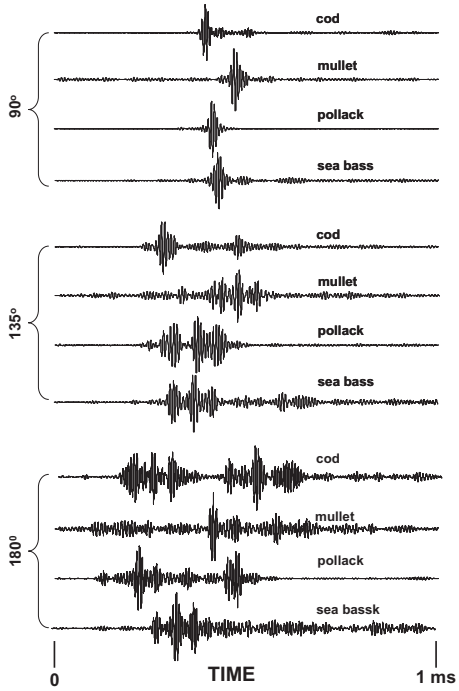


FIG. 5. Echo waveforms using the dolphin-like biosonar signal for the four species of fish examined at the broadside, 135°, and tail aspects. Each waveform was normalized.

and estimated Q values of approximately 10 at 30 kHz to approximately 40 at 150 kHz, varying with center frequency according to the equation

$$Q = q_0 f_c^k, \tag{4}$$

where q_0 is a quality at $f_c = 1$, and k determined the degree of Q dependency on f_c. Constant bandwidth filters would have $k=0$ and constant Q filters would have $k=1$. Popov *et al.* (2006) found that k between 0.83 and 0.86 fit their threshold results best. The gammatone filter bank with Q values that varied according to Eq. (4) was used to analyze the results from the porpoise-like biosonar signal shown in Fig. 4(b).

III. RESULTS

Polar plots of the target strength (based on the energy flux density in the echoes and incident signal) as a function of the aspect angle from 0° to 360° were shown in the previous publication of Au *et al.* (2007). The polar plots of target strength were in general very similar in shape for all the specimens measured and would probably not provide much information on the species of fish producing the echoes.

Examples of the echo waveforms generated with the simulated dolphin biosonar signal for the four fish species are shown in Fig. 5 at aspect angles of 90° (broadside aspect), 135°, and 180° (tail aspect). Each waveform is normalized to its maximum value. The simplest echoes occurred at the broadside aspect and consisted mainly of the specular reflection from the surface of the swimbladder facing the transducer and some secondary components from other structures in the fish. Even at the broadside aspect, the echo waveforms can be distinguished from one another. As the

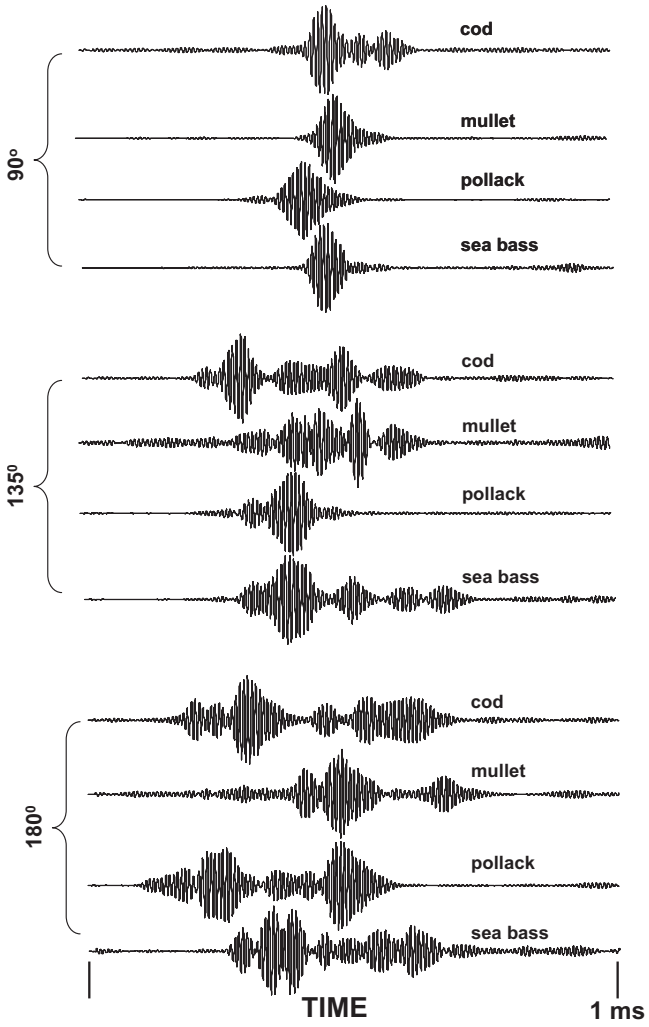


FIG. 6. Echo waveforms using the porpoise-like biosonar signal for the four species of fish examined at the broadside, 135°, and tail aspects. Each waveform was normalized.

aspect angle increased away from the broadside aspect, the echoes became longer in duration and more complex in structure as seen by the presence of more and larger secondary echo components. The difference in the echo structure between species became more distinguishable at these aspect angles than at the broadside aspect. Differences between the echoes from the four fish species at 135° and 180° include differences in the number of secondary echo components (highlights) and differences in the relative amplitude and spacing between the highlights. At the tail aspect, the echo duration was the longest which is consistent with typical swimbladder geometry. The x-ray images from the dorsal aspects in Fig. 3 show that the swimbladders are aligned with the longitudinal axis of the fish and are typically tilted dorsoventrally. Therefore, the incident signal entering a fish from the tail aspect will travel the maximum distance propagating from the tail-end of the swimbladder to the front-end.

A similar set of echo results as in Fig. 5 but for a porpoise-like biosonar signal is shown in Fig. 6. Even with a longer and narrower biosonar signal, the echoes returning to a porpoise show differences between species that could provide discrimination cues. At the broadside aspects, the secondary echo components are not resolvable with the narrow

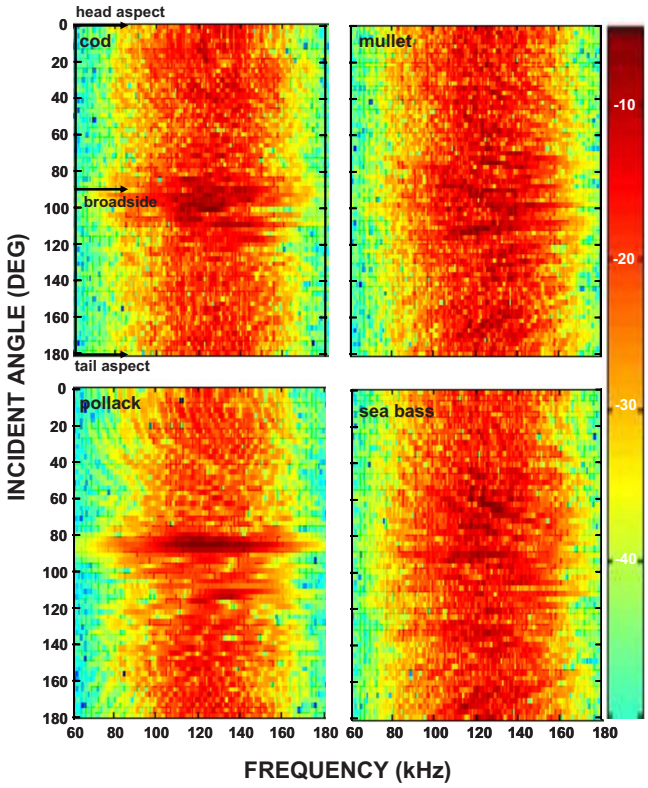


FIG. 7. (Color online) (Color online) Polargrams (frequency spectra versus polar angle of the echoes) for the four species of fish obtained with the dolphin-like biosonar signal. The frequency spectra for different polar angles are shown with the amplitude color coded according to the color bar on the right.

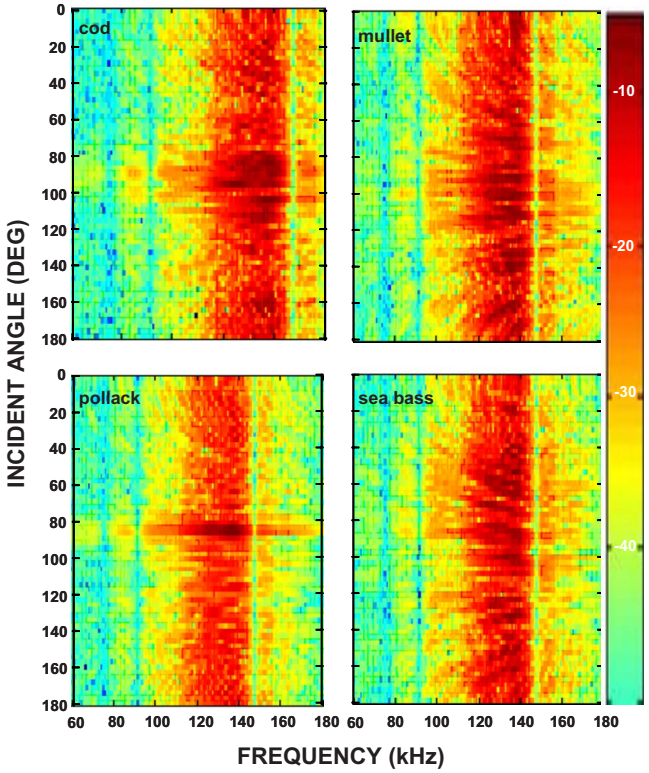


FIG. 8. (Color online) (Color online) Polargrams for the four species of fish obtained with the porpoise-like biosonar signal. The frequency spectra for different polar angles are shown with the amplitude color coded according to the color bar on the right.

band porpoise-like signal, except for the cod. As in the dolphin signal case, the differences become more pronounced as the aspect angle was increased from the broadside orientation. The same types of differences involving the number of highlights, the relative amplitude of highlights, and the time delay between highlights existed for the porpoise signal as observed in the dolphin signal.

Polargrams, which are the frequency spectra of the echoes as a function of the polar angle about one side of each fish species, are shown in Fig. 7 for the dolphin-like biosonar signal and in Fig. 8 for the porpoise-like biosonar signal. The amplitude of each spectrum is coded in color as shown in the color bar to one side of the figure. A similar kind of polargram can be drawn in which the envelope of the echo for each polar angle can be drawn as was done by Reeder *et al.* (2004). Perhaps the best way to visualize the polargrams is to step back and look at the pattern of changes in the spectra as the polar angle varies. Each polargram has a slightly different manner in which the echo spectra change with angle and this pattern may be used by dolphins and porpoises to discriminate a specific species of fish. One feature of the polargrams is the presence of diagonal stripes that indicate how information from different frequencies varies in a pattern as the fish aspect angle changed. These are caused by changes in the high-light separation time as the polar angle changes which will cause local maxima and minima in the spectrum to shift. The shift in local maxima and minima in the spectrum is reflected by the diagonal stripes. The polargrams clearly show differences in the spectra of the echoes from the difference fish species that can be utilized by dolphins and porpoises in discriminating between these four species of fish. In a natural situation, the predator-prey geometry will constantly change from ping to ping and the polargram can be used to gain an appreciation of how the spectra of the echoes will change as the predator-prey geometry changes continuously and dynamically.

The time-frequency representations of the echoes (Fig. 5) from the four fish species produced by analyzing the echoes with the gammatone filter bank of Fig. 4(a) are shown in Figs. 9 and 10 for the broadside and 135° incident angles, respectively. The frequency values along the vertical axis correspond to the center frequencies of some of the individual gammatone filters shown in Fig. 4(a). The time-frequency representations show how the spectra of the echoes develop as a function of time as the echoes propagate into the dolphin auditory system. Even for the broadside aspect, differences in the time-frequency representations can be seen between species. The time-frequency plot for the mullet and sea bass had the narrowest frequency extent. The frequency extent of the cod and pollack was similar and larger than for the mullet and sea bass. The differences become more apparent for the 135° aspect angle. Time-frequency representations of echoes (Fig. 6) associated with the porpoise signal and porpoise hearing model for the broadside and 135° aspects are shown in Figs. 11 and 12, respectively. Differences in the time-frequency plots are obvious and are likely exploited by porpoises to discriminate between different fish species. As would be expected, the time-frequency

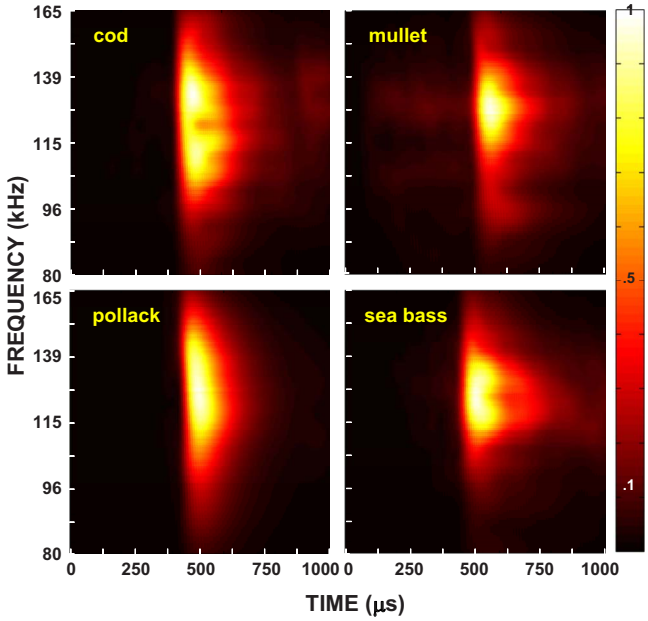


FIG. 9. (Color online) Time-frequency representation of the echoes using the dolphin-like biosonar signal at the broadside aspect.

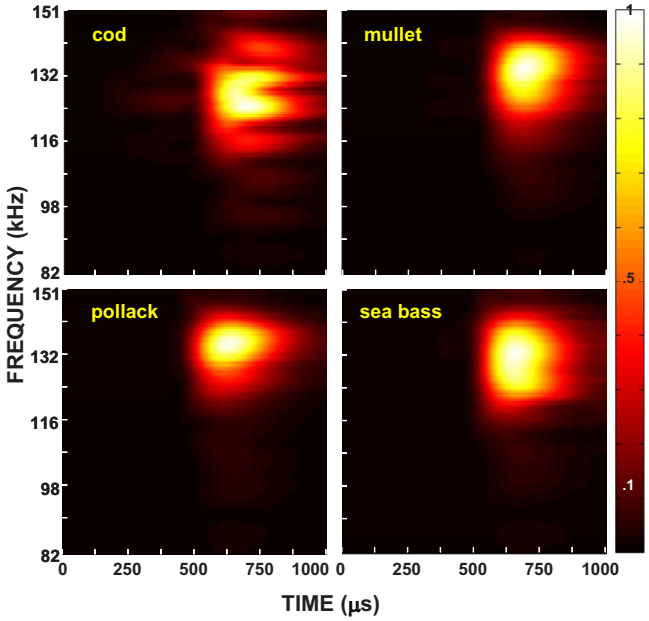


FIG. 11. (Color online) Time-frequency representation of the echoes using the porpoise-like biosonar signal at the broadside aspect.

plots for the porpoise signal are different from those for the dolphin because of the different incident signals and gammatone filter bandwidths.

IV. DISCUSSION AND CONCLUSIONS

The results obtained in this study resembled the broadband measurements performed by Au and Benoit-Bird (2003) for deep dwelling snappers (commercially referred to as bottom fish) in Hawaiian waters. In both studies, the echo structures were complex with many echo components originating from different parts of the fish anatomy. Reeder *et al.* (2004) focused on the backscatter process using broadband frequency modulated signals to measure the echoes from the fish, alewife, and were able to identify some of the sources of the secondary reflections which should be similar for the fish used in this study. The results in this study were analyzed and discussed from both an echo structure perspective in the time domain and in the frequency domain with the polargrams and in both domains simultaneously with the time-frequency hearing model plots. The multiple highlight feature of the echoes made analysis and interpretation in the time domain very insightful because the secondary echoes could be easily observed.

The polargrams showing how the frequency spectra of the echoes changed with the aspect angle of the fish also

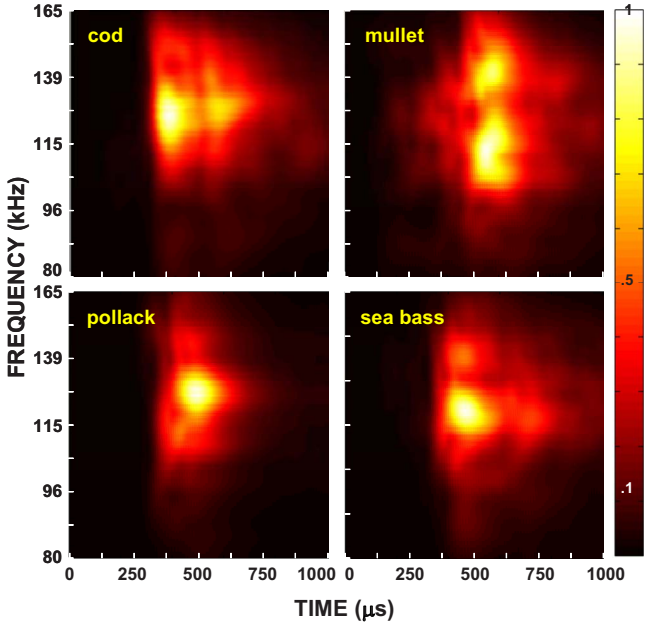


FIG. 10. (Color online) Time-frequency representation of the echoes using the dolphin-like biosonar signal at the 135° aspect.

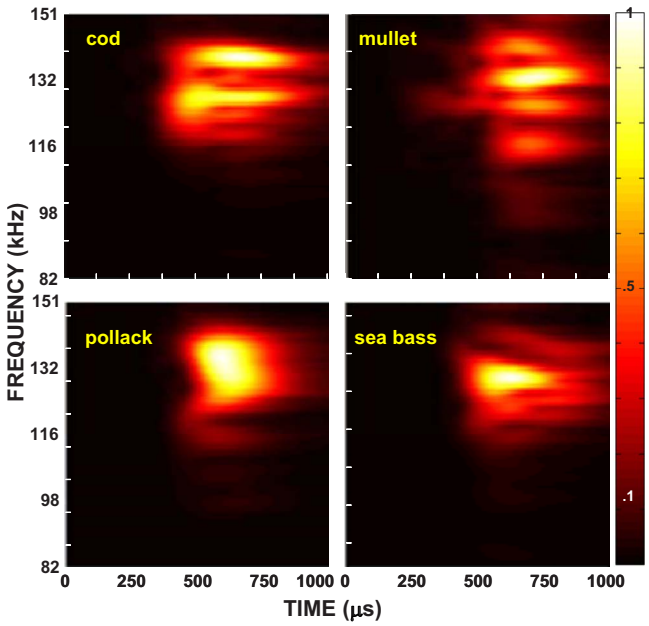


FIG. 12. (Color online) Time-frequency representation of the echoes using the porpoise-like biosonar signal at the 135° aspect.

provide insights into the differences between species because one can see how certain frequency peaks vary as a function of aspect angle. Therefore, whether the data are analyzed in the time or frequency domain is immaterial because species difference cues can easily be seen in both domains. However, the polargrams have the advantage of being compact, allowing the effects of aspect to be readily seen for all aspects on one side of the fish. Reeder *et al.* (2004) presented graphs similar to polargrams plotting the compressed pulse output (CPO) as a function of time-delay as the aspect angle changed. The CPO at any given angle is the envelope of the cross-correlation function between the echo from the subject with an echo from a calibrated sphere. Just as in the polargram, there were peaks in the CPO that varied as a function of the aspect angle. Combining the time and frequency information is likely most appropriate because the auditory system of mammals probably utilizes simultaneous time-frequency information rather than information in only one domain. A future follow-up to this study is to use the echoes collected in this study in a dolphin auditory model with a classification algorithm (e.g., Au, 1994; Branstetter *et al.*, 2007) as well as in a human listener experiment as has been done for echoes done with other targets (Au, 1993; Delong *et al.*, 2007).

The time-frequency representations based on the gammatone filter banks modeled the processing of the echoes by the auditory periphery. The results from only two aspect angles strongly suggest that the echoes from the four species contain sufficient species-specific information to facilitate discrimination by echolocating odontocetes. Although auditory images from only two angles are presented for brevity, similar species-specific patterns are apparent at different fish orientations.

Species-specific differences in the echo structure of backscattered acoustic signals from the four species of fish used in this study are apparent. The data indicate that the echo structures vary in amplitude, time separation between highlights, number of highlights, and overall duration depending on the angle of incident of echolocation signals. These results suggest a very complex backscattering process with various types of aspect-dependent information available. So the most obvious question is whether or not a dolphin or porpoise can handle the aspect-dependent fluctuations associated with reflections from different species of fish. From a slightly different perspective one could ask whether or not a dolphin or porpoise can generalize from fluctuating broadband echoes the species of a potential prey. The task for an odontocete is to detect, localize, recognize, and track a moving prey. Since both predator and prey move, the acoustic geometry will be continuously and dynamically changing, causing the echo structure to fluctuate from ping-to-ping. If an odontocete utilizes the echo structure information to hunt for specific prey, the odontocete auditory classification scheme would need to match and generalize a large variety of echo exemplars (some of which will be novel) to specific species categories. Such a generalization capability is not out of the question and has been demonstrated by a dolphin matching three dimensional, aspect-dependent targets (Helweg *et al.*, 1996) that were allowed to freely rotate, resulting in within target echo variability. Despite the large variability of within target echoes, the dolphin was successful at discriminating between the targets.

Just as the aspect-dependent echoes could be discriminated by echolocating *Tursiops truncates* in study of Helweg *et al.* (1996) it would not be far-fetched to assume that the same will be true for aspect-dependent echoes from fishes. However, in order to unequivocally demonstrate the capability for aspect-dependent discrimination of fish echoes, a rigorous psychophysical study would be needed. Such an experiment can be conducted with electronic generated phantom echoes (Aubauer *et al.* 2000; Ibsen *et al.*, 2007). The apparent aspect of fish echoes could be varied from ping-to-ping in a similar manner, as was done by Delong *et al.* (2007) for human listening experiments using echoes generated by simulated dolphin clicks and rotating targets.

It should be emphasized that although our results suggest that intraspecies difference in the echo structure exists for the four species examined, this type of information is probably not the only information used by an echolocating predator to detect, localize, and recognize specific species. There are potentially a multitude of cues that would be available through the echolocation process. The swimming behavior and dynamics of potential prey can be determined by examining the change in the echo amplitude and timing as well as the echo structure from ping-to-ping. The depth of potential prey could also be determined by the echolocation process. An odontocete will no doubt use as many available cues that are present. Furthermore, different cues may have different weights depending if the odontocetes forage in relatively open waters or in shallow waters, the variety of species present in the habitat, and the relative food value of these species.

ACKNOWLEDGMENTS

We thank Sander van der Hel for assistance in conducting the experiments. Jan van der Veen, Sea aquarium "het Arsenaal," The Netherlands, lent us the study animals. Gijs Rutjes (Coppens International) provided some of the sea bass. We thank Brigitte Kastelein and the volunteers for logistical support. The facilities of the research station were made available, thanks to Dick Vethaak (RIKZ), Roeland Allewijn (RIKZ), and Wanda Zevenboom (North Sea Directorate). This work was supported by the US Office of Naval Research, Mardi Hastings, Program Manager, and the Netherlands Ministry for Agriculture, Nature, and Food Quality (DKW-program 418: North Sea and Coast). This project complied with the Dutch standards for animal experiments (Chris Pool, Head of the Committee for Animal Experiments of RIKZ) and was conducted under University of Hawaii Animal Care Protocol 04-019. This is HIMB Contribution No. 1343.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).
Au, W. W. L. (**1994**). "Comparison of sonar discrimination: Dolphin and an artificial neural network," J. Acoust. Soc. Am. **95**, 2728–2735.
Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Acoustic backscattering by Hawaiian lutjanid snappers. II. Broadband temporal and spectral structure," J. Acoust. Soc. Am. **114**, 2767–2774.

Au, W. W. L., and Hastings, M. C. (**2008**). *Principles of Marine Bioacoustics* (Springer-Verlag, New York).
Au, W. W. L., and Pawloski, D. (**1992**). "Cylinder wall thickness difference discrimination by an echolocating Atlantic bottlenose dolphin," J. Comp. Physiol., A **172**, 41–47.
Au, W. W. L., and Turl, C. W. (**1991**). "Material composition discrimination of cylinders at different aspect angles by an echolocating dolphin," J. Acoust. Soc. Am. **89**, 2448–2451.
Au, W. W. L., Benoit-Bird, K. J., and Kastelein, R. A. (**2007**). "Modeling the detection range of fish by echolocating bottlenose dolphins and harbor porpoises," J. Acoust. Soc. Am. **121**, 3954–3962.
Au, W. W. L., Schusterman, R., and Kersting, D. A. (**1980**). "Sphere-cylinder discrimination via echolocation by Tursiops truncatus," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 859–862.
Au, W. W. L., Ford, J. K. B., Horne, J. K., and Newman-Allman, K. A. (**2004**). "Echolocation signals of free-ranging killer whales (*Orcinus orca*) and modeling of foraging for chinook salmon (*Oncorhynchus tshawytscha*)," J. Acoust. Soc. Am. **56**, 1280–1290.
Aubauer, R., Au, W. W. L., Nachtigall, P. E., Pawloski, J. L., Pawloski, D. A., and DeLong, C. (**2000**). "Classification of electronically generated phantom targets by an Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **107**, 2750–2754.
Barta, R. E. (**1969**). "Acoustical pattern discrimination by an Atlantic bottlenosed dolphin," Naval Undersea Center, San Diego, CA.
Branstetter, B. K., Mecado, E., III, and Au, W. W. L. (**2007**). "Representing multiple discrimination cues in a computational model of the bottlenose dolphin auditory system," J. Acoust. Soc. Am. **122**, 2459–2468.
Busnel, R.-G., and Dziedzic, A. (**1967**). "Resultats metrologiques experimentaux de l'echolocaiton chez le *Phocaena phocaena* et leur comparison avec cues de cdrtaines chauves—souris," in *Animal Sonar Systems: Biology and Bionics*, edited by R. G. Busnel (Labaoratoire de Physiologie Acoustique, Jouy-en-Josas, France), Vol. **1**, pp. 307–335.
Delong, C., Au, W., Harley, H., Roitblat, H., and Pytka, L. (**2007**). "Human listeners provide insights into echo features used by dolphins to discriminated among objects," J. Comp. Psychol. **121**, 306–319.
Evans, W. W., and Powell, B. A. (**1967**). "Discrimination of different metallic plates by an echolocating delphinid," in *Animal Sonar Systems: Biology and Bionics*, edited by R. G. Busnel (Labaoratoire de Physiologie Acoustique, Jouy-en-Josas, France), pp. 363–382.
Foote, K. G. (**1980**). "Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackerel target strengths," J. Acoust. Soc. Am. **67**, 2084–2089.
Foote, K. G., and Ona, E. (**1985**). "Swimbladder cross sections and acoustic target strengths of 13 pollack and 2 saithe," Fiskeridir. Skr., Ser. Havunders. **18**, 1–57.
Ford, J. K. B., and Ellis, G. M. (**2006**). "Selective foraging by fish-eating killer whales *Orcinus orca* in British Columbia," Mar. Ecol.: Prog. Ser. **316**, 185–199.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Helweg, D. A., Au, W. W. L., Roitblat, H. L., and Nachtigall, P. E. (**1996**). "Acoustic basis for recognition of aspect-dependent three-dimensional targets by an echolocating bottlenose dolphin," J. Acoust. Soc. Am. **99**, 2409–2420.
Ibsen, S. D., Au, W. W. L., Nachtigall, P. E., Delong, C. D., and Breeze, M. (**2007**). "Changes in signal parameters over time for an echolocating Atlantic bottlenose dolphin performing the same target discrimination task," J. Acoust. Soc. Am. **122**, 2446–2450.
Johnson, C. S. (**1968**). "Masked tonal thresholds in the bottlenose porpoise," J. Acoust. Soc. Am. **44**, 965–967.
Kellogg, W. N. (**1958**). "Echo ranging in the porpoise," Science **128**, 982–988.
Lemonds, D. W. (**1999**). "Auditory filter shapes in an Atlantic bottlenose dolphin (*Tursiops truncates*)," Ph.D. dissertation, University of Hawaii at Manoa.
Madsen, P. T., Kerr, I., and Payne, R. (**2004**). "Echolocation clicks of two free-ranging delphinids with different food preferences: False killer whales (*Pseudorca crassidens*) and Risso's dolphin (*Grampus griseus*)," J. Exp. Biol. **207**, 1811–1823.
Nachtigall, P. E. (**1980**). "Odontocete echolocation performance on object size, shape and material," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 71–95.
Norris, K. S., Prescott, J. H., Asa-Dorian, P. V., and Perkins, P. (**1961**). "An experimental demonstrated echolocation behavior I the porpoise, *Tursiop truncates* (Montagu)," Biol. Bull. **120**, 163–176.
Pack, A. A., Herman, L. M., Hoffman-kuhnt, M., and Branstetter, B. K. (**2002**). "The object behind the echo: Dolphins (*Tursiops truncates*) perceive object shape globally through echolocation," Behav. Processes **58**, 1–26.
Patterson, R. D. (**1994**). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409–1418.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. H. (**1992**). *Complex Sounds and Auditory Images* (Pergamon, Oxford).
Popov, V. V., Supin, A. Ya., Ding, W., and Wang, K. (**2006**). "Nonconstant quality of auditory filters in the porpoises, *Phocoena phocoena and Neophocaena phocenoides* (Cetacea, Phocoenidae)," J. Acoust. Soc. Am. **119**, 3173–3180.
Reeder, B. D., Jech, J. M., and Stanton, T. K. (**2004**). "Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling," J. Acoust. Soc. Am. **116**, 747–761.
Schevill, W. E., and Lawrence, B. (**1956**). "Food-finding by a captive porpoise (*Tursiops truncates*)," Breviora (Mus. Comp. Zool, Harvard) **53**, 1–15.
Simmonds, J., and MacLennan, D. (**2005**). *Fisheries Acoustics: Theory and Practice*, 2nd ed. (Blackwell, Oxford, UK).
Slaney, M. (**1993**). "An efficient implementation of the Patterson-Holdsworth filter bank," Apple Technical Report No. 35, Advanced Technology Group, Apple Computer, Inc., Cupertino, CA.

Localization and tracking of phonating finless porpoises using towed stereo acoustic data-loggers

Songhai Li
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

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

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

(Received 3 November 2008; revised 4 May 2009; accepted 11 May 2009)

Cetaceans produce sound signals frequently. Usually, acoustic localization of cetaceans was made by cable hydrophone arrays and multichannel recording systems. In this study, a simple and relatively inexpensive towed acoustic system consisting of two miniature stereo acoustic data-loggers is described for localization and tracking of finless porpoises in a mobile survey. Among 204 porpoises detected acoustically, 34 individuals (~17%) were localized, and 4 of the 34 localized individuals were tracked. The accuracy of the localization is considered to be fairly high, as the upper bounds of relative distance errors were less than 41% within 173 m. With the location information, source levels of finless porpoise clicks were estimated to range from 180 to 209 dB re 1 μPa pp at 1 m with an average of 197 dB (N=34), which is over 20 dB higher than that estimated previously from animals in enclosed waters. For the four tracked porpoises, two-dimensional swimming trajectories relative to the moving survey boat, absolute swimming speed, and absolute heading direction are deduced by assuming the animal movements are straight and at constant speed in the segment between two consecutive locations.
 [DOI: 10.1121/1.3147507]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Nd [WA] Pages: 468–475

I. INTRODUCTION

Cetaceans are included as top trophic-level predators in food-chains of ecosystem (Baumgartner and Mate, 2003; Tynan, 2004). They occupy unique statuses in models of ecosystem dynamics. Researches on population status, ecology, behavior, and conservation of cetaceans in the wild are increasingly popular. Traditional visual observation methods for cetacean researches can only allow detection of a fraction of the animals present and observation of the surface behaviors, both due to the brief appearances of animals at the surface when breathing, and limited transparency of water. The weather condition, circadian pattern, and unintentional variation of efforts among observers cause additional biases in visual observation.

To aquatic life, cetaceans possess highly developed sound production and hearing capabilities (Herman, 1980). Furthermore, all investigated odontocetes possess a sophisticated echolocation system (Au, 1993). Small odontocetes, such as porpoises, frequently produce series of high-frequency echolocation clicks (i.e., click trains) for navigation, orientation, and prey capture (Au, 1993; Akamatsu *et al.*, 2005a, 2007). Therefore, it is not surprising that many researches of cetaceans focus on their acoustics, and passive acoustic methods are widely used for cetacean observation.

Passive acoustic methods in cetacean localization and observation can be distinguished to tagging acoustic systems, fixed acoustic systems, and mobile acoustic systems. Tagging acoustic systems with depth- and/or acceleration-meters have been widely used in the studies of humpback whales (Stimpert *et al.*, 2007), sperm whales (Johnson and Tyack, 2003; Zimmer *et al.*, 2003, 2005; Miller *et al.*, 2004a, 2004b), and even small odontocetes, such as porpoises (Akamatsu *et al.*, 2005a, 2005b, 2005c, 2007). The tagging acoustic systems do have enabled scientists to gain knowledge of underwater behaviors of cetaceans. However, the systems themselves can disrupt or alter the natural behaviors of the tagged animals, especially for the small odontocetes. Also the tagging procedures, in some of which the animals need to be captured (Akamatsu *et al.*, 2005b), are time consuming and difficult to implement.

Fixed acoustic systems, which may be left in stationary place for long time periods, are often used for monitoring of population status and ecological dynamics of cetaceans (Mellinger *et al.*, 2007). Recently, fixed acoustic systems with multiple hydrophone sensors, which compose arrays, are widely used for localization and behavior observation of cetaceans (Fox *et al.*, 2001; Au and Benoit-Bird, 2003; Wiggins, 2003; Kimura *et al.*, 2009). While these systems enable scientists to better understand the presence and seasonal occurrence patterns, as well as underwater acoustic activity, they are limited to be in small range. Furthermore, there are still many hurdles for the fixed systems to estimate abun-

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

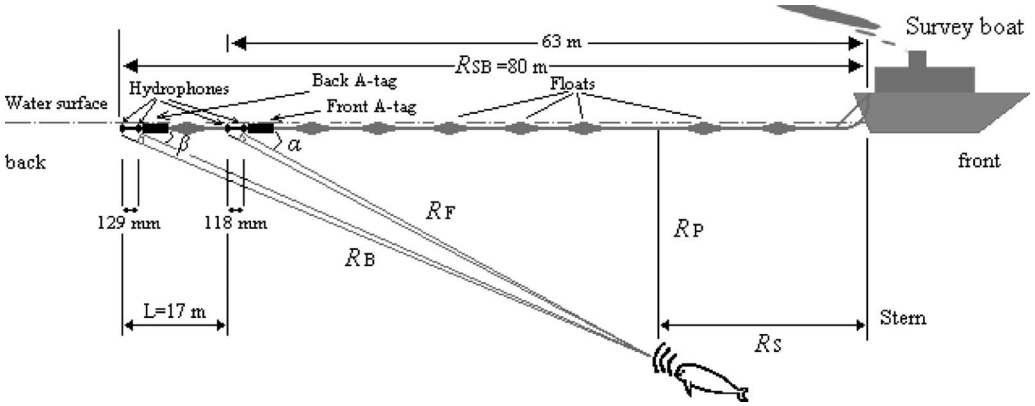


FIG. 1. A linear line array consists of two miniature stereo acoustic data-loggers (A-tags), which were towed 63 m (front A-tag) and 80 m (back A-tag) behind the survey boat, respectively. R_P, R_F, R_B, and R_S correspond to the distance of the phonating animal to the cruise line (i.e., perpendicular distance to the survey boat), the front A-tag, the back A-tag, and the stern of the survey boat along the cruise line, respectively.

dance of animals, which is the ultimate aim for ecological studies and management of the target animals (Mellinger *et al.*, 2007).

Mobile acoustic systems often consist of cabled hydrophones, which are towed behind a ship or affixed to a mobile platform to detect animals in a large area. The mobile systems are often used in joint visual and acoustic surveys, to detect animals with increased levels in accuracy. Experimentally, in joint visual-acoustic surveys, mobile acoustic systems can usually detect one to ten times as many cetacean groups as visual ones (Barlow and Taylor, 2005; Rankin *et al.*, 2007; Akamatsu *et al.*, 2008). However, for comparison between visual and acoustic observations to determine detection performance of each method, determination of number of animals traveling together, distance, and bearing angle to the animal/group are necessary. Previous attempts at acoustical location determination for cetaceans have generally used cabled hydrophone arrays and differences in time-of-arrival measurements (Miller and Tyack, 1998; Gillespie and Chappell, 2002). In the mobile acoustic systems, at least three hydrophones in a towed linear line array are necessary to determine the two-dimensional location of a phonating animal by evaluating the travel time differences of the input signal. These systems with long cables from the hydrophones to the recording devices on board are not easy to set up and handle on a moving platform. Multichannel recording and signal evaluation are also time consuming and difficult to implement because of the huge amount of data size.

This study describes a portable and operable acoustic system, which only requires two miniature acoustic data-loggers (A-tags; see below) detecting the high-frequency echolocation click events of odontocetes for localization and potential behavior observation of the animals in mobile survey. In each A-tag, there are two miniature acoustic sensors (i.e., hydrophones) with about 120 mm apart to record the travel time difference of each click. With the location information the source level of click signals is estimated. And also, the potential applications of this acoustic system are discussed.

II. MATERIALS AND METHODS

A. Subject and equipment

The subject is a freshwater subspecies of finless porpoises, living only in main stream and tributaries of the middle and lower reaches of Yangtze River and its conjoint large lakes, such as Poyang Lake and Dongting Lake. To document the population status of this subspecies, a joint visual-acoustic survey was performed between November and December 2006 in the main stream of the middle and lower reaches of Yangtze River (Akamatsu *et al.*, 2008; Zhao *et al.*, 2008).

In one of the survey boats, two miniature stereo acoustic data-loggers (A-tags; ML200-AS2, Marine Micro Technology, Saitama, Japan; Akamatsu *et al.*, 2008), which are 21 mm in diameter, less than 350 mm in length including the external hydrophones, and 72 g in weight, were towed 63 m (front A-tag) and 80 m (back A-tag) behind the boat, respectively, in a linear line array for localization and behavior observation of the porpoises (Fig. 1). Each A-tag contains a CPU (PIC18F6620, Microchip, USA) for system control and signal processing, a 128 Mbyte flash memory for data storage, a miniature high-frequency pulse event recorder, and a CR2 lithium battery cell, encased in a waterproof tube. The present A-tags are slightly modified from the previous model, but had identical signal processing (see Akamatsu *et al.*, 2005b). Each A-tag has two external hydrophones, apart each other with 118 and 129 mm for the front A-tag and back A-tag, respectively (Fig. 1). The hydrophone sensitivity is −201 dB re 1 V/μPa at 120 kHz (100–160 kHz within 5 dB), which is close to the dominant frequency of sonar signal of finless porpoises (Li *et al.*, 2005). An electronic band pass filter (55–235 kHz) is included to eliminate noise outside the frequency bands of porpoise sonar signals. Every 0.5 ms (i.e., using a 2 kHz sampling operation), the A-tags record and store the intensity of a received pulse and the travel time difference of each pulse to the two hydrophones with a resolution of 271 ns (one count in Fig. 2), which can be used to estimate the bearing angle to a sound source (Akamatsu *et al.*, 2008).

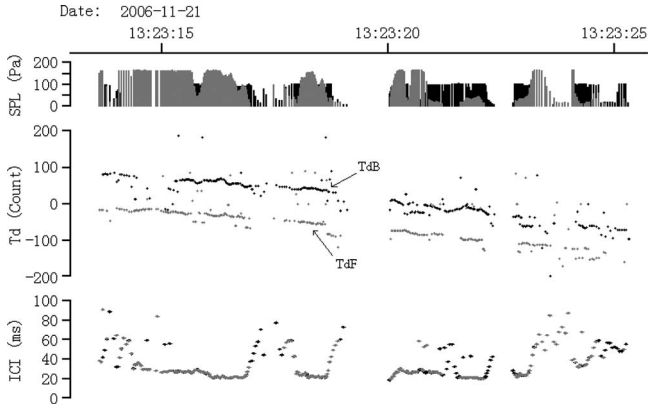


FIG. 2. Echolocation click trains from single porpoise passing by the two acoustic data-loggers (A-tags); the gray one corresponds to the front A-tag and the black one corresponds to the back A-tag. Top panel: The received SPL in pascals. Middle panel: The travel time difference of clicks (Td) in count (one count equals to travel time difference of 271 ns). Lower panel: ICI in milliseconds. Td_F and Td_B in the middle panel correspond to the travel time differences of clicks in the front A-tag and the back A-tag, respectively. Note that the trace of the travel time difference (Td) changes from positive to negative, corresponding to an individual passing from bow to stern relative to the data-loggers.

The A-tags were mounted on a towing cable, on which floats were placed at about 5 m interval to keep the cable to near the water surface (Fig. 1). To stabilize the position of data-loggers and to prevent them from swinging, a 5 m-length and 5 mm-diameter nylon rope was added behind the back A-tag.

B. Localization and tracking of phonating animals

Free ranging finless porpoises frequently produce series of high-frequency echolocation clicks (i.e., click trains), which usually contain over five to up to several hundreds of clicks. They produce click trains every 5 s in average (Akamatsu *et al.*, 2005a), with regular or gradual change in the sound intensity and interclick interval (ICI) changing typically between 20 and 70 ms (Akamatsu *et al.*, 1998). These characteristics can distinguish porpoises click trains from the noise of background, survey boat, and other cargo ships passing nearby, which have randomly changing ICIs and sound intensities. During the survey, the speed of the survey boat was kept to be approximately 15 km/h, much faster than the average swimming speed of finless porpoises (4.3 km/h; Akamatsu *et al.*, 2002). This means the porpoises will always pass the survey boat from bow to stern. When passing animals vocalize, the travel time differences of porpoise clicks to the two hydrophones in one A-tag that corresponded to the bearing angle will change from positive to negative (Fig. 2).

In theory, when one vocalizing animal passes by the A-tag, there would be one smooth gradual change trace of travel time differences, which changed from positive to negative (Fig. 2). And when the passing animals are two or more with both or all vocalizing, there would be two or more traces with gradual change in the travel time differences when two or more animals separated each other outside of the resolution of the acoustical location by A-tag. The number of independent traces could be used for counting of passing animals (see Akamatsu *et al.*, 2008).

When click trains of porpoises were received by both the front A-tag and back A-tag (Fig. 1), there would be two parallel traces of travel time differences recorded by front A-tag (Td_F) and back A-tag (Td_B), respectively (Fig. 2). The two parallel traces of travel time difference (Td), which correspond to two independent bearing angles to the phonating animal, could be used for localization and tracking of phonating animals. By measuring the travel time differences Td_F and Td_B, and taking advantage of the fact that the distance between phonating animal and the A-tags was significantly longer than the distances between two hydrophones of each A-tag and the opening angles from the animal to two hydrophones in each A-tag could be neglected, the two bearing angles α and β of phonating animals to the two A-tags could be determined by using the following equations (see Fig. 1):

$$\cos\alpha = \frac{Td_F}{Td_{F\,\max}}, \tag{1}$$

$$\cos\beta = \frac{Td_B}{Td_{B\,\max}}, \tag{2}$$

where Td_F is travel time difference of porpoise click to the two hydrophones of the front A-tag in count (one count equals to travel time difference of 271 ns), Td_B is the travel time difference recorded by the back A-tag in count, and $Td_{F\,\max}$ and $Td_{B\,\max}$ correspond to maximum time difference when the sound came from 0° for each A-tag. Since the distances between two hydrophones in front A-tag and back A-tag were 118 and 129 mm (Fig. 1), respectively, $Td_{F\,\max}$ and $Td_{B\,\max}$ in count can be calculated by $118\times10^{-3}/c/271\times10^{-9}$ and $129\times10^{-3}/c/271\times10^{-9}$, respectively, where c is the sound speed in water, which was calculated from the Medwin equation (Medwin, 1975) to be 1465 m s^{-1}, by setting the salinity (<1 ppt) and temperature (15 °C) measurements made in the survey season in the Yangtze River. Using triangulation and Eq. (3), the distances of the phonating animals to the cruise line R_P (i.e., perpendicular distance), the front A-tag R_F, the back A-tag R_B, and the stern of the survey boat along the cruise line R_S (see Fig. 1) are given by Eqs. (4)–(7), respectively,

$$\frac{R_P}{\tan(\pi-\alpha)} + \frac{R_P}{\tan\beta} = L, \tag{3}$$

$$R_P = \frac{L\times\sin\alpha\sin\beta}{\sin(\alpha-\beta)}, \tag{4}$$

$$R_F = \frac{L\times\sin\beta}{\sin(\alpha-\beta)}, \tag{5}$$

$$R_B = \frac{L\times\sin\alpha}{\sin(\alpha-\beta)}, \tag{6}$$

$$R_S = \frac{L\times\sin\alpha\times\cos\beta}{\sin(\alpha-\beta)} - R_{\mathrm{SB}}, \tag{7}$$

where L is the distance between the two A-tags, which is 17 m (Fig. 1); and R_{SB} is the distance of the back A-tag to the

stern of the survey boat, which is 80 m (Fig. 1). R_P and R_S would determine the two-dimensional location of the phonating animals relative to the stern of the survey boat. R_F and R_B would be used for estimation of source levels of porpoise clicks.

In practice, while the traces of the time differences changed from positive to negative, the traces did not ideally change smoothly and gradually, but showed some fluctuation and bounce (Fig. 2). Both ambiguity at the trigger point in the waveform of porpoise clicks between two hydrophones in one A-tag and swing of A-tags due to water flow can have contributed the fluctuation of coordination. When the trigger points of two hydrophones are in different cycle of a click waveform, the ambiguity would be over 1 cycle, corresponding to over 8.5 μs in time (see Li *et al.*, 2005). This would reduce an increase of over 8.5 μs (corresponds to 31 counts in Fig. 2) in absolute value of travel time difference (Td). When the trigger points are in a same cycle, the ambiguity would be less than 1/4 cycle according to the characteristics of click waveforms, which corresponds to a less than 2.2 μs (8 counts in Fig. 2) increase of absolute value of Td. In localization of animals, to avoid the effect of the ambiguity and swing, only the click trains containing at least three consecutive clicks with Td change less than 1.4 μs (i.e., five counts in Fig. 2) in both the two A-tags were selected. The average of Td of the three consecutive clicks would be used for ultimate localization of the phonating animals. In addition, the over 1 cycle ambiguity at the trigger point between two hydrophones, which brought on an increase of over 31 counts in the absolute value of Td, could be kept away from the localization of phonating animals by selecting click cluster (over 3 clicks) with Td closer to 0 when there was a change of over 31 counts in Td among click clusters. For each animal identified by acoustics, if the localization could be determined for more than two times, the animal would be tracked acoustically.

C. Behavior observation

Once phonating animals were tracked, i.e., were localized for more than two times, the two-dimensional swimming trajectories of the animals relative to the survey boat were reconstructed by assuming the movements were straight and at constant speed in the segment between two consecutive locations. In the meantime, when the time duration of the given segment is longer than 0.5 s, the absolute speed and absolute heading direction in the segment were determined. If the authors assume a porpoise at a perpendicular distance to the survey boat R_{P1}, longitudinal distance to the stern of the survey boat along the cruise line R_{S1}, and instantaneous time t_1 heads straight to a perpendicular distance R_{P2}, longitudinal distance R_{S2}, and instantaneous time t_2, at a constant speed V, then the perpendicular and longitudinal speeds of the animal V_P and V_L could be given by the following equations:

$$V_P = (R_{P2} - R_{P1})/(t_2 - t_1), \tag{8}$$

$$V_L = V_{SB} - (R_{S2} - R_{S1})/(t_2 - t_1), \tag{9}$$

where V_{SB} is the speed of the survey boat, which was monitored by hand-held GPS and could be considered constant in a short segment. Thus, the porpoise speed $V = \sqrt{V_P^2 + V_L^2}$ and heading direction could be estimated when crossing the given segment.

D. Source level measurement

The received intensity of porpoise click by the A-tags is termed sound pressure level (SPL). The source level (SL) defined as the SPL at 1 m from phonating porpoise on its acoustic axis could be estimated by Eq. (10) by assuming spherical spreading, which is typical of spreading observed in dolphin and porpoise sonar (Au, 1993) and using the above calculated distances between the phonating animals and A-tags.

$$\mathrm{SL} = \mathrm{SPL} + 20 \log R + \lambda R, \tag{10}$$

where R is the distance between phonating animal and A-tags, and λ is the frequency-dependent absorption coefficient of water in dB/m. In this case, it was estimated to be $\sim$0.004 dB/m in the freshwater at 15 °C and 125 kHz, the peak frequency of finless porpoise (Li *et al.*, 2005), with Fisher and Simmons' model (Fisher and Simmons, 1977).

Since dolphins and porpoises emit echolocation clicks directionally (Au, 1993) and it is very difficult, or almost impossible, to accurately determine whether the phonating animal points its acoustic axis at one of the A-tags with the present system, the received intensities of clicks (i.e., SPLs) were very likely acquired from both directly on and off the axis of porpoise sonar. In this paper, a term "apparent source levels" (ASLs) was introduced, which equals the sound intensity at 1 m from a directional source in an unknown direction (Villadsgaard *et al.*, 2007). Echolocation signals acquired from off the beam axis are lower in sound levels, relative to the source signals (Au, 1993). The directionality of porpoise sonar could have resulted in an underestimation of the on-axis source levels. The present ASLs should be regarded as conservative estimates of the true source levels.

In the determination of source levels, three policies were adopted: (1) only the click with maximum intensity in one click train was selected for level estimation; (2) source level was estimated by the front A-tag and back A-tag, respectively, and the higher one was selected as the final value; and (3) to maintain the independence of data, only one SL was estimated for each localized and/or tracked animal.

III. RESULTS

A. Localization and tracking of phonating animals

In the whole survey, the acoustic system with two towed A-tags, deploying about 120 h, detected 204 porpoises, in which 34 individuals ($\sim$17%) were localized acoustically, based on the selection criteria of travel time differences (Td). Figure 3(b) shows the two-dimensional locations of the 34 localized porpoises relative to the stern of the moving survey boat. Due to the symmetry of the linear line array consisting of A-tags, the animals can be on either side of the cruise line.

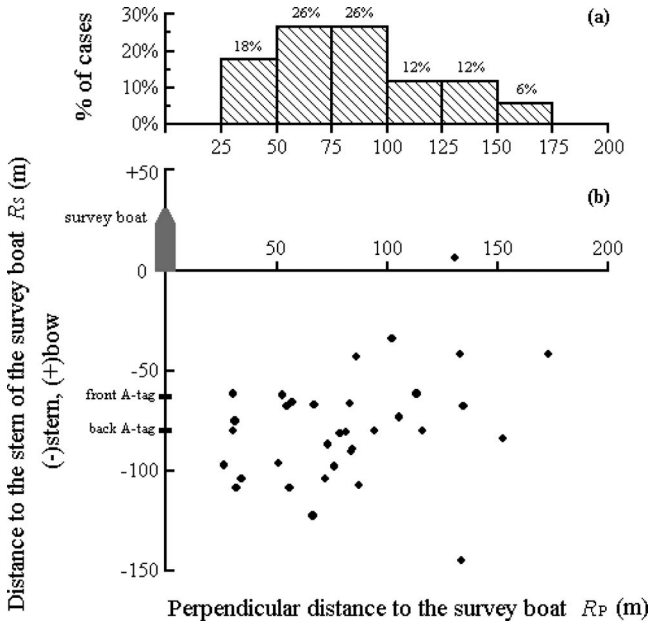


FIG. 3. Two-dimensional localization of phonating porpoises. (a) Histogram of distribution of perpendicular distance to the survey boat and (b) two-dimensional locations of 34 localized porpoises relative to the stern of the moving survey boat.

Most of the localizations distribute around the A-tag array within 150 m. The maximum detected perpendicular distance to the survey boat is 173 m, and 32 of 34 localizations have perpendicular distances (over 94%) distributing between 25 and 150 m [Fig. 3(a)]. No animals were detected within 25 m of the cruise line.

Among the localized individuals, four were tracked (i.e., localized acoustically for more than two times). The four tracked individuals had been localized for three, six, eight, and nine times with time spans of 10, 4, 10, and 9 s, respectively (Fig. 4). During these tracks the porpoises continuously emitted echolocation click trains detected by both the A-tags (see Fig. 2).

B. Swimming speeds and heading directions of animals

For the four tracked animals, the two-dimensional swimming trajectories of animals relative to the moving survey boat were reconstructed (Fig. 4). In Fig. 4, the absolute speed (and not relative) V, perpendicular-oriented speed V_P (x axis in Fig. 4), and longitudinal-oriented speed V_L (y axis in Fig. 4) of the animals in each segment between two consecutive localizations are presented along with the two-dimensional trajectories in format of V (V_P,V_L), i.e., the numeral outside the parenthesis is the V, the former numeral inside the parenthesis is the V_P, and the latter numeral inside the parenthesis is the V_L. The marks "+" and "−" represent the directions of V_P and V_L along the x and y axes (see the upper right corner in Fig. 4). The absolute heading directions of the animals in each segment are sketched by arrowheads in Fig. 4. The

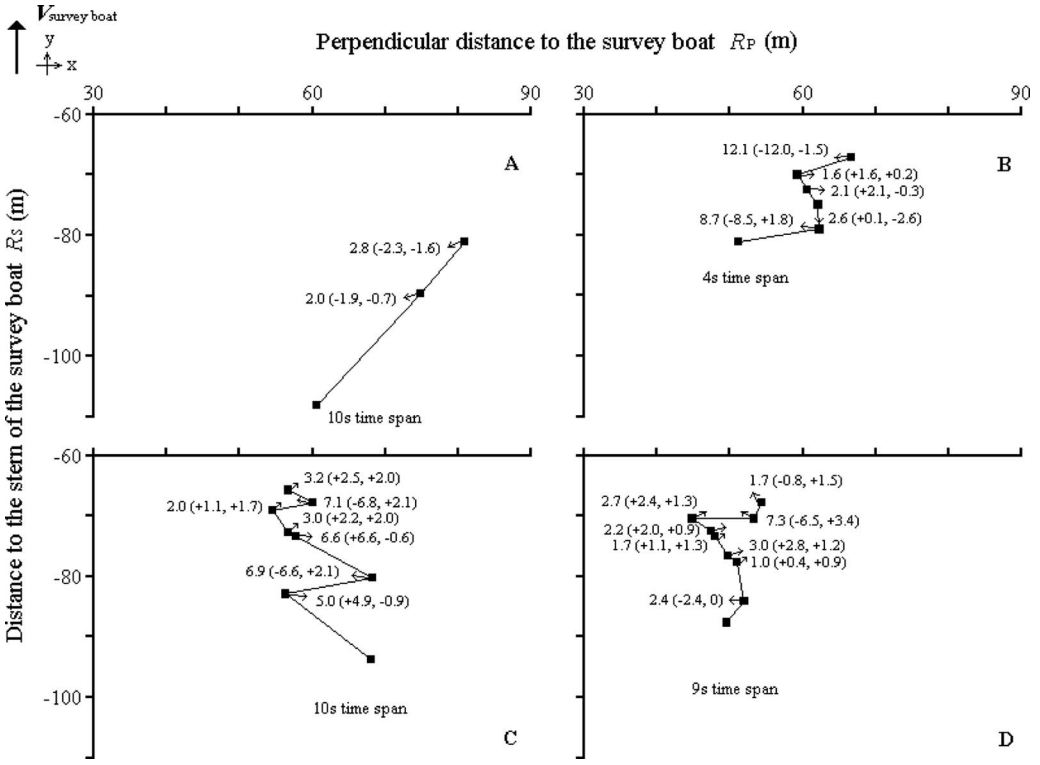


FIG. 4. Two-dimensional swimming trajectories of four animals relative to the moving survey boat. In each segment between two consecutive locations, the animal movements were assumed straight and at constant speed. Along with the two-dimensional trajectories, absolute speed (the numeral outside the parenthesis), perpendicular-oriented speed (the former numeral inside the parenthesis), and longitudinal-oriented speed (the latter numeral inside the parenthesis) are indicated. The marks "+" and "−" represent the directions of perpendicular-oriented speed and longitudinal-oriented speed along the x and y axes (see the up right corner). The absolute heading directions of the animals in each segment are sketched by arrowheads.

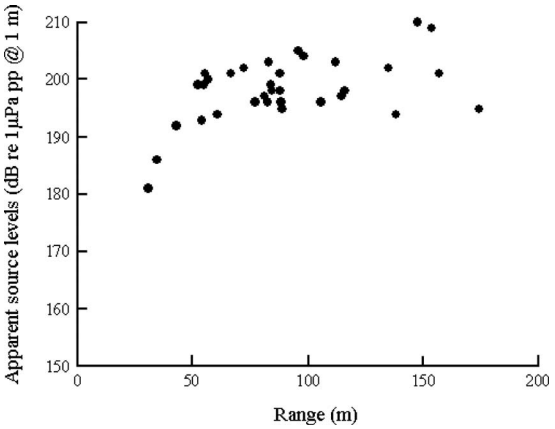


FIG. 5. The calculated ASL of porpoise clicks as a function of the distances between localized animals and A-tags.

fastest absolute speed V of the animals is 12.1 m s^{-1} [Fig. 4(b)], and most of the speed V (16 of 22, i.e., ~73%) are between 1.0 and 3.2 m s^{-1}.

C. Source levels of clicks

ASLs from 180 to 209 dB re 1 μPa pp at 1 m were estimated with an average of 197 dB (N=34). A scatter plot of ASLs as a function of the distances between localized animals and A-tags is presented in Fig. 5.

IV. DISCUSSION

A. Error estimation

Assuming the sound speed calculated from the Medwin equation (Medwin, 1975) is valid and constant, the accuracy of the localization in the present study mainly lies on errors of travel time differences (Td) of porpoise clicks to the two hydrophones of each A-tag. The errors of Td could be both from ambiguity at the trigger point between two hydrophones and swing of A-tags due to water flow. By selecting Td based on the criteria described above, and averaging Td of three consecutive clicks, the authors presume that the errors of Td have been well controlled less than five counts, corresponding to ~1.4 μs. Also, the measurement errors of distances between two A-tags could contribute minor effect to the accuracy of the localization. The accuracy of the present localization can be estimated with the total error differential of the distances, which is the sum of the partial derivatives of all variables multiplied by the error bounds of the variables [see Eqs. (11)–(13)]. The location errors were evaluated by using the relative distance errors of R_P, R_F, and R_B, which represent the distance of phonating animal to the cruise line (i.e., perpendicular distance to the survey boat), the front A-tag, and the back A-tag, respectively. The relative distance error was defined as the quotient of the total error differential of distance and the estimated distance as follows:

$$\frac{\Delta R_P}{R_P} = \frac{|\partial R_P/\partial Td_F| \cdot |\Delta Td_F| + |\partial R_P/\partial Td_B| \cdot |\Delta Td_B| + |\partial R_P/\partial L| \cdot |\Delta L|}{R_P}, \tag{11}$$

$$\frac{\Delta R_F}{R_F} = \frac{|\partial R_F/\partial Td_F| \cdot |\Delta Td_F| + |\partial R_F/\partial Td_B| \cdot |\Delta Td_B| + |\partial R_F/\partial L| \cdot |\Delta L|}{R_F}, \tag{12}$$

$$\frac{\Delta R_B}{R_B} = \frac{|\partial R_B/\partial Td_F| \cdot |\Delta Td_F| + |\partial R_B/\partial Td_B| \cdot |\Delta Td_B| + |\partial R_B/\partial L| \cdot |\Delta L|}{R_B}. \tag{13}$$

In the present error analysis, the error bounds of $|\Delta L|$ =0.1 m and $|\Delta Td|$=5 count for travel time differences Td_F and Td_B are assumed. The error estimations of location are shown as scatter plots of $\Delta R_P/R_P$ to R_P, $\Delta R_F/R_F$ to R_F, and $\Delta R_B/R_B$ to R_B in Fig. 6. All the relative distance errors $\Delta R_P/R_P$, $\Delta R_F/R_F$, and $\Delta R_B/R_B$ depend on the relevant distances of the porpoise in a similar behavior and level, and tend to increase with increasing distances. The highest relative distance error is 41% for the present localization, and when the distances are within 100 m, the relative distance errors are even less than 30%.

The absorption coefficient λ is very low (only ~0.004 dB/m) in the present condition, and the uncertainty in its calculation does not contribute much to the source level measurement. By assuming spherical spread, the SL measurement error ΔSL can be expressed by

$$\Delta\mathrm{SL} = |20 \log(1 \pm \Delta R/R)|. \tag{14}$$

Considering a highest relative distance error of 41%, the upper bound of SL measurement error would be ~4.6 dB.

B. Application

The most important and powerful feature of this localization method is its potential application in *distance sampling* methodology, which was originally developed for visual survey to investigate population size of animals (Buckland *et al.*, 1993). In distance sampling, for reliable estimation of absolute density (i.e., number of animals in unit area), an accurate measurement of distance between the animal and the survey cruise line (i.e., perpendicular distance) is essential (Buckland *et al.*, 1993). The present localization using a towed linear line array consisting of two

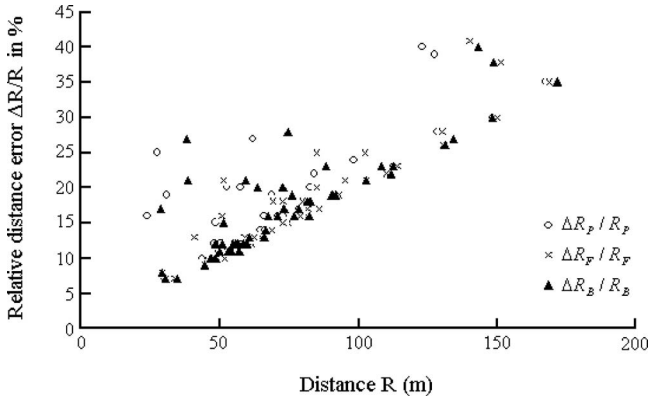


FIG. 6. Relative error in distance estimation of R_P, R_F, and R_B, which correspond to the distance of the phonating animal to the cruise line (i.e., perpendicular distance to the survey boat), the front A-tag, and the back A-tag, respectively.

A-tags is able to localize the phonating animals with a relative distance error less than 41% within 173 m, and when the distance is within 100 m, the relative error is even less than 30% (Fig. 6). The distance estimation by this acoustic localization method could be considered fairly accurate. The accurate distance estimation along with a high animal detection capability (see Akamatsu *et al.*, 2008) contributes the possibility for this localization method to apply in distance sampling methodology. A successful application of the towed acoustic system in distance sampling raises great perspectives for conduction of moving survey at regular intervals to monitor population status and estimate population size of selected species, such as finless porpoises, in long-term base. However, since the localization method can only localize animals in two dimensions, the methodology is not directly applicable to a deep water environment, where depth component cannot be neglected.

A second application of the present acoustic localization system is the acoustic tracking and behavior observation of animals in mobile survey and in large area. This is very useful for evaluation of habitat selection and ship or boat effect on behavior of the animals. Usually, acoustic tracking and behavior observation of marine mammals were done using fixed hydrophone arrays (Fox *et al.*, 2001; Au and Benoit-Bird, 2003; Wiggins, 2003), which were restricted in local area. In the present mobile survey, four finless porpoises were successfully tracked in their natural habitat. Two-dimensional swimming trajectories relative to the moving survey boat, absolute swimming speed V, and absolute heading direction of the tracked animals were deduced by assuming the animal movements were straight and at constant speed in the segment between two consecutive locations (Fig. 4). According to the distribution of the values of speed V (Fig. 4), they could be qualitatively divided into two subsets. One is between 1.0 and 3.2 m s^{-1} with an average of 2.1 m s^{-1}, and the other one is between 5.0 and 12.1 m s^{-1} with an average of 9.0 m s^{-1}. The former is slightly higher than the speed measured by Yang and Chen (1996) when animals were traveling, and Akamatsu *et al.* (2002). However, it should be noticed that the animals in Yang and Chen, 1996 and Akamatsu *et al.*, 2002 were living in a stagnant water environment, while the animals here are living in a running water environment. The latter speed is obviously higher than the one measured by Akamatsu *et al.* (2002), whereas, it is comparable to the speed measured by Yang and Chen (1996) when animals were in fright. Figures 4(b)–4(d) showed that the animals changed their heading direction frequently, and in most of the cases when there was an obvious change in heading direction between two conjoint segments, the speed V of the animals was changed to be very high, which was over 5 m s^{-1}. One explanation is the higher speed implies that the animals were trying to move away from the survey boat. Alternatively, the higher speed could simply be artifact due to the errors in the location determination of the animals.

A third application of this localization system is the estimation of SLs of porpoise or dolphin clicks in the wild. This parameter is very important for studies of sonar and social behaviors of these animals. Previous researches on SLs of odontocete clicks were mainly for animals in captivity (Au, 1993) or in small enclosed waters (Li *et al.*, 2006), and might not substantially represent the SLs produced by animals in their natural habitat.

In this study, ASLs from 34 located finless porpoises in their natural habitat are estimated. The ASLs are over an order of magnitude higher than those reported for this species in an enclosed waters—Tongling Reserve, which is only 1600 m in length and 80–220 m in width (see Li *et al.*, 2006). The ASLs of 180–209 dB with an average of 197 dB re 1 μPa pp are also much higher than previous reports on other porpoise species, which were usually 160–170 dB re 1 μPa pp (Møhl and Andersen, 1973; Awbrey *et al.*, 1979). For other odontocete species, such as bottlenose dolphin (*Tursiops truncatus*) and beluga (*Delphinapterus leucas*), it was observed that the same individual in open bay was able to produce signals about 40 dB more intense than the signals produced when it was in captivity (Au, 1993). Recent field recordings of harbor porpoises (*Phocoena phocoena*) also indicated that the ASLs of harbor porpoise clicks could be up to 205 dB with an average of 191 dB re 1 μPa pp (Villadsgaard *et al.*, 2007). Probably, the flexibility in SLs of sonar signals depending on environments is not unique for dolphins, but also for porpoises.

V. CONCLUSIONS

The use of the present towed acoustic system consisting of two miniature stereo acoustic data-loggers (A-tags) provided a simple and relatively inexpensive way to acquire valuable information on odontocete location, two-dimensional moving trajectory, behavior, and sound SLs in moving survey. The localization method with the upper bound of relative distance error less than 41% within 170 m could be considered to be fairly accurate. This gives the towed acoustic system a potential in the application of distance sampling methodology, where accurate distance estimation is essential, to calculate absolute densities of selected animals in shallow water environment.

ACKNOWLEDGMENTS

The authors thank K. Sasamori, S. Dong, J. Cheng, Y, Zhou, L. A. Barrett, M. Richlen, R. L. Pitman, B. S. Stewart, R. R. Reeves, B. Taylor, S. T. Turvey, J. R. Brandon, X. Wang, and Z. Wei for their assistance in the wild survey. This research was supported by National Basic Research Program of China (Grant No. 2007CB411600), the Chinese National Natural Science Foundation (Grant No. 30730018), the Ocean Park Conservation Foundation of Hong Kong (OPCFHK), the President Fund of the Chinese Academy of Sciences, Special Funds for Presidential Scholarships of the Chinese Academy of Sciences (Grant No. 082Z01), and Research and Development Program for New Bio-industry Initiatives, Japan. Sponsorship of the survey was provided by Baiji.org Foundation, SeaWorld-Busch Gardens Conservation Fund, Budweiser Wuhan, Anhueser-Busch Inc., SGS, DEZA, BAFU, Manfred Hermsen Stiftung Foundation, U.S. NOAA Fisheries, Hubbs-SeaWorld Research Institute, and Wuhan Baiji Conservation Foundation.

Akamatsu, T., Matsuda, A., Suzuki, S., Wang, D., Wang, K., Suzuki, M., Muramoto, H., Sugiyama, N., and Oota, K. (**2005b**). "New stereo acoustic data logger for tagging on free-ranging dolphins and porpoises," Mar. Technol. Soc. J. **39**, 3–9.

Akamatsu, T., Teilmann, J., Miller, L. A., Tougaard, J., Dietz, R., Wang, D., Wang, K., Siebert, U., and Naito, Y. (**2007**). "Comparison of echolocation behaviour between coastal and riverine porpoises," Deep-Sea Res., Part II **54**, 290–297.

Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (**1998**). "Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*) and bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **104**, 2511–2516.

Akamatsu, T., Wang, D., and Wang, K. (**2005c**). "Off-axis sonar beam pattern of free-ranging finless porpoises measured by a stereo pulse event data logger," J. Acoust. Soc. Am. **117**, 3325–3330.

Akamatsu, T., Wang, D., Wang, K., Li, S., Dong, S., Zhao, X., Barlow, J., Stewart, B. S., and Richlen, M. (**2008**). "Estimation of the detection probability for Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*) with a passive acoustic method," J. Acoust. Soc. Am. **123**, 4403–4411.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005a**). "Biosonar behaviour of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

Akamatsu, T., Wang, D., Wang, K., Wei, Z., Zhao, Q., and Naito, Y. (**2002**). "Diving behavior of freshwater finless porpoises (*Neophocaena phocaenoides*) in an oxbow of the Yangtze River, China," ICES J. Mar. Sci. **59**, 438–443.

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

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature (London) **423**, 861–863.

Awbrey, F. T., Norris, J. C., Hubbard, A. B., and Evans, W. E. (**1979**). "The bioacoustics of the Dall's porpoise-salmon drift net interaction," Hubbs/Sea World Research Institute Technical Report, 79-120.

Barlow, J., and Taylor, B. L. (**2005**). "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21**, 429–445.

Baumgartner, M. F., and Mate, B. R. (**2003**). "Summertime foraging ecology of North Atlantic right whales," Mar. Ecol.: Prog. Ser. **264**, 123–135.

Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (**1993**). *Distance Sampling: Estimating Abundance of Biological Populations* (Chapman and Hall, London).

Fisher, F. H., and Simmons, V. P. (**1977**). "Sound absorption in sea water," J. Acoust. Soc. Am. **62**, 558–564.

Fox, C. G., Matsumoto, H., and Lau, T. K. A. (**2001**). "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," J. Geophys. Res. **106**, 4183–4206.

Gillespie, D, and Chappell, O (**2002**). "An automatic system for detecting and classifying the vocalizations of harbour porpoisesBioacoustics **13**, 37–61.

Herman, L. M. (**1980**). *Cetacean Behavior* (Wiley, New York).

Johnson, M., and Tyack, P. (**2003**). "A digital recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.

Kimura, S., Akamatsu, T., Wang, K., Wang, D., Li, S., and Dong, S. (**2009**). "Comparison of stationary acoustic monitoring and visual observation of finless porpoises," J. Acoust. Soc. Am. **125**, 547–553.

Li, S., Wang, D., Wang, K., and Akamatsu, T. (**2006**). "Sonar gain control in echolocating finless porpoises (*Neophocaena phocaenoides*) in an open water," J. Acoust. Soc. Am. **120**, 1803–1806.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*)," J. Acoust. Soc. Am. **117**, 3288–3296.

Medwin, H. (**1975**). "Speed of sound in water: A simple equation for realistic parameters," J. Acoust. Soc. Am. **58**, 1318–1319.

Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (**2007**). "An overview of fixed passive acoustic observation methods for cetacean," Oceanogr. **20**, 36–45.

Miller, P., Johnson, M., and Tyack, P. (**2004a**). "Sperm whale behavior indicates the use of echolocation click buzzes 'creaks' in prey capture," Proc. R. Soc. London, Ser. B **271**, 2239–2247.

Miller, P., Johnson, M., Tyack, P., and Terray, E. (**2004b**). "Swimming gaits, passive drag and buoyancy of diving sperm whales *physeter macrocephalus*," J. Exp. Biol. **207**, 1953–1967.

Miller, P., and Tyack, P. (**1998**). "A small towed beamforming array to identify vocalizing resident killer whales (*Orcinus orca*) concurrent with focal behavioral observations," Deep-Sea Res., Part II **45**, 1389–1405.

Møhl, B., and Andersen, S. (**1973**). "Echolocation: High-frequency component in the click of the harbour porpoise (*Phocoena ph.* L.)," J. Acoust. Soc. Am. **54**, 1368–1372.

Rankin, S., Norris, T. F., Smultea, M. A., Oedekoven, C., Zoidis, A. M., Silva, E., and Rivers, J. (**2007**). "A visual sighting and acoustic detections of Minke whales, Balaenoptera acutoroshata (cetacea: balaenopteridae), in nearshore Hawaiian waters," Pac. Sci. **61**, 395–398.

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*)," Biol. Lett. **3**, 467–470.

Tynan, C. T. (**2004**). "Cetacean populations on the SE Bering Sea shelf during the late 1990s: Implications for decadal changes in ecosystem structure and carbon flow," Mar. Ecol.: Prog. Ser. **272**, 281–300.

Villadsgaard, A., Wahlberg, M., and Tougaard, J. (**2007**). "Echolocation signals of wild harbour porpoises, Phocoena phocoena," J. Exp. Biol. **210**, 56–64.

Wiggins, S. M. (**2003**). "Autonomous acoustic recording packages (ARPs) for long-term monitoring of whale sounds," Mar. Technol. Soc. J. **37**, 13–22.

Yang, J., and Chen, P. (**1996**). "Movement and behavior of finless porpoise (Neophocaena phocaenoides asiaeorientialis) at Swan oxbow, Hubei province," Acta Hydrobiol. Sin. **20**, 32–40.

Zhao, X., Barlow, J., Taylor, B. L., Pitman, R. L., Wang, K., Wei, Z., Stewart, B. S., Turvey, S. T., Akamatsu, T., Reeves, R. R., and Wang, D. (**2008**). "Abundance and conservation status of the Yangtze finless porpoise in the Yangtze River, China," Biol. Conserv. **141**, 3006–3018.

Zimmer, W., Johnson, M., D'Amico, A., and Tyack, P. (**2003**). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*physeter macrocephalus*)," IEEE J. Ocean. Eng. **28**, 13–28.

Zimmer, W., Tyack, P., Johnson, M., and Madsen, P. (**2005**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bentthorn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485.

Underwater hearing sensitivity of harbor seals (*Phoca vitulina*) for narrow noise bands between 0.2 and 80 kHz

Ronald A. Kastelein,[a] Paul Wensveen, and Lean Hoek
Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands

John M. Terhune
Department of Biology, The University of New Brunswick, P.O. Box 5050, Saint John, New Brunswick E2L 4L5, Canada

(Received 16 December 2008; revised 28 March 2009; accepted 17 April 2009)

The underwater hearing sensitivities of two 1.5-year-old female harbor seals were quantified in a quiet pool built specifically for acoustic research, by using a behavioral psychoacoustic technique. The animals were trained to respond when they detected an acoustic signal and not to respond when they did not ("go/no-go" response). Fourteen narrowband noise signals (1/3-octave bands but with some energy in adjacent bands), at 1/3-octave center frequencies of 0.2–80 kHz, and of 900 ms duration, were tested. Thresholds at each frequency were measured using the up-down staircase method and defined as the stimulus level resulting in a 50% detection rate. Between 0.5 and 40 kHz, the thresholds corresponded to a 1/3-octave band noise level of ~60 dB re 1 μPa (SD $\pm$ 3.0 dB). At lower frequencies, the thresholds increased to 66 dB re 1 μPa and at 80 kHz the thresholds rose to 114 dB re 1 μPa. The 1/3-octave noise band thresholds of the two seals did not differ from each other, or from the narrowband frequency-modulated tone thresholds at the same frequencies obtained a few months before for the same animals. These hearing threshold values can be used to calculate detection ranges of underwater calls and anthropogenic noises by harbor seals.
 [DOI: 10.1121/1.3132522]

PACS number(s): 43.80.Lb, 43.80.Nd [WA] Pages: 476–483

I. INTRODUCTION

The harbor seal (*Phoca vitulina*) has an extensive geographic distribution in coastal regions of temperate areas of the Northern Hemisphere. It leads an amphibious life, resting and pupping on land, but migrating, foraging, and carrying out courtship under water (Burns, 2002). The vocalizations of harbor seals have been described as growls and short broadband pulsed calls (Hanggi and Schusterman, 1994; Van Parijs and Kovacs, 2002). During the breeding season, territorial males produce underwater vocalizations that likely play a role in male-male competition and mate attraction (Van Parijs and Kovacs, 2002). In some locations, males establish acoustic display areas (Hayes *et al.*, 2004). To estimate the range over which a male's calls can be detected, information is needed about the source level of the call, the local transmission losses over distance, and the detection threshold of the listening seal.

Many human activities create significant underwater noise which may interfere with seals' abilities to hear ecologically important sounds. Most anthropogenic underwater sounds are not tonal, but consist of noise of various bandwidths, or of a combination of broadband noise and tonal sounds (Richardson *et al.*, 1995).

In nature, biotic and abiotic noises can be biologically important signals or maskers that reduce an animal's ability to detect biological signals. The underwater hearing sensitivity of harbor seals has been measured for tonal signals (Møhl, 1968; Terhune, 1988; Turnbull and Terhune, 1993; Kastak and Schusterman, 1998; Southall *et al.*, 2005; Kastelein *et al.*, 2009) and for frequency swept tones (Turnbull and Terhune, 1994), but not for narrow bandwidth noises. For human adults with normal hearing, narrowband noise thresholds (1/3-octave) are very similar to pure-tone thresholds (Simon and Northern, 1966; Cox and McDaniel, 1986). Also, using narrowband noises as stimuli in test pools has the great advantage over the use of tonal signals that more stable received sound pressure levels (SPLs) can be created at the location of the test animal's head, as noise bands cause less interference than tonal signals and do not create standing waves.

It is not clear how the harbor seal's hearing for noise bands compares to that for tonal signals, however. Therefore, the aim of the present study was to determine the absolute hearing thresholds of two harbor seals for narrow noise bands and to compare the hearing thresholds with those obtained only a few months before for tonal signals in the same animals (Kastelein *et al.*, 2009).

II. MATERIALS AND METHODS

A. Study animals

The study animals were two female harbor seals (identified as 01 and 02). During the study they aged from 18 to 21 months old and their bodyweight was around 40 kg. More information about the study animals can be found in Kastelein *et al.*, 2009.

[a]Author to whom correspondence should be addressed. Electronic mail: researchteam@zonnet.nl

 0001-4966/2009/126(1)/476/8/$25.00

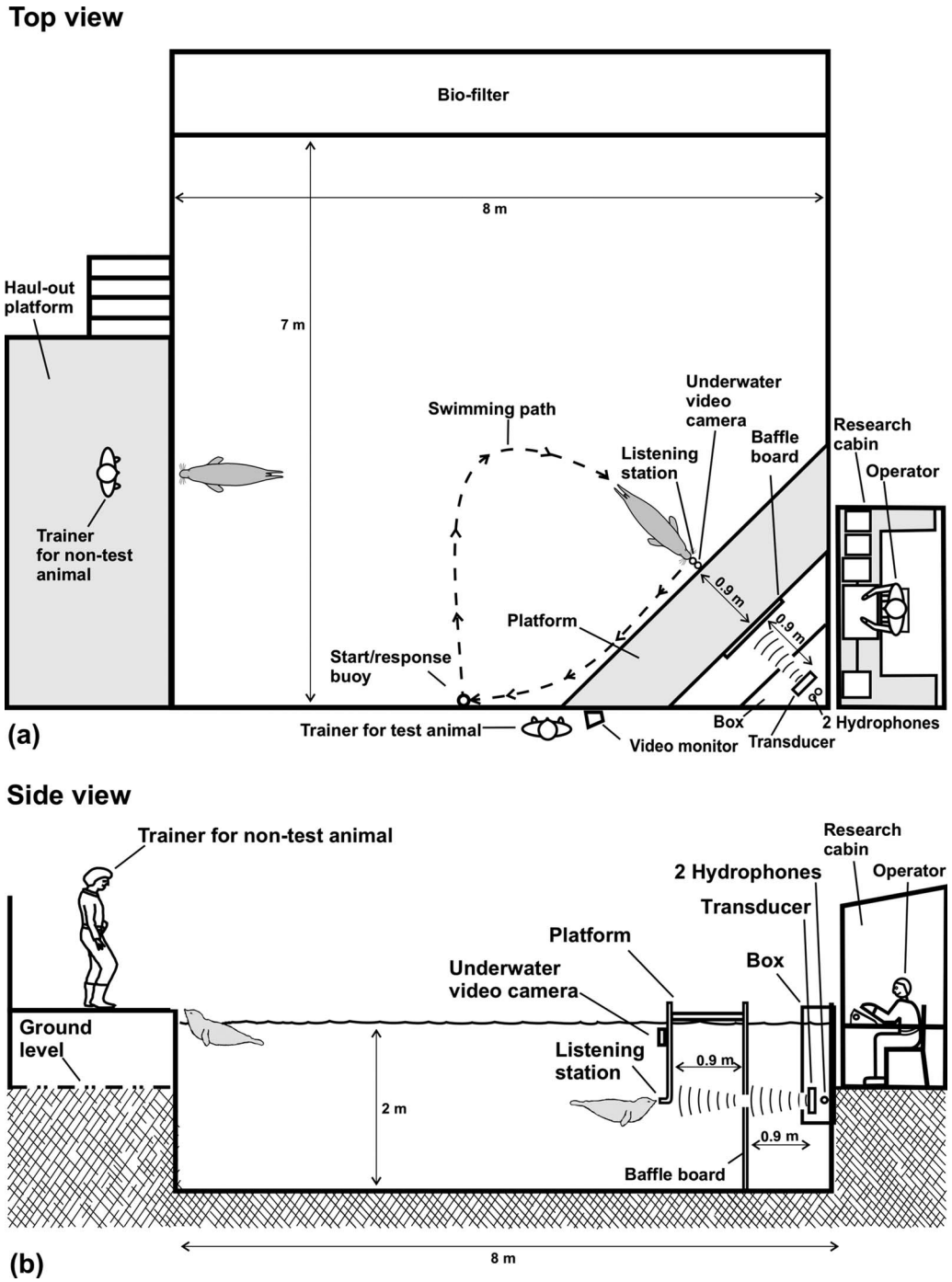


FIG. 1. The study area, showing the test harbor seal in position at the underwater listening station; (a) top view and (b) side view, both to scale.

B. Study area and staff

The study was conducted at SEAMARCO's Research Institute, The Netherlands. The institute is in a remote area which was specifically selected for acoustic research. The measurements were conducted in an outdoor pool (8 m × 7 m, 2 m deep) with an adjacent haul-out platform (Fig. 1). Several measures in the design of the pool were taken to make the pool as quiet as possible and reduce reflections above 25 kHz. More detailed information about the study area can be found in Kastelein *et al.*, 2009.

During test sessions, the seals were tested alternately. The animal not being tested was trained to keep very still and quiet for 15 min in the water next to the haul-out platform (this was quieter than staying on the platform, where a scratch of a flipper nail could trigger a prestimulus response in the animal being tested). The operator and the equipment used to produce the stimuli were in a research cabin next to the pool, out of sight of both animals (Fig. 1).

C. Test stimuli

The equipment used to configure and emit outgoing signals is shown in Fig. 2. The seals' hearing sensitivity was tested for narrow (approximately 1/3-octave) noise bands. White noise was produced by a waveform generator (Hewlett

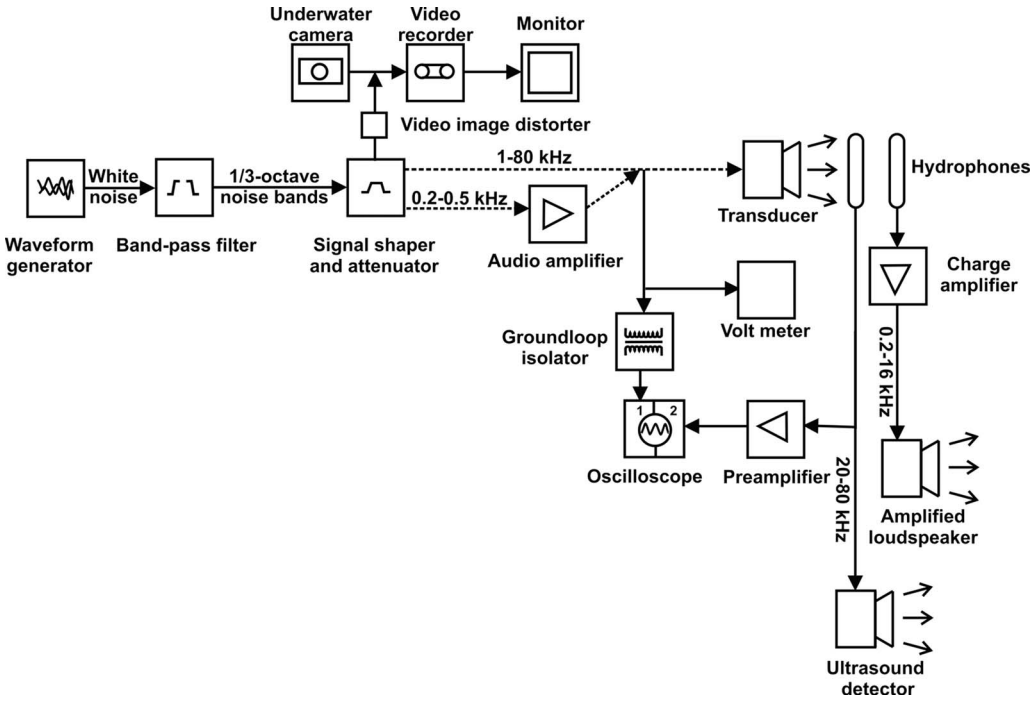


FIG. 2. Block diagram of the transmitting and listening systems.

Packard 33120A) and filtered by a brick-wall band-pass filter (Krohn-Hite 3901; roll-off rate: 115 dB/octave) at 1/3-octave band frequency limits (according to the American National Standards Institute). This resulted in noise bands with most energy in the center frequency band and some energy in both adjacent 1/3-octave bands. Center frequencies of the noise bands were 0.2, 0.25, 0.5, 1, 2, 4, 8, 16, 25, 31.5, 40, 50, 63, and 80 kHz (Table I).

A modified audiometer for testing human aerial hearing (Midimate, model 602) was used as a signal shaper (to control the duration and amplitude of the signals). The stationary portion of all signals was 900 ms in duration. The rise and fall times of the signals were 50 ms, to prevent transients. The signal duration probably exceeded the integration time of the harbor seal's hearing system (Terhune, 1988). The sound level (noise band level) at the seal's head while it was at the listening station could be varied in 5 dB increments. This step size was determined by the audiometer; 5 dB steps are generally used in human audiometry. The 0.2–0.5 kHz signals from the audiometer were amplified by an audio amplifier (Sony TA-F335 R).

A directional transducer (Ocean Engineering Enterprise DRS-12; 30 cm diameter) was used to project the signals into the water. In order to eliminate harmonic distortion, an

TABLE I. The mean calibration SPLs and standard deviations (SDs), at the location of the seals' head, of the 1/3-octave noise bands centered at 14 frequencies. Also shown is the power sum of the three 1/3-octave noise bands used to calculate the 50% detection hearing thresholds shown in Table II and Fig. 3(b).

Center frequency (kHz)	1/3-octave frequency range (kHz)	Mean (n=4) received SPL ± SD (dB re 1 μPa, rms)			Power sum all three 1/3-octave bands (dB re 1 μPa, rms)
		1/3 octave lower band	1/3-octave center band	1/3-octave higher band	
0.2	0.18–0.22	73 ± 3	79 ± 1	74 ± 3	81
0.25	0.22–0.28	80 ± 3	81 ± 4	73 ± 3	84
0.5	0.45–0.56	88 ± 2	98 ± 2	90 ± 5	99
1	0.89–1.1	72 ± 0	81 ± 0	81 ± 0	84
2	1.8–2.2	79 ± 1	87 ± 1	80 ± 1	88
4	3.5–4.5	91 ± 2	100 ± 1	98 ± 1	102
8	7.1–8.9	100 ± 1	104 ± 1	97 ± 1	106
16	14–18	107 ± 1	108 ± 1	101 ± 2	111
25	23–28	106 ± 1	110 ± 1	103 ± 2	112
31.5	28–36	107 ± 2	113 ± 3	108 ± 3	115
40	36–45	111 ± 3	115 ± 3	104 ± 1	117
50	45–56	109 ± 2	115 ± 1	108 ± 3	117
63	56–71	115 ± 1	123 ± 1	121 ± 1	126
80	71–89	124 ± 1	129 ± 1	122 ± 2	131

impedance matching transformer was not used. Multi-path arrivals and standing waves can introduce both temporal and spatial variations in the free field SPL at the listening station. To avoid this, the transducer was placed in a corner of the pool in a protective wooden box lined with rubber with an irregular surface [this was less important for the present study than the previous tonal study (Kastelein *et al.*, 2009), because in the present study noise bands were used]. The transducer was hung by four nylon cords from the cover of the box and made no contact with the box. A stainless steel weight was fixed to the lower part of the transducer to compensate for its buoyancy. The transducer was 1.85 m from the tip of the L-shaped listening station (Fig. 1) and was positioned so that the acoustic axis of the projected sound beam was pointed at the center of the study animal's head while it was at the listening station. To reduce reflections from the bottom of the tank and water surface reaching the listening station, a baffle board was placed exactly halfway between the transducer and the animal. The board consisted of 2.4 m high, 1.2 m wide, 4 cm thick plywood, covered with a 2 cm thick closed cell rubber mat on the side facing the transducer. A 30-cm-diameter hole was made in the board with its center level with the seal's head and the transducer (1 m below the water surface). As an indicator of the condition of the transducer, its capacity was checked once a week with a capacity meter (SkyTronic 600.103). Throughout the study period, the capacity remained constant.

D. Stimuli level calibration and background noise measurement

Great care was taken to make the seal's listening environment as quiet as possible. Nobody was allowed to move within 15 m of the pool during sessions. Underwater background noise levels were measured five times during the 4-month study period, under the same conditions as during the test sessions (in various weather conditions with wind speed below Beaufort 3).

The equipment used to measure the background noise in the pool consisted of a Bruel & Kjær (B&K) 8101 hydrophone, a voltage amplifier system (TNO TPD, 0–300 kHz), and a dual spectrum analyzer system (0.025–160 kHz). The system was calibrated with a pistonphone (B&K 4223) and a white noise signal (0.025–40 kHz) which was inserted into the hydrophone preamplifier. Measurement results were corrected for the frequency sensitivity of the hydrophone and the frequency response of the measurement equipment. The customized analyzer consisted of an A/D-converter (Avisoft UltraSoundGate 116, 0–250 kHz) coupled to a notebook computer (sampling rate of 500 kHz). The digitized recordings were analyzed by two parallel analysis systems: (1) a fast Fourier transform narrowband analyzer (0.025–160 kHz) and (2) a 1/3-octave band analyzer (0.025–160 kHz).

1/3-octave band background noise levels were determined in the range 0.025–100 kHz. The background noise in the pool was very low [below sea state 0 at frequencies below 7 kHz, Fig. 3(a)].

The received sound level (dB re 1 μPa, rms) of each noise band was measured at the seal's head position (while at the listening station). During trials, the seal's head position was carefully monitored and was consistent to within a few centimeters. The sound level received by the seal was measured in three 1/3-octave bands: the 1/3-octave test band and the two adjacent 1/3-octave bands (so the analysis range was 1 octave per noise band). Because the adjacent 1/3-octave bands contributed to the total energy in the test signal, the power sum of the three 1/3-octave bands was used to determine the auditory threshold. This summation resulted in 1–3 dB being added to the sound levels of the central 1/3-octave noise bands (Table I). There was little variation in level between calibration sessions, because noise bands produce little interference with their surroundings (i.e., few standing waves) and because the bandwidth of the noise was wide enough to create a uniform field (Dillon and Walker, 1982).

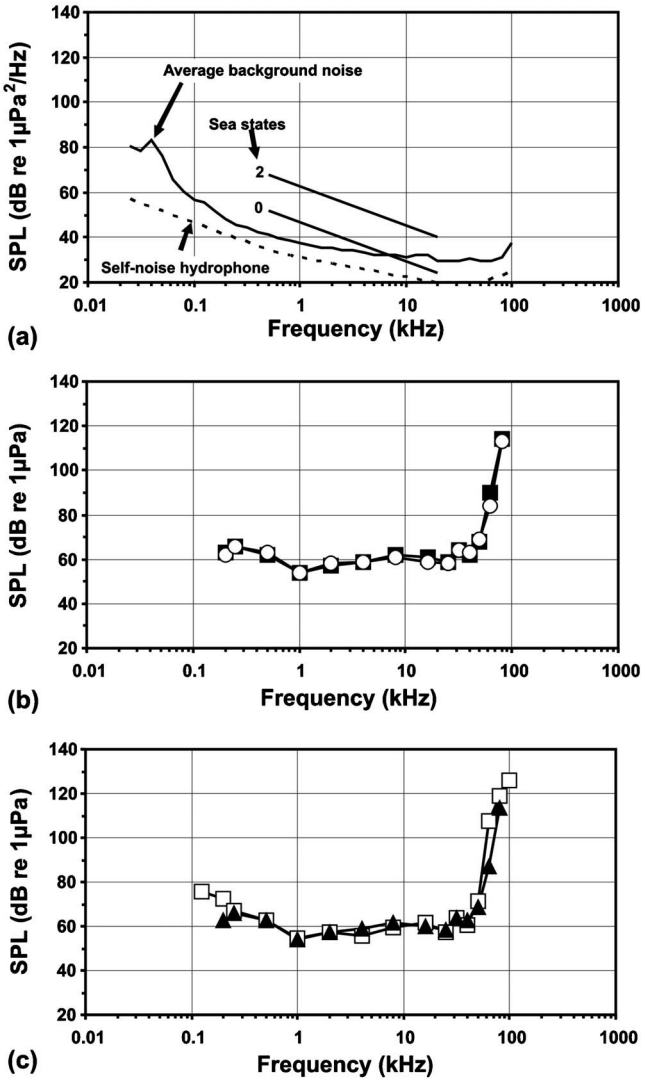


FIG. 3. (a) Power averaged background noise level in the pool in dB re 1 μPa^2/Hz (derived from 1/3-octave band levels; $n=5$) and the self-noise of the B&K 8101 hydrophone (and amplifier). Also shown are the sea noise levels of sea states 0 and 2 (Knudsen *et al.*, 1948). (b) The mean 50% detection thresholds (dB re 1 μPa, rms) for 1/3-octave noise bands obtained for female harbor seals 01 (■) and 02 (○) in the present study (for details see Table II). (c) The mean underwater hearing threshold (dB re 1 μPa, rms) of the two study animals in the present study for 1/3-octave noise bands (▲), and the mean underwater hearing thresholds for tonal signals (□) of the same animals, with the same equipment, in the same environment 4 months earlier (Kastelein *et al.*, 2009).

TABLE II. The mean 50% detection thresholds, SDs, and total number of reversal pairs of 18–21-month-old female harbor seals 01 and 02 for 14 1/3-octave noise band signals, and their prestimulus response rates calculated from signal-present and signal-absent trials.

	Harbor seal 01			Harbor seal 02		
Center frequency (kHz)	Total no. of reversal pairs	Mean 50% detection threshold (SPL in dB re 1 μPa, rms)±SD	Prestimulus response rate (%)	Total no. of reversal pairs	Mean 50% detection threshold (SPL in dB re 1 μPa, rms)±SD	Prestimulus response rate (%)
0.2	135	63±3	5	131	62±4	9
0.25	123	66±3	7	131	66±3	8
0.5	93	62±4	9	92	63±4	9
1	105	54±4	10	100	54±4	7
2	119	57±4	8	114	58±4	6
4	108	59±4	6	120	59±4	5
8	128	62±3	8	112	61±4	6
16	112	61±4	8	113	59±4	7
25	131	59±3	5	111	58±4	6
31.5	108	64±4	8	111	64±4	8
40	120	62±4	7	107	63±3	8
50	118	68±4	7	107	69±4	5
63	116	90±4	7	120	84±4	7
80	103	114±4	8	134	113±3	3

The received sound levels (Table I) were calibrated at a level of 17–54 dB (depending on frequency) above the threshold levels (Table II) found in the present study. The attenuation linearity of the audiometer was checked several times during the study and varied by less than 2 dB.

E. Experimental procedure

The experimental procedure was the same as used in the tonal audiogram study (Kastelein *et al.*, 2009). The seals were trained to respond ("go") in the presence of a signal and to withhold the response ("no-go") in the absence of a signal. A trial began when the animal not being tested was near the platform with one trainer and the animal being tested was positioned with its head at the start/response buoy at the edge of the pool next to the research trainer [Fig. 1(a)]. When the trainer gave the animal a vocal command accompanied by a gesture (pointing downwards), the animal descended to the listening station (an L-shaped, 32-mm-diameter, water-filled polyvinylchloride tube with an end cap), so that its external auditory meatus was 200 cm from the sound source and 100 cm below the water surface [i.e., mid-water; Fig. 1(b)]. Each animal was trained to position its nose against the listening station so that its head axis was in line with the projected beam axis of the transducer. The listening station was not connected to the sound box, and the transducer was suspended within the box by four cords, so the animals were not able to use vibration via contact conduction to the nose to detect the signals. The animals' positions could be viewed from above by means of an underwater camera (Mariscope, Micro), which was attached to the listening station. The images were visible to the trainer near the start/response buoy (who was out of the study animal's view when it was at the listening station) and to the operator in the research cabin.

Two trial types were conducted during each experimental session: signal-present trials and signal-absent trials. In signal-present trials, the stimulus was presented unpredictably between 4 and 10 s after the animal was positioned correctly at the listening station. A minimum waiting time of 4 s was chosen because it took about 4 s for the waves, created by the animal's descent, to dissipate. If the animal detected the sound, it responded by leaving the listening station ("go" response) at any time during the signal's duration and returning to the start/response buoy [Fig. 1(a)]. The signal operator then indicated to the trainer that the response was correct (a hit), after which the trainer gave a vocal signal and the seal received a fish reward. If the animal did not respond to the signal, the signal operator indicated to the trainer that the animal had failed to detect the signal (a miss). The trainer then indicated to the animal (by tapping softly on the side of the pool) that the trial had ended, thus calling the animal back to the start/response buoy. No reward was given following a miss. If the animal moved away from the listening station to the start/response buoy before a signal was produced (a prestimulus response), the signal operator indicated to the trainer to end the trial without rewarding the animal. After a prestimulus response, the animal was ignored for 8–10 s by the trainer.

In signal-absent, or catch, trials the signal operator handsignaled to the trainer to end the trial after a random interval of 4–10 s from when the seal had stationed (determined by a random number generator). The trial was terminated when the trainer blew very softly on a whistle. The tapping on the pool wall and whistle blowing were done softly to reduce the difference in the seal's exposure level (between the test signals and the acoustic signals from the trainer). We believe this helped the animals to focus on very faint sounds throughout the sessions. If the animal responded correctly by remaining at the listening station until the whistle was blown (a correct rejection), it then returned to the start/response buoy and received a fish reward. If the seal left the listening station before the whistle was blown (a prestimulus response), the signal operator indicated to the trainer to end the trial without rewarding the animal. The same amount of fish was given as a reward for correct "go" and "no-go" re-

sponses. In both signal-present and signal-absent trials, the trainer was unaware of the trial type when she sent the animal to the listening station. After sending the animal to the listening station, the trainer stepped out of the seal's view.

A session generally consisted of 30 trials per animal and lasted for about 15 min per animal. The seals were always tested in the same order, one immediately after the other. Sessions consisted of 70% signal-present and 30% signal-absent trials presented in random order, and only one signal frequency was presented each day. For each session, one of four data collection sheets was used. Each sheet comprised a different random series of trial types. Each seal had its own set of four data collection sheets. In each session, the signal amplitude was varied according to the simple up-down staircase procedure, a conventional psychometric technique (Robinson and Watson, 1973). This is a variant of the method of limits, which results in a 50% correct detection threshold (Levitt, 1971). During preliminary sessions a rough threshold per test frequency was determined. During subsequent experimental sessions, the starting SPL of the signal was 10–15 dB above the estimated threshold. Following each hit, the signal amplitude of the next signal-present trial was reduced by 5 dB. Following each miss, the signal level was increased of the next signal-present trial by 5 dB. Prestimulus responses did not lead to a change in signal amplitude for the next trial. A switch in the seal's response from a detected signal (a hit) to an undetected signal (a miss), or vice versa, is called a reversal.

Thresholds were determined for 14 noise bands. To prevent the animals' learning process from affecting the threshold levels, the center frequency was varied each day; adjacent frequencies were tested on successive days (going from low to high and from high to low frequencies, and so forth). This way the difference in frequency between days was limited, reducing the animals' potential need to adapt to a new frequency. During the course of the study (and during the previous hearing study in this pool), it became apparent that the thresholds obtained for higher frequencies (≥40 kHz) were not influenced by the wind force. Therefore, those frequencies were tested under relatively high wind force conditions. The 0.2–0.5 kHz signals were only tested under wind force conditions below Beaufort 2, because a quieter environment was needed to achieve consistent thresholds for these signals. Four measurement sessions were conducted daily, five days per week (at 0900, 1100, 1400, and 1600 h). Data were collected between December 2007 and March 2008.

Before each session, the acoustic equipment producing the stimuli was checked to ensure that it was functioning properly and the stimuli were being produced accurately (Fig. 2). Also the background noise level was checked to make sure it was low enough for testing. This was done as described by Kastelein *et al.* (2009).

F. Analysis

The seals had participated in a similar hearing study just before the present study (Kastelein *et al.*, 2009), and therefore the mean session thresholds were stable from the beginning. No warm-up trials were used but it sometimes took several reversals before a stable threshold was reached within a session. In this case the first 1-4 reversals were not included in the analysis. Prestimulus response rates are calculated from signal-present and signal-absent trials (incorrect detections on stimulus-absent trials). Sessions with more than 20% prestimulus responses (i.e., 6 or more out of the usual 30 trials per session) would have been omitted from the analysis. However, such sessions did not occur during the entire study period.

The 50% correct detection thresholds for each animal were calculated by taking the mean of ~115 reversal pairs per frequency, obtained in ten sessions (Table II). Per frequency, the hearing thresholds of the two seals for each 1/3-octave noise band were compared by using a paired t-test, in which the sample size was the number of frequencies (14). The mean 1/3-octave hearing thresholds of the seals were compared to the mean tonal thresholds of the same animals by using a separate paired t-test.

III. RESULTS

The seals' sensitivity for each test band was stable over the 4-month study period. The mean prestimulus response rate (for both signal-present and signal-absent trials) varied between 3% and 10%, depending on the animal and frequency (Table II). The 50% detection threshold plots of both seals were U-shaped as is typical for mammals [Fig. 3(b) and Table II]. The bottom part of the U was rather flat and wide, and the low-frequency sensitivity decreased gradually, while the high-frequency cutoff was steep. The range of best hearing was from 0.5 to 40 kHz (threshold level <10 dB from the maximum sensitivity: 54 dB re 1 μPa, rms), where thresholds were approximately 60 dB re 1 μPa (SD±3.0 dB). The threshold levels increased to 114 dB re 1 μPa at 80 kHz. The 1/3-octave noise band thresholds of the two seals were similar for each frequency ($t=1.17$, $d.f.=13$, $P=0.26$).

Except for 0.2 kHz, the mean values of the thresholds of the two seals per frequency for 1/3-octave noise bands were similar ($t=1.43$, $d.f.=13$, $P=0.18$) to those for narrowband FM tonal signals [Kastelein *et al.*, 2009; Fig. 3(c)].

IV. DISCUSSION AND CONCLUSIONS

The hearing thresholds shown in the present study were based on the power sum of the 1/3-octave test band and the two adjacent 1/3-octave bands. Whether this threshold calculation is correct is debatable, but the summation resulted only in an additional 1–3 dB being added to the sound levels of the central 1/3-octave noise bands (Table I).

Kastelein *et al.* (2009) presented FM tonals (the modulation range of the signals was ±1% of the center frequency) rather than constant frequency tonals to reduce the possibility of standing waves developing, and thus, as with the current study, presented the subjects with stable sound levels in the test area. Except for one frequency, the mean thresholds for 1/3-octave noise bands were similar to those obtained for narrowband FM tonal signals from the same seals four months earlier [Kastelein *et al.*, 2009; Fig. 3(c)]. Only at 0.2 kHz, the noise band thresholds were 9 dB lower than the

tonal thresholds. The lower threshold of the noise band (0.18–0.22 kHz) could be the result of the energy in this band above 0.2 kHz, as in this low-frequency part of the hearing curve the seals' hearing becomes better as frequency increases. Another possible explanation is that this noise band has energy below 0.2 kHz, and perhaps particle motion influenced the threshold.

Results from human audiometry support the validity of using 1/3-octave noise bands as a measure of hearing sensitivity. For human adults with normal hearing, narrowband noise thresholds (1/3-octave) are very similar to pure-tone thresholds (Simon and Northern, 1966; Berger, 1981; Cox and McDaniel, 1986).

The hearing thresholds obtained in the present study were probably not masked. Fig. 3(a) shows the power averaged background noise in the seal pool (as Leq spectrum level), measured during 4 calibration sessions conducted equally spread over the study period. The conditions during these measurements varied between absolutely no wind and a wind force of Beaufort 3. Theoretical masked detection thresholds were calculated based on the background noise levels [Fig. 3(a)], and the harbor seal's critical ratios (from a smooth line through the data points of Turnbull and Terhune, 1990 and Southall *et al.*, 2000). For pure tones the noise-limited theoretical masked detection threshold is calculated as background noise (spectrum level) plus the critical ratio. The theoretical masked detection thresholds for tones lie just below the hearing thresholds found in the present study.

Sound detection thresholds measured in very quiet surroundings are limited by the sensitivity of the listener. At sea state 0, in an area with little nearby vessel traffic, a broadband noise source from a ship or pile driver may be audible at great distances. Sea state is influenced mainly by waves and wind at medium to high frequencies, and by vessel traffic at lower frequencies (see Wenz curves in Urick, 1983). In nature, except on particularly quiet days with no nearby shipping or other anthropogenic noise sources, the hearing thresholds for the 1/3-octave noise bands found in the present study would be masked by ambient noise.

Coupling source levels of noise bands with sound transmission losses in the environment and harbor seal detection threshold levels would allow modeling of the detection ranges of various noises. This is an important step in understanding at what distance harbor seals can detect, and possibly respond to, ecologically relevant sounds, such as sounds from conspecifics, and anthropogenic sounds.

ACKNOWLEDGMENTS

We thank students Krista Krijger, Tess van der Drift, Janna Loot, and Alejandra Vargas, volunteers Menno van den Berg, Jesse Dijkhuizen, Cathy Philipse, Saskia Roose, Joke de Lange, and Petra van der Marel for their help in collecting the data, and Rob Triesscheijn for making the figures. We thank Veenhuis Medical Audio (Marco Veenhuis and Herman Walstra) for donating and modifying the audiometer. We thank Bert Meijering (Director of Sea Bait farm Topsy Baits, Wilhelminadorp, The Netherlands) for providing space for SEAMARCO's Institute and Hein Hermans for providing technical support to run the facility. We thank Willem Verboom (JunoBioacoustics) for calibrating the equipment and for his comments on this manuscript. We also thank Nancy Jennings (Dotmoth.co.uk, Bristol, UK), Charles Greene (Greenridge Science, Missouri, MO), Christ de Jong (TNO Science and Industry, Delft, The Netherlands), Michael Ainslie (TNO Defence, Security and Safety, The Hague, The Netherlands), Jim Finneran (Space and Naval Warfare Systems Center, San Diego, CA), Dorian Houser (Biomimetica, San Diego, CA), and two anonymous reviewers for their valuable constructive comments on this manuscript. This study was conducted by SEAMARCO as a subcontractor of IMARES (contacts: Han Lindeboom and Reinier Hille Ris Lambers). It was funded by We@Sea and Noordzee Wind EIA (wind turbine parks at sea), and RIKZ Middelburg, The Netherlands (contacts: Belinda Kater and Martine van den Heuvel-Greve; acoustic disturbance of harbor seals in the Westerscheldt). We thank director Just van den Broek and curator of animals Henk Brugge (both from Ecomare) for making the harbor seals available for this project. The seals' training and testing were conducted under authorization of the Netherlands Ministry of Agriculture, Nature and Food Quality, Department of Nature Management, with endangered Species Permit No. FF/75A/2005/048.

Berger, E. H. (**1981**). "Re-examination of the low-frequency (50–1000 Hz) normal threshold of hearing in free and diffuse sound fields," J. Acoust. Soc. Am. **70**, 1635–1645.

Burns, J. J. (**2002**). "Harbor seal and spotted seal," in *Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Würsig, and J. G. M. Thewissen (Academic, San Diego), pp. 552–560.

Cox, R. M., and McDaniel, M. (**1986**). "Reference equivalent threshold levels for pure tones and 1/3-octave noise bands: Insert earphone and TDH-49 earphone," J. Acoust. Soc. Am. **79**, 443–446.

Dillon, H., and Walker, G. (**1982**). "Comparison of stimuli used in sound field audiometric testing," J. Acoust. Soc. Am. **71**, 161–172.

Hanggi, E. B., and Schusterman, R. (**1994**). "Underwater acoustic displays and individual variation in male harbor seals, *Phoca vitulina*," Anim. Behav. **48**, 1275–1283.

Hayes, S. A., Costa, D. P., Harvey, J. T., and Le Boeuf, B. J. (**2004**). "Aquatic mating strategies of the male Pacific harbor seal (*Phoca vitulina richardii*): Are males defending the hotspot?," Marine Mammal Sci. **20**, 639–656.

Kastak, D., and Schusterman, R. J. (**1998**). "Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology," J. Acoust. Soc. Am. **103**, 2216–2228.

Kastelein, R. A., Wensveen, P. J., Hoek, L., Verboom, W. C., and Terhune, J. M. (**2009**). "Underwater detection of tonal signals between 0.125 and 100 kHz by harbor seals (*Phoca vitulina*)," J. Acoust. Soc. Am. **125**, 1222–1229.

Knudsen, V. O., Alford, R. S., and Emling, J. W. (**1948**). "Underwater ambient noise," J. Mar. Res. **3**, 410–429.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Møhl, B. (**1968**). "Auditory sensitivity of the common seal in air and water," J. Aud. Res. **8**, 27–38.

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Robinson, D. E., and Watson, C. S. (**1973**). "Psychophysical methods in modern psychoacoustics," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **2**, pp. 99–131.

Simon, G. R., and Northern, J. L. (**1966**). "Automatic noiseband audiometry," J. Aud. Res. **6**, 403–407.

Southall, B. L., Schusterman, R. J., and Kastak, D. (**2000**). "Masking in three pinnipeds: Underwater, low-frequency critical ratios," J. Acoust. Soc. Am. **108**, 1322–1326.

Southall, B. L., Schusterman, R. J., Kastak, D., and Reichmuth-Kastak, C.

(**2005**). "Reliability of underwater hearing thresholds in pinnipeds," ARLO **6**, 243–249.
Terhune, J. M. (**1988**). "Detection thresholds of a harbor seal to repeated underwater high-frequency, short duration sinusoidal pulses," Can. J. Zool. **66**, 1578–1582.
Turnbull, S. D., and Terhune, J. M. (**1990**). "White noise and pure tone masking of pure tone thresholds of a harbour seal listening in air and underwater," Can. J. Zool. **68**, 2090–2097.
Turnbull, S. D., and Terhune, J. M. (**1993**). "Repetition enhances hearing detection thresholds in a harbour seal (*Phoca vitulina*)," Can. J. Zool. **71**, 926–932.
Turnbull, S. D., and Terhune, J. M. (**1994**). "Descending frequency swept tones have lower thresholds than ascending frequency swept tones for a harbor seal and human listeners," J. Acoust. Soc. Am. **96**, 2631–2636.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).
Van Parijs, S. M., and Kovacs, K. M. (**2002**). "In-air and underwater vocalizations of the eastern Canadian harbour seals, *Phoca vitulina*," Can. J. Zool. **80**, 1173–1179.

Auditory evoked potentials in a stranded Gervais' beaked whale (*Mesoplodon europaeus*)

James J. Finneran[a)]
US Navy Marine Mammal Program, SSC Pacific, Biosciences Division, Code 71510, 53560 Hull Street, San Diego, California 92152

Dorian S. Houser
BIOMIMETICA, 7951 Shantung Drive, La Mesa, California 92071

Blair Mase-Guthrie and Ruth Y. Ewing
National Marine Fisheries Service, 75 Virginia Beach Drive, Miami, Florida 33149

Robert G. Lingenfelser
The Marine Mammal Conservancy, P.O. Box 1625, Key Largo, Florida 33037

(Received 7 November 2008; revised 17 April 2009; accepted 21 April 2009)

Efforts to identify the specific causal mechanisms responsible for beaked whale strandings coincident with naval exercises have been hampered by lack of data concerning the hearing abilities of beaked whales and their physiological and behavioral responses to sound. In this study, auditory capabilities of a stranded Gervais' beaked whale (*Mesoplodon europaeus*) were investigated by measuring auditory evoked potentials. Click-evoked potentials, auditory thresholds as a function of frequency, and the modulation rate transfer function were determined. The evoked potentials and modulation rate transfer function were similar to those measured in other echolocating odontocetes; the upper limit of functional hearing was 80–90 kHz.
 [DOI: 10.1121/1.3133241]

PACS number(s): 43.80.Lb, 43.80.Nd [WA] Pages: 484–490

I. INTRODUCTION

Mass strandings of beaked whales temporally and spatially coincident with naval sonar exercises have raised numerous questions about the potential adverse effects of underwater sound on these animals (Houser *et al.*, 2001; U.S. Department of Commerce and U.S. Department of the Navy, 2001; Jepson *et al.*, 2003; Fernández *et al.*, 2005). At present, there is much speculation regarding the specific causal mechanisms responsible for these strandings (e.g., Jepson *et al.*, 2003; Fernández *et al.*, 2005), but few actual data, primarily because of the difficulties historically associated with studying deep-diving animals that are generally not kept under human care. Recent investigations have begun to overcome some of these difficulties through the use of recording tags that archive movements, dive patterns, and acoustic signals of wild marine mammals (e.g., Johnson and Tyack, 2003). Application of these tags has revealed unique dive behaviors in beaked whales (Tyack *et al.*, 2006) and shed light on certain features of the echolocation system (e.g., Madsen *et al.*, 2005). For example, echolocation clicks recorded from Cuvier's beaked whales (*Ziphius cavirostris*) consisted of relatively long duration (~200 μs), frequency-modulated (FM) pulses, with frequency content (−10 dB bandwidth) from 31 to 53 kHz (Zimmer *et al.*, 2005). Recordings from Blainville's beaked whales (*Mesoplodon densirostris*) revealed FM clicks with durations of ~250 μs duration and −10 dB bandwidth from 26 to 51 kHz, as well as shorter duration (~100 μs) clicks with −10 dB bandwidth from 25 to at least 80 kHz (Johnson *et al.*, 2006).

Despite increasing progress in understanding the characteristics of the biosonar transmitting system in beaked whales, direct data regarding the biosonar receiving system, i.e., the auditory system, are scarce. In the only previous study of beaked whale hearing, Cook *et al.* (2006) measured auditory evoked potentials (AEPs) in a stranded juvenile Gervais' beaked whale (*Mesoplodon europaeus*). AEPs are small voltages formed by synchronous neural discharges that are time-locked to a sound stimulus. Previous studies have demonstrated that AEPs may be safely measured in odontocete cetaceans using noninvasive surface electrodes (e.g., Popov and Supin, 1985; Popov and Supin, 1990b), that no specific training is required (e.g., Nachtigall *et al.*, 2005), and that AEP thresholds are correlated with behavioral measurements of hearing (Szymanski *et al.*, 1999; Yuen *et al.*, 2005; Finneran and Houser, 2006; Houser and Finneran, 2006a; Finneran *et al.*, 2007c; Schlundt *et al.*, 2007). Using sinusoidally amplitude-modulated (SAM) tone stimuli, Cook *et al.* (2006) measured AEP amplitudes in *M. europaeus* as a function of the stimulus modulation rate. The resulting modulation rate transfer function (MRTF) was similar to those measured in dolphins and other odontocetes (Dolphin *et al.*, 1995; Supin and Popov, 1995; Klishin *et al.*, 2000; Mooney *et al.*, 2006; Finneran *et al.*, 2007b), with peaks in the response near 600 Hz and 1 kHz. Cook *et al.* (2006) also measured hearing thresholds; however, the measurement system was limited to frequencies of 80 kHz and below; there-

a)Author to whom correspondence should be addressed. Electronic mail: james.finneran@navy.mil

 0001-4966/2009/126(1)/484/7/$25.00

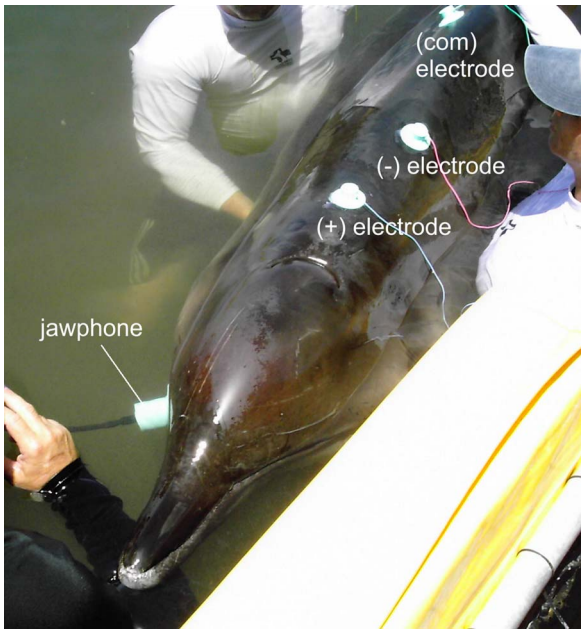


FIG. 1. (Color online) Photograph of the beaked whale subject showing the positions of the three surface electrodes and jawphone.

fore, the upper cutoff frequency of hearing—the highest frequency at which the whale could effectively hear—was not identified.

This report describes the results of AEP measurements conducted on a stranded adult, female Gervais' beaked whale. There were two primary differences between this study and the previous work of Cook *et al.* (2006): (1) The whale was an adult, rather than a juvenile, and (2) the measurement system possessed bandwidth to 200 kHz, allowing the upper limit of hearing to be defined. Three specific tests were conducted: (1) recording transient, click-evoked potentials, (2) measuring hearing thresholds as a function of stimulus frequency, and (3) measuring AEP amplitude as a function of stimulus modulation rate (to derive the MRTF). The primary goal of the measurements was to define the upper cutoff frequency of hearing.

II. METHODS

A. Subject

The subject was an adult female Gervais' beaked whale, 4.6 m (15 ft) in length with an estimated weight of 725 kg (1600 lb). The whale was found on the morning of 20 June 2008, near Islamorada, FL and brought to the Marine Mammal Conservancy (MMC), in Key Largo, FL, for rehabilitation. Species identification was verified through a genetic analysis conducted at the Southeast Fisheries Science Center Marine Mammal Molecular Genetics Laboratory.

Hearing tests were conducted in a shallow (~1.5 m depth) lagoon at the MMC on 21 June 2008. The subject was held near the water surface by MMC staff and volunteers, with the top of the head and blowhole above the water (Fig. 1). The exposed surface of the whale was kept moist using water from the lagoon.

The whale was euthanized on 23 June 2008 after tests revealed multiple organ failure and fulminant hepatic disease. The necropsy bacteriology results revealed coagulase negative *Staphylococcus* in multiple organs (stomach, uterine horn, cerebral meninges, kidney, and spleen). A diffuse meningeal opacity was noted from which coagulase negative *Staphylococcus* and *Staphylococcus aureus* were cultured. The *Staphylococcus* sp isolates obtained at necropsy were likely the result of postmortem overgrowth subsequent to a terminal bacterial sepsis. Incidentally, the whale presented with a mild subdural hematoma immediately dorsal to the cerebellum on the right occipital lobe. Microscopic examinations of central nervous system (CNS) tissues consisted of tissue fixation in 10% neutral buffered formalin followed by routine histological processing. The tissues were embedded in paraffin, sectioned at 5 μm, and stained with hematoxylin and eosin. Additional special histochemical stains were also conducted including Prussian Blue, Masson's Trichrome, and Lipofuscin. On microscopic examination, the leptomeninges covering the cerebellum were markedly thickened by diffuse collagen bundles interspersed with edema and scattered golden brown pigment laden macrophages that had variable ceroid-lipofuscin staining. Sections from other regions of the cerebrum had varying amounts of perivascular darker brown pigment deposits, which also had variable reduced positive staining for ceroid-lipofuscin. The eighth cranial nerve had prominent perivascular macrophages with abundant refractile golden pigment deposits, which stained intensely positive for ceroid-lipofuscin. The Prussian Blue stain for iron was negative on all CNS sections stained.

B. AEP measurements

Stimulus generation and evoked response recording were performed using the evoked response study tool (EVREST) (Finneran, 2008, 2009). This system is centered on a rugged notebook computer with a PCI expansion chassis containing a multifunction data acquisition board (National Instruments NI PCI-6251). Stimuli were digitally generated, converted to analog with a 1 MHz update rate and 16-bit resolution, low-pass filtered at 200 kHz (eight-pole Butterworth, Krohn-Hite 3C series), and attenuated (custom, 0–70 dB range) before being applied to a "jawphone"—a piezoelectric sound projector (ITC 1042) embedded in a silicon rubber suction cup. The jawphone was placed on the whale's lower right jaw, in the region estimated to be the site of high-frequency sound reception (Fig. 1). A jawphone was used, rather than a non-contact sound projector, because the whale's head movements were at times large and would have made accurate calibration of the received sound levels difficult. The jawphone was previously calibrated by measuring the underwater sound pressure produced at a distance of 15 cm, derived from anatomical measurements in dolphins. This calibration technique has produced reasonable agreement, in dolphins, between AEP thresholds measured in-air with jawphones and those behaviorally measured underwater (Finneran and Houser, 2006). For beaked whales, this jawphone calibration approach has not been directly validated; however, to allow comparison with the beaked whale thresholds obtained by Cook *et al.* (2006) and dolphin thresholds ob-

tained using jawphones calibrated under similar conditions, received sound pressure levels (SPLs) at threshold were estimated from the jawphone calibrations validated in dolphins.

Stimuli consisted of clicks and SAM tones. Clicks were (positive) rectangular pulses with a duration of 50 μs, designed to produce a broadband stimulus that would excite a large population of neurons and thus produce a relatively large AEP. SAM tones evoke the auditory steady-state response (ASSR), also known as the envelope following response. The ASSR is produced when stimuli are presented sufficiently fast that transient evoked responses overlap to form a steady-state signal. The ASSR is a periodic signal with a fundamental frequency related to the SAM tone modulation frequency and may be analyzed in the frequency domain using established techniques for objective, statistically-based response detection (e.g., Stapells *et al.*, 1987; Dobie and Wilson, 1989; Dobie and Wilson, 1996; Finneran *et al.*, 2007a). Because SAM tones may possess relatively narrow frequency bandwidth, ASSR measurement has become a common approach for frequency specific audiometry in marine mammals (e.g., Dolphin, 1995; Supin and Popov, 1995; Supin *et al.*, 2001; Nachtigall *et al.*, 2005; Yuen *et al.*, 2005; Cook *et al.*, 2006; Houser and Finneran, 2006b; Houser *et al.*, 2008). The SAM tones in this study were 22 ms in duration, with 1-ms cosine rise/fall envelopes. For threshold testing carrier frequencies varied from 20 to 160 kHz and the modulation rate was fixed at 1 kHz (based on data from Cook *et al.*, 2006). The MRTF measurements used a 40-kHz carrier and modulation rates from 0.4 to 3 kHz. SPLs for the MRTF measurements were 25 dB above the threshold measured at 40 kHz.

AEPs were measured using three 10-mm gold cup surface electrodes embedded in silicon suction cups and placed on the head and back (Fig. 1). The noninverting (+) electrode was located on the dorsal midline approximately 10 cm posterior to the blowhole. The inverting (−) electrode was located approximately 12 cm posterior to the (+) electrode. A ground (com) electrode was placed approximately 12 cm anterior of the dorsal fin. Electrodes were coupled to the skin surface using conductive paste. The electrode signals passed into a biopotential amplifier (Grass ICP-511) that amplified ($\times 10^5$) and filtered (0.3–3 kHz) the voltage between the (+) and (−) electrodes. The biopotential amplifier output was digitized at a rate of 10 kHz with 16-bit resolution, then synchronously averaged (in software) over 30 ms sweeps synchronized with the stimulus onset. Sweeps with peak instantaneous voltage above 12–18 μV were excluded from analysis (the artifact rejection threshold was initially 12 μV but was increased during the testing period as more artifacts occurred). The number of sweeps rejected in this fashion depended on the whale's movements and normally ranged near 10% of the total sweeps, but was at times as large as 50%.

Click-evoked potentials were averaged over 1024 sweeps. For threshold measurements with SAM stimuli, the presence or absence of an evoked response was determined after integral multiples of 256 sweeps were collected. If a response was detected, the measurement was complete; if not, additional 256 sweeps were collected and the process repeated (using all the available sweeps). The maximum number of sweeps collected was 1024. At each integral multiple of 256 sweeps, coherent averaging in the frequency domain was used to obtain 16 unique "subaverages," each created from an equal number of consecutive sweeps, and a single "grand average" created from all of the sweeps. Magnitude-squared coherence (MSC), which is the ratio of the power in the grand average to the average power of the subaverages (a ratio of signal power to signal-plus-noise power), was then calculated (Dobie and Wilson, 1989; Dobie and Wilson, 1996; Finneran *et al.*, 2007a). If the MSC was greater than the critical value (α=0.01, Brillinger, 1978), the response at a particular modulation frequency was considered to be detected. The MRTF measurements utilized the MSC calculation as described, but always used 1024 sweeps (to allow the amplitudes to be properly compared).

Since the amount of time available for testing was limited, threshold testing focused on delivering near-threshold stimuli to identify the transition between stimuli producing detectable responses and those for whom responses could not be detected. As a result, the number of suprathreshold stimuli was limited and useful input-output curves were not obtained. Threshold testing began at SPLs estimated to be approximately 20 dB above the lowest threshold reported by Cook *et al.* (2006). After each measurement, the SPL was adjusted using a modified up/down staircase approach. If a response was detected, the SPL was reduced by the step size ΔL; if a response was not detected, the SPL was increased by ΔL. The initial step size was 10 dB. After each reversal (a transition from detection to nondetection or vice versa), the step size was reduced according to the rules

$$\Delta L_{k+1} = 0.4\Delta L_k \text{ (for reversals following detections)},$$
$$\Delta L_{k+1} = 0.45\Delta L_k \text{ (for reversals following nondetections)}, \quad (1)$$

where ΔL_k is the step size for the kth measurement. The staircase was terminated when the step size for the next measurement was $<$3 dB. The threshold was defined as the mean of the SPLs corresponding to the lowest detection and the next highest nondetection. The staircase approach with changing step size was designed to rapidly approach threshold while lessening the chance of repeat testing at identical stimulus levels.

For the MRTF measurements, ASSR amplitudes and phase angles were corrected for the bioamplifier frequency response; phase measurements were also corrected for the time delay between stimulus onset and the start of the analysis window. Since sounds were delivered via a jawphone attached to the subject, no phase correction was applied for the sound propagation delay. Measured ASSR phase angles were unwrapped by adding $\pm 2\pi$ rad if the phase difference between adjacent values exceeded $\pm\pi$ rad. Linear regression was performed on the phase vs modulation frequency data and the slope of the regression line, $\Delta\theta/\Delta f_m$, was used to calculate the group delay T_d as follows:

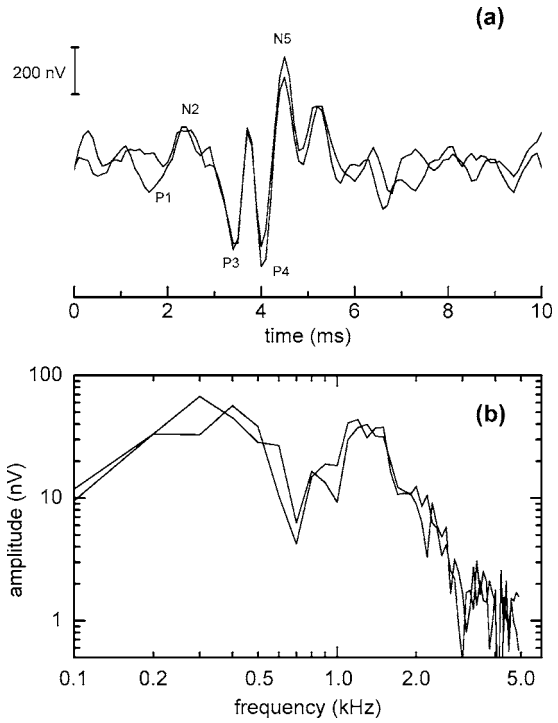


FIG. 2. (a) Time waveforms and (b) frequency spectra for two click-evoked potentials. For comparison to previously published data, waveforms are displayed with upward deflection indicating negativity of the (+) electrode relative to the (−) electrode.

$$T_d = \frac{\Delta\theta/\Delta f_m}{2\pi}, \tag{2}$$

where $\Delta\theta/\Delta f_m$ has units of rad/Hz and T_d is expressed in s (see Dolphin and Mountain, 1992; Supin and Popov, 1995). The regression was only performed at modulation frequencies of 1 kHz and below; above this frequency, the frequency spacing was too large to prevent whole-cycle phase ambiguity.

III. RESULTS

Figure 2 shows two overlaid click-evoked potentials recorded from the beaked whale. The waveforms were qualitatively similar to those measured from dolphins and other odontocetes (Supin *et al.*, 2001), though there are differences between the specific structure of the peaks and the corresponding latencies. To ease comparison to previously published data, waveform peaks were labeled analogously to those presented by Popov and Supin (1985, 1990b). Latencies of the various peaks (estimated from the overlaid waveforms in Fig. 2) were 1.8 (P1), 2.4 (N2), 3.4 (P3), 4.0 (P4), and 4.5 ms (N5). Waveform amplitudes were relatively small (<1 μV, P4-N5) compared to those evoked from dolphins stimulated with the same jawphone at comparable excitation voltages (up to 5–10 μV), and the signal-to-noise ratios were relatively low. Click-evoked potential frequency spectra were also similar to those seen in other odontocetes, with frequency content extending to ~2–3 kHz, peaks near 0.5 and 1 kHz, and a notch near 0.8 kHz.

Auditory thresholds (Fig. 3) were relatively consistent from 20 to 80 kHz, with best sensitivity at 40 kHz. Thresholds rose sharply above 80 kHz and no AEPs were detected at 100, 113, or 160 kHz at the highest SPLs that could be generated. The whale's upper cutoff frequency was therefore 80–90 kHz. Thresholds from the present study at 40 and 56 kHz were close to those reported by Cook *et al.* (2006), while thresholds at 20 and 80 kHz were lower and higher, respectively; however, the threshold differences are within the range of variation observed in large scale AEP testing in dolphins (Houser and Finneran, 2006b; Popov *et al.*, 2007; Houser *et al.*, 2008). The beaked whale thresholds were ~20–30 dB higher than the mean AEP thresholds from ten bottlenose dolphins tested under similar conditions and identified as having "normal" hearing, defined as having an upper cutoff frequency at or above 140 kHz (Houser and Finneran, 2006b). The differences between the dolphin and beaked whale thresholds may reflect an over-estimate of the effective stimulus levels—stimulus levels were based on a calibration distance relevant to dolphins and the larger transmission path in the beaked whale likely reduced the effective stimulus level.

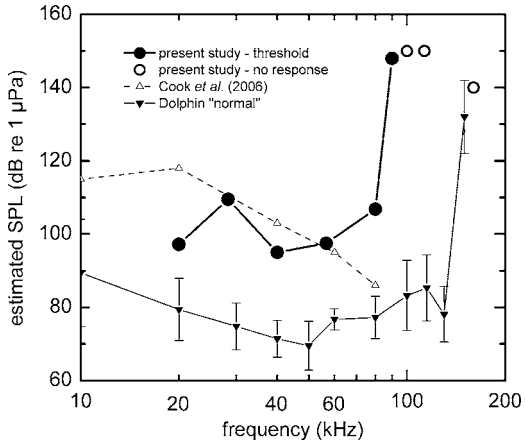


FIG. 3. Hearing thresholds as a function of sound frequency. Circles–*M. europaeus*, present study. Filled circles indicate estimated thresholds, defined as the mean of the SPLs corresponding to the lowest detected AEP and the next highest nondetection. Open circles indicate the highest SPLs tested with no detected AEPs (MSC with 16 subaverages and α=0.01,1024 sweeps). Open triangles–*M. europaeus*, Cook *et al.* (2006). Filled triangles–*T. truncatus*, Houser and Finneran (2006b). Symbols indicate mean thresholds for ten individuals with upper cutoff frequency ≥140 kHz. Error bars represent ± 1 SD.

The MRTF amplitude and phase are presented in Fig. 4. The MRTF amplitude was similar to that measured in other odontocetes, with a general low-pass shape, upper cutoff around 2–3 kHz, and peaks near 0.5 and 1.0 kHz. The MRTF amplitude data closely resembled those presented by Cook *et al.* (2006), who reported peaks near 0.6 (0.5 kHz was not tested) and 1.0 kHz for a juvenile Gervais' beaked whale. The MRTF phase angle varied linearly with modulation frequency from 0.4 to 1.0 kHz, with a slope of −0.027 rad/Hz. The resulting group delay was 4.3 ms, similar to those previously reported for dolphins and belugas (dolphins: ~4 ms, Supin and Popov, 1995; dolphins: ~3.5 ms, Finneran *et al.*, 2007b; dolphin and beluga:~5 ms from 80–350 Hz, Dolphin *et al.*, 1995). The group delay of 4.3 ms suggests that the responses to modulation rates from 0.4 to 1.0 kHz originated in the auditory brainstem.

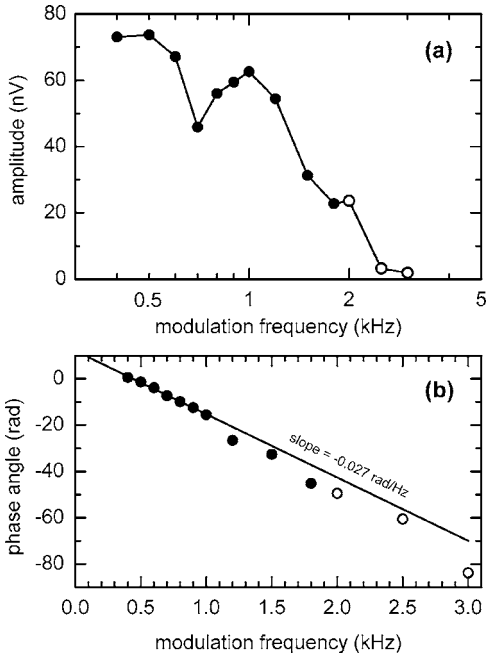


FIG. 4. MRTF (a) amplitude and (b) phase angle measured for the beaked whale. Filled symbols indicate detected responses, and open symbols indicate no detection (MSC, with 16 subaverages and α=0.01,1024 sweeps). Note the difference in abscissa scales. The solid line in (b) is a linear fit to the phase data up to 1 kHz, above which the frequency spacing was too large to prevent whole-cycle phase ambiguity.

IV. DISCUSSION

The specific causal mechanisms for beaked whale mass strandings linked to naval exercises have yet to be identified. Since the most salient feature of these exercises involved the use of relatively high-powered underwater acoustic sources, questions persist about whether beaked whales possess more sensitive auditory systems than other odontocetes. The data presented here, along with previous results from Cook *et al.* (2006), do not reveal any unique auditory system characteristics; instead, these data reveal many similarities between the auditory system of *M. europaeus* and smaller echolocating odontocetes, such as dolphins. The differences that have been observed are most likely a result of the differences in size between the species.

The click-evoked potential waveforms and spectra from the present study were similar to those previously obtained from other odontocetes, with the major differences concerning the amplitudes and latencies of the individual components and the fine-scale temporal and spectral structures. In the present study, latencies of the major peaks in the click-evoked potential waveforms were up to 0.8 ms larger than those reported for dolphins (Popov and Supin, 1985; Popov and Supin, 1990b; Finneran *et al.*, 2007b), within −0.3 to +0.4 ms relative to those in belugas (Popov and Supin, 1990a), and 2.3–2.6 ms smaller than those measured in killer whales (Szymanski *et al.*, 1998). The relationship between the latencies is most likely due to the differences in body size: ~200 kg for dolphins, ~650 kg for belugas, ~700 kg for Gervais' beaked whales, and ~2000 kg for killer whales.

In contrast to the relationships between latencies, where the Gervais' beaked whale data were closer to those of dolphins than to killer whales, P4-N5 amplitudes in the present study (on the order of 1 μV) were substantially smaller than those seen in dolphins (~10 μV, Popov and Supin, 1990a; Popov and Supin, 1990b) and were more similar to those seen in killer whales (~1 μV, Szymanski *et al.*, 1998). These relationships can be potentially explained by the ratio of brain to body mass for each species, which provides a first approximation to relative AEP amplitudes (Szymanski *et al.*, 1998; Supin *et al.*, 2001). Brain mass to body mass ratios for bottlenose dolphins are ~0.009–0.01, while those for Gervais' beaked whales and killer whales are ~0.001–0.003 (Ridgway and Brownson, 1984; Marino, 1998; Szymanski *et al.*, 1998), suggesting that AEPs in *M. europaeus* would be expected to be roughly three to ten times smaller than those in dolphins, but about the same amplitude as those in killer whales.

The MRTF amplitude and phase angle data for *M. europaeus* were also similar to those of other odontocetes, indicating similar capabilities for following the temporal envelopes of rapidly fluctuating sounds. The MRTF amplitude data were nearly identical to those presented by Cook *et al.* (2006) and confirmed that 1 kHz was an appropriate choice for the modulation rate during threshold testing.

The upper cutoff frequency of hearing was lower than that typically observed in bottlenose dolphins (Johnson, 1967; Brill *et al.*, 2001; Houser and Finneran, 2006b; Houser *et al.*, 2008) and was comparable to that observed in killer whales (Szymanski *et al.*, 1999). The observed range of hearing encompasses the reported frequency range of echolocation clicks for beaked whales (Johnson *et al.*, 2006) and is consistent with the need to process echo returns at the frequencies of click production; however, since no other data exist for the upper cutoff frequency in *M. europaeus*, it is impossible to identify if this individual possessed normal hearing for the species. Age-related hearing loss (presbycusis), as well as more profound hearing deficits, potentially related to genetic or environmental factors, has been observed in bottlenose dolphins (Ridgway and Carder, 1997; Andre *et al.*, 2003; Houser and Finneran, 2006b; Houser *et al.*, 2008). Given that the subject of this study was a mature adult, it might have suffered from age-related hearing loss. Other factors may have also contributed to hearing loss, including those directly related to the stranding. Although the hematoma was probably too mild to have affected hearing, necropsy results revealed meningeal thickening due to diffuse collagen bundles interspersed with edema and ceroid-lipofuscin laden macrophages. The meningeal fibrosis may represent an age-related change in this species or be sequelae to meningeal inflammation or hemorrhage. The variable staining of the lipopigment in the different sections of the brain suggests that there was compositional variation in the deposits, possibly reflecting different pathogenic mechanisms in their generation. It is unknown if meningeal fibrosis accumulation of this magnitude would be sufficient to result in sensorineural hearing loss as has been documented in humans and animals subsequent to meningitis. Bacterial meningitis is known to be a potential cause of sensorineural hearing loss in humans and terrestrial mammals (Martini and Trevisi, 2004; Klein *et al.*, 2008; Ruben, 2008); it is un-

known if meningitis results in similar pathologies in cetaceans. For these reasons, the upper cutoff frequency identified for the subject of the present study must be treated with some caution; additional hearing tests of *M. europaeus* will be required to determine the normal range of hearing for this species.

Evoked potential testing of stranded/rehabilitating cetaceans presents a variety of challenges. Maintaining control of large, wild, aquatic animals, such as the beaked whale tested in this study (~1600 lb), is difficult. As a result, physiological background noise within the electroencephalogram tends to be somewhat high. For this study, spectral amplitudes of the background physiological noise (assessed at frequencies near 1 kHz after 1024 averages) were typically 10–15 nV and the residual background noise, estimated using the "sweep-to-sweep variance" method (Elberling and Don, 1984), ranged from 80–120 nV rms after 1024 sweeps. Although these noise levels are not excessively high, when coupled with small AEP amplitudes (presumably because of the large size of the subject), the result was a generally low signal-to-noise ratio (e.g., Fig. 2), making detections difficult near threshold.

Another constraint to testing stranded/rehabilitating marine mammals is the uncertainty regarding the amount of time the subject will be available. Scientific research permits for working with wild cetaceans may prescribe fixed time limits and testing may be interrupted or suspended at any moment if there are concerns regarding the animal's health. Under these circumstances, tradeoffs are often made to balance the fidelity of the acquired data against the time required to obtain the data. For the present study, the upper cutoff frequency was of prime interest, so limited time was spent in adjusting electrode and jawphone positions to optimize the measured AEPs. Once a satisfactory signal was obtained, threshold testing began. It is possible that further manipulation of electrode and jawphone positions could have resulted in larger AEP amplitudes, though, based on the subject's size and the brain/body mass comparisons above, it seems unlikely that the potential increase in amplitude would have been large.

For this study, a jawphone was used because of the freedom the animal had to move its head; regardless of whale movements, the jawphone would maintain the sound source at the same position relative to the animal. Unfortunately, absolute thresholds obtained with jawphones have limited meaning unless validated for the species via behavioral or direct-field threshold measurements. Since this has not been done for *M. europaeus*, the thresholds presented in this paper must be considered only as estimates. If future opportunities provide for longer term rehabilitation of a stranded beaked whale (and thus opportunities for more thorough testing), testing should be performed using direct-field stimulation and include lower frequencies as well. The specific threshold techniques would likely need to be modified at the lower frequencies, where SAM tones become increasingly less effective at producing the ASSR. In these cases, repetitive or single tone pips may yield better results (e.g., Popov *et al.*, 2007).

V. CONCLUSIONS

The upper limit for effective hearing in a stranded Gervais' beaked whale was 80–90 kHz, substantially lower than that seen in dolphins (~120–150 kHz), but similar to that measured in killer whales. Since this is the only beaked whale for which an upper limit of hearing has been identified, it is not known if this is normal or if this individual possessed high-frequency hearing loss. Other features of the auditory system, such as click-evoked potentials and the MRTF, were similar to those of dolphins and other odontocetes.

ACKNOWLEDGMENTS

The volunteers and staff of the MMC provided valuable technical and logistical support for the hearing tests. T. Rowles, Coordinator of the Marine Mammal Health and Stranding Response Program, was responsible for initially coordinating the AEP tests between the authors and the regional stranding network. Genetic identification was performed by P. Rosel and A. Viricel at the NMFS SEFSC Marine Mammal Molecular Genetics Laboratory. This work was performed under National Marine Fisheries Permit No. 1095-1837-00. Financial support was provided by the U.S. Office of Naval Research.

Andre, M., Supin, A., Delory, E., Kamminga, C., Degollada, E., and Alonso, J. M. (**2003**). "Evidence of deafness in a striped dolphin, *Stenella coeruleoalba*," Aquat. Mamm. **29**, 3–8.

Brill, R. L., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). "Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones," J. Acoust. Soc. Am. **109**, 1717–1722.

Brillinger, D. R. (**1978**). "A note on the estimation of evoked response," Biol. Cybern. **31**, 141–144.

Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., Houser, D., and Mann, D. A. (**2006**). "Beaked whale auditory evoked potential hearing measurements," J. Comp. Physiol. [A] **192**, 489–495.

Dobie, R. A., and Wilson, M. J. (**1989**). "Analysis of auditory evoked potentials by magnitude-squared coherence," Ear Hear. **10**, 2–13.

Dobie, R. A., and Wilson, M. J. (**1996**). "A comparison of *t* test, *F* test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am. **100**, 2236–2246.

Dolphin, W. F. (**1995**). "Modulation rate transfer functions to low-frequency carriers in three species of cetaceans," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 25–47.

Dolphin, W. F., Au, W. W., Nachtigall, P. E., and Pawloski, J. (**1995**). "Modulation rate transfer functions to low-frequency carriers in three species of cetaceans," J. Comp. Physiol. [A] **177**, 235–245.

Dolphin, W. F., and Mountain, D. C. (**1992**). "The envelope following response: Scalp potentials elicited in the Mongolian gerbil using sinusoidally AM acoustic signals," Hear. Res. **58**, 70–78.

Elberling, C., and Don, M. (**1984**). "Quality estimation of averaged auditory brainstem responses," Scand. Audiol. **13**, 187–197.

Fernández, A., Edwards, J., Martín, V., Rodríguez, F., Espinosa de los Monteros, A., Herráez, P., Castro, P., Jaber, J. R., and Arbelo, M. (**2005**). " 'Gas and fat embolic syndrome' involving a mass stranding of beaked whales exposed to anthropogenic sonar signals," Vet. Pathol. **42**, 446–457.

Finneran, J. J. (**2008**). "Evoked response study tool (EVREST) user's guide," SSC San Diego Technical Document No. 3226 (SSC San Diego, San Diego, CA).

Finneran, J. J. (**2009**). "Evoked response study tool (EVREST): A portable, rugged system for single and multiple auditory evoked potential measurements," J. Acoust. Soc. Am. (in press).

Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose

dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **119**, 3181–3192.
Finneran, J. J., Houser, D. S., and Schlundt, C. E. (**2007a**). "Objective detection of bottlenose dolphin (*Tursiops truncatus*) steady-state auditory evoked potentials in response to AM/FM tones," Aquat. Mamm. **33**, 43–54.
Finneran, J. J., London, H. R., and Houser, D. S. (**2007b**). "Modulation rate transfer functions in bottlenose dolphins (*Tursiops truncatus*) with normal hearing and high-frequency hearing loss," J. Comp. Physiol. [A] **193**, 835–843.
Finneran, J. J., Schlundt, C. E., Branstetter, B., and Dear, R. L. (**2007c**). "Assessing temporary threshold shift in a bottlenose dolphin (*Tursiops truncatus*) using multiple simultaneous auditory evoked potentials," J. Acoust. Soc. Am. **122**, 1249–1264.
Houser, D. S., and Finneran, J. J. (**2006a**). "A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods," J. Acoust. Soc. Am. **120**, 1713–1722.
Houser, D. S., and Finneran, J. J. (**2006b**). "Variation in the hearing sensitivity of a dolphin population obtained through the use of evoked potential audiometry," J. Acoust. Soc. Am. **120**, 4090–4099.
Houser, D. S., Gomez-Rubio, A., and Finneran, J. J. (**2008**). "Evoked potential audiometry of 13 Pacific bottlenose dolphins (*Tursiops truncatus gilli*)," Marine Mammal Sci. **24**, 28–41.
Houser, D. S., Howard, R., and Ridgway, S. (**2001**). "Can diving-induced tissue nitrogen supersaturation increase the chance of acoustically driven bubble growth in marine mammals?," J. Theor. Biol. **213**, 183–195.
Jepson, P. D., Arbelo, M., Deaville, R., Patterson, I. A. R., Castro, P., Baker, J. R., Degollada, E., Ross, H. M., Herráez, P., Pocknell, A. M., Rodriguez, E., Howie, F. E., Espinosa, A., Reid, R. J., Jaber, J. R., Martin, V., Cunningham, A. A., and Fernandez, A. (**2003**). "Gas-bubble lesions in stranded cetaceans," Nature (London) **425**, 575–576.
Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 247–260.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., and Tyack, P. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densitorstris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.
Johnson, M. P., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.
Klein, M., Koedel, U., Kastenbauer, S., and Pfister, H. W. (**2008**). "Nitrogen and oxygen molecules in meningitis-associated labyrinthitis and hearing impairment," Infection **36**, 2–14.
Klishin, V. O., Popov, V. V., and Supin, A. Y. (**2000**). "Hearing capabilities of a beluga whale, *Delphinapterus leucas*," Aquat. Mamm. **26**, 212–228.
Madsen, P. T., Johnson, M., Aguilar de Soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.
Marino, L. (**1998**). "A comparison of encephalization between odontocete cetaceans and anthropoid primates," Brain Behav. Evol. **51**, 230–238.
Martini, A., and Trevisi, P. (**2004**). "Classification and epidemiology," in *Genetic Hearing Loss*, edited by P. J. Willems (Dekker, New York), pp. 49–64.
Mooney, T. A., Nachtigall, P. E., and Yuen, M. M. L. (**2006**). "Temporal resolution of the Risso's dolphin, *Grampus griseus*, auditory system," J. Comp. Physiol. [A] **192**, 373–380.
Nachtigall, P. E., Yuen, M. M. L., Mooney, T. A., and Taylor, K. A. (**2005**). "Hearing measurements from a stranded infant Risso's dolphin, *Grampus griseus*," J. Exp. Biol. **208**, 4181–4188.
Popov, V., and Supin, A. (**1990a**). "Electrophysiological studies of hearing in some cetaceans and a manatee," in *Sensory Abilities in Cetaceans*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 405–415.
Popov, V. V., and Supin, A. Y. (**1985**). "Determining hearing characteristics in dolphins using evoked potentials of brain stem," Dokl. Akad. Nauk SSSR **283**, 496–499.
Popov, V. V., and Supin, A. Y. (**1990b**). "Auditory brainstem responses in characterization of dolphin hearing," J. Comp. Physiol. [A] **166**, 385–393.
Popov, V. V., Supin, A. Y., Pletenko, M. G., Tarakanov, M. B., Klishin, V. O., Bulgakova, T. N., and Rosanova, E. I. (**2007**). "Audiogram variability in normal bottlenose dolphins (*Tursiops truncatus*)," Aquat. Mamm. **33**, 24–33.
Ridgway, S. H., and Brownson, R. H. (**1984**). "Relative brain sizes and cortical surface areas in odontocetes," Acta Zool. Fenn. **172**, 149–152.
Ridgway, S. H., and Carder, D. A. (**1997**). "Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin," J. Acoust. Soc. Am. **101**, 590–594.
Ruben, R. (**2008**). "Bacterial meningitic deafness: Historical development of epidemiology and cellular pathology," Acta Oto-Laryngol. **128**, 388–392.
Schlundt, C. E., Dear, R. L., Green, L., Houser, D. S., and Finneran, J. J. (**2007**). "Simultaneously measured behavioral and electrophysiological hearing thresholds in a bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **122**, 615–622.
Stapells, D. R., Makeig, S., and Galambos, R. (**1987**). "Auditory steady-state responses: threshold prediction using phase coherence," Electroencephalogr. Clin. Neurophysiol. **67**, 260–270.
Supin, A. J., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, Boston, MA).
Supin, A. Y., and Popov, V. V. (**1995**). "Envelope-following response and modulation transfer function in the dolphin's auditory system," Hear. Res. **92**, 38–46.
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.
Szymanski, M. D., Supin, A. Y., Bain, D. E., and Henry, K. R. (**1998**). "Killer whale (*Orcinus orca*) auditory evoked potentials to rhythmic clicks," Marine Mammal Sci. **14**, 676–691.
Tyack, P. L., Johnson, M., Soto, N. A., Sturlese, A., and Madsen, P. T. (**2006**). "Extreme diving of beaked whales," J. Exp. Biol. **209**, 4238–4253.
U.S. Department of Commerce and U.S. Department of the Navy (**2001**). *Joint Interim Report Bahamas Marine Mammal Stranding Event*, 14–16 March (Department of Commerce, Washington, DC).
Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **118**, 2688–2695.
Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.

Evoked response study tool: A portable, rugged system for single and multiple auditory evoked potential measurements

James J. Finneran
U.S. Navy Marine Mammal Program, SSC Pacific, Code 71510, 53560 Hull Street, San Diego, California 92152

(Received 1 January 2009; revised 6 April 2009; accepted 12 May 2009)

Although the potential of using portable auditory evoked potential systems for field testing of stranded cetaceans has been long recognized, commercial systems for evoked potential measurements generally do not possess the bandwidth required for testing odontocete cetaceans and are not suitable for field use. As a result, there have been a number of efforts to develop portable evoked potential systems for field testing of cetaceans. This paper presents another such system, called the evoked response study tool (EVREST). EVREST is a Windows-based hardware/software system designed for calibrating sound stimuli and recording and analyzing transient and steady-state evoked potentials. The EVREST software features a graphical user interface, real-time analysis and visualization of recorded data, a variety of stimulus options, and a high level of automation. The system hardware is portable, rugged, battery-powered, and possesses a bandwidth that encompasses the audible range of echolocating odontocetes, making the system suitable for field testing of stranded or rehabilitating cetaceans.
 [DOI: 10.1121/1.3148214]

PACS number(s): 43.80.Vj, 43.80.Lb, 43.64.Ri [MCH] Pages: 491–500

I. INTRODUCTION

Studies of marine mammal audition have traditionally relied on behavioral response paradigms. In these techniques, subjects are trained to produce a specific behavioral response when presented with a sound stimulus and to withhold this response, or produce a different response, in the absence of the stimulus. Behavioral methods are well understood; however, the amount of time and level of access required to train subjects have limited the number of species for whom testing is practical and the number of individuals within a species for whom testing has been conducted.

To overcome the inherent limitations of behavioral techniques, measurements of auditory evoked potentials (AEPs) are becoming increasingly popular. AEPs are small voltages automatically generated by the brain in response to sound stimuli. AEPs reflect neuronal discharges that are synchronized with the onset of a sound and thus provide information about the effects of sound on the neural activity within the auditory pathway. AEP measurements have become important clinical tools for auditory screening of infants and difficult to test human patients. Since they do not require specific training for hearing tests, AEP measurements have also been increasingly used to examine auditory functions in marine mammals (see Dolphin, 1997, 2000; Supin *et al.*, 2001), to test larger numbers of subjects (Houser and Finneran, 2006b; Popov *et al.*, 2007; Houser *et al.*, 2008), and to obtain data from previously untested species (Szymanski *et al.*, 1999; Nachtigall *et al.*, 2005; Popov *et al.*, 2005; Cook *et al.*, 2006; Nachtigall *et al.*, 2007, 2008).

The potential of using portable AEP systems for field testing of stranded cetaceans was recognized early (e.g., Ridgway and Carder, 1983). Although there are portable, commercial off-the-shelf (COTS) systems for human and terrestrial mammal AEP measurements, these systems are primarily designed for clinical settings and are not ideal for field measurements. More significantly, the frequency responses of commercial systems do not typically extend to the ultrasonic range required for testing odontocetes (i.e., up to 150–200 kHz). As a result, there have been a number of efforts to develop custom, portable AEP systems for field testing of marine mammals, particularly cetaceans (e.g., Ridgway and Carder, 1983; Carder and Ridgway, 1994; Helweg *et al.*, 1997; Finneran and Houser, 2004; Taylor *et al.*, 2006; Delory *et al.*, 2007; Taylor *et al.*, 2007). This paper presents another such system, called the evoked response study tool (EVREST). The primary differences between the current AEP system and its predecessors include the (1) rugged nature, compact size, and level of hardware integration; (2) variety of possible stimuli; (3) capability for offline processing of the "raw" evoked potential data stream; (4) inclusion of a variety of objective response detection (ORD) methods; (5) potential for automated measurements; and the (6) capability of very high sample and update rates. This paper describes the EVREST hardware and software components and presents some example data collected with the system. As a complete description of the EVREST software is beyond the scope of this paper, only the most significant features are described here; complete details may be found in Finneran, 2008.

II. BACKGROUND

A. Transient and steady-state evoked potentials

The EVREST system is designed to calibrate sound stimuli, measure single and multiple transient and steady-state AEPs, analyze the resulting data, and archive tabular and graphical results. Transient evoked responses are typi-

cally elicited by clicks or tone-pips. Steady-state evoked potentials are produced when stimuli are presented at a sufficiently high rate that transient AEPs overlap and produce a steady-state response, called the auditory steady-state response (ASSR) or envelope following response. The ASSR may be generated by repetitive stimuli (e.g., sequences of clicks or tone-pips), as well as amplitude modulated (AM) or frequency modulated (FM) tones. The ASSR is a harmonic signal with a fundamental frequency related to the modulation rate of sinusoidally modulated stimuli or the repetition rate of click/tone-pip sequences; thus ASSRs may be analyzed in the frequency domain by examining the spectral amplitude at the appropriate frequency. If multiple sinusoidally modulated stimuli, each with a unique carrier and modulation rate, are presented simultaneously, the ASSR to each carrier frequency occurs at the corresponding modulation rate. The responses to the individual stimulus components may therefore be independently tracked and used to simultaneously assess hearing at multiple frequencies. This technique is called the multiple ASSR method, and has been used to test hearing in a number of terrestrial and marine mammals (e.g., Picton *et al.*, 1987; Regan and Regan, 1988; Lins *et al.*, 1995; Lins and Picton, 1995; Dolphin, 1996; Popov *et al.*, 1997, 1998; Finneran and Houser, 2007; Finneran *et al.*, 2007). If sufficient isolation exists between the left/right ears, the multiple ASSR technique may also be used to test both ears simultaneously (Lins and Picton, 1995; John *et al.*, 1998).

B. Objective measurement of evoked potentials

Evoked potentials are measured using surface or subcutaneous electrodes placed on the head and neck or back, with the specific style varying with species. AEP amplitudes depend on the species and stimulus, but are generally less than a few microvolts and are therefore dominated by the background physiological noise. For this reason, hundreds or thousands of stimuli are typically presented, and the evoked responses synchronously averaged, to improve the signal to noise ratio (under the assumptions of a deterministic evoked response and stationary background noise).

For threshold testing, ORD procedures are often used to determine the presence/absence of the evoked response. ORD techniques can be performed in the time domain or frequency domain. Time domain ORD methods include those based on template matching, correlation, and statistical tests based on variance ratios (Schimmel *et al.*, 1974; Mason *et al.*, 1977; Weber and Fletcher, 1980; Elberling and Don, 1984; Supin *et al.*, 2001). Frequency domain methods compare the AEP amplitude and/or phase in some specified frequency band to a noise estimate obtained either from a control trial or concurrently with the AEP measurement (Dobie and Wilson, 1993, 1995). Frequency domain ORD techniques are particularly attractive for use with the ASSR since it possesses a known fundamental frequency (Dobie and Wilson, 1994).

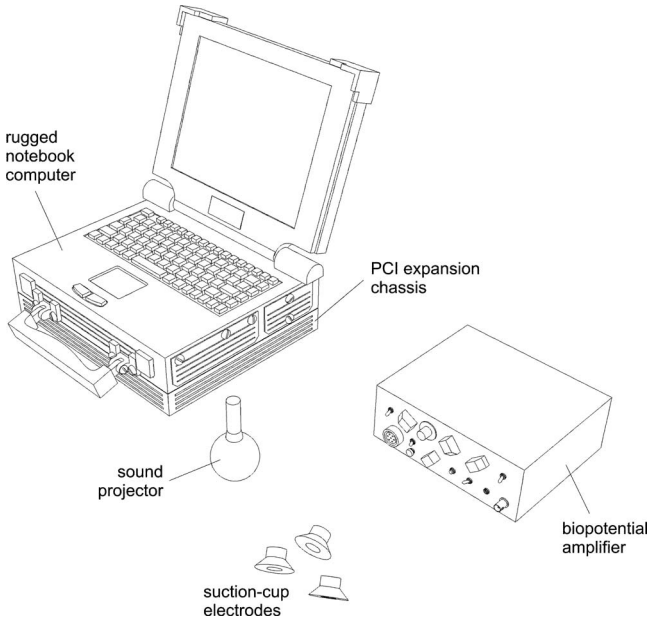


FIG. 1. Schematic of the EVREST hardware system configured for cetacean testing. The system consists of four main components: (1) a rugged notebook computer with PCI expansion chassis, (2) biopotential amplifier, (3) sound projector, and (4) electrodes.

C. Example data

To illustrate the operation of the EVREST system, AEPs were measured in a 24-year old female bottlenose dolphin (*Tursiops truncatus*). Evoked potential measurements were conducted in-air with the subject resting on a foam mat. Sounds were presented to the subject using a "jawphone"—a piezoelectric sound projector (ITC 1042) embedded in a silicon suction cup—attached to the subject's left jaw in the pan region. Received sound pressure levels (SPLs) were estimated from underwater direct field measurements with a calibrated hydrophone (B&K 8105) placed 15 cm from the jawphone face. In-air evoked potential thresholds measured with jawphones calibrated in this fashion have provided reasonable approximations to underwater behavioral thresholds in five dolphins, especially with respect to audible bandwidth (Finneran and Houser, 2006; Houser and Finneran, 2006a). Since the primary intent of the measurements was to demonstrate the operation of the EVREST system, the specific calibration method is of secondary importance and the received levels cited here should be considered approximate.

III. HARDWARE SYSTEM

The EVREST hardware (Fig. 1) consists of four primary components: (1) a rugged notebook computer with a multifunction data acquisition (DAQ) card and custom signal conditioning circuitry, (2) a biopotential amplifier, (3) a sound projector, and (4) electrodes. The biopotential amplifier is a COTS model (Grass ICP-511) with adjustable gain (up to 2×10^5) and bandpass filters. The biopotential amplifier is housed in a watertight plastic case (IP67, MIL C-4150J) with external input/output (I/O) connectors. Power is supplied from two 12 V, 2.3 A h, lead-acid batteries. The total size of the plastic case is $30\times23\times13$ cm^3; the total weight of the biopotential amplifier, batteries, and case is 5 kg (11 lb).

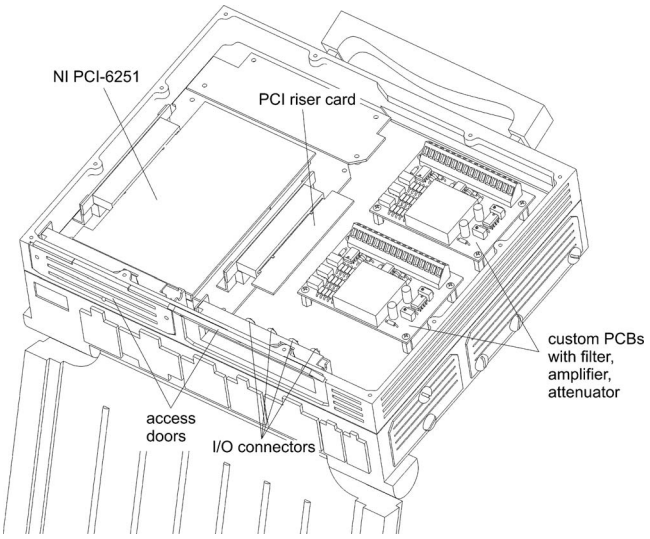


FIG. 2. Bottom view of the EVREST notebook PCI chassis. The chassis holds a multifunction PCI DAQ card and custom electronic filters, programmable attenuators, and amplifiers.

Battery life under normal use is >8 h. The specific style of sound projector depends on the application: Underwater tests have been performed using moving coil and piezoelectric underwater sound projectors; airborne tests have been performed using moving coil headphones (pinnipeds) and piezoelectric transducers embedded in suction cups (odontocetes). The specific electrode style also depends on the application, with surface electrodes embedded in suction cups used on odontocetes and subcutaneous electrodes used on pinnipeds. For the example data presented here, three 10-mm gold cup electrodes embedded in suction cups were used: The noninverting electrode was placed 5–10 cm behind the blowhole near the midline, the inverting electrode was placed just behind the right external auditory meatus (contralateral to the jawphone), and the ground electrode was placed on the back near the dorsal fin.

The rugged notebook computer is a COTS model with a 1.6 GHz Pentium M processor and 512 Mbytes of random access memory. The laptop chassis is an IP 54/NEMA 12 enclosure with sealed ports and connectors, sealed rubber keyboard and trackpad, and 12.1-in. sunlight readable screen. The computer features a 97 W h Li-ion main battery and a 65 W h Li-ion auxiliary battery. The batteries slide into recessed compartments in the chassis and have integral, sealed access doors matching the chassis exterior. The batteries are held by thumbscrews; thus they may be easily replaced in the field (though the computer must be turned off). The most salient feature of the computer is an expansion chassis capable of holding two 3/4-length peripheral component interconnect (PCI) cards (Fig. 2). Within the expansion chassis, one PCI slot is occupied by a multifunction DAQ card (National Instruments PCI-6251) with analog and digital I/O. The DAQ card features 16-bit analog inputs and output, with output update rates up to 2 MHz and an input sampling rate up to 1.25 MHz. The area near the second PCI slot is occupied by two custom printed circuit boards (PCBs), each containing a filter, amplifiers, and digitally-controlled attenuators. Power for the PCBs is provided by extracting ± 12 V from the second PCI slot via a modified right-angle PCI extender card. External connections to the analog I/O signals are made using panel-mounted connectors (IP67/NEMA 6P) accessed via a door at the rear of the chassis. Analog and digital signals to/from the DAQ card are internally routed within the expansion chassis, allowing the access door near the DAQ card to remain closed. The total size of the computer with expansion chassis is $31 \times 25 \times 12$ cm^3; the weight of the unit, including the DAQ card, custom PCBs, and the primary and auxiliary batteries, is 7.7 kg (17 lb). Battery life with the primary and auxiliary batteries under normal use is approximately 4 h.

Figure 3 is a block diagram of the signal flow from one output channel through the signal conditioning PCB. Attenuation is provided using individual 10, 20, and 40 dB attenuation stages, each consisting of a "bridged-T" style resistor network. Solid state relays (Omron G6K) are used to switch individual stages in/out of the circuit, allowing attenuations from 10 to 70 dB in 10 dB steps. The relays are controlled by digital outputs from the PCI-6251. The EVREST software automatically engages the relays as necessary to keep the minimum DAC output voltage above a user-specified limit. The signal conditioning PCBs accept a variety of fixed-frequency modules (Krohn-Hite 3A, 3B, 3C series); the primary role of the filter is to eliminate components at or above one-half the update rate. The data presented in this paper were collected with an eight-pole Butterworth style, low-pass filter with cutoff at 200 kHz. The line driver allows reactive loads such as piezoelectric transducers to be operated without distortion.

The frequency response and linearity of the system were assessed by varying the stimulus voltage from +10 to −110 dBV and the frequency from 1 Hz to 200 kHz and measuring the resulting output voltage. Figure 4(a) shows the frequency response of the system with no attenuation. In this particular case, the high-frequency response is limited by the 200 kHz low-pass filter, while the low-frequency response remains flat down to 1 Hz. The voltage drop across the attenuator

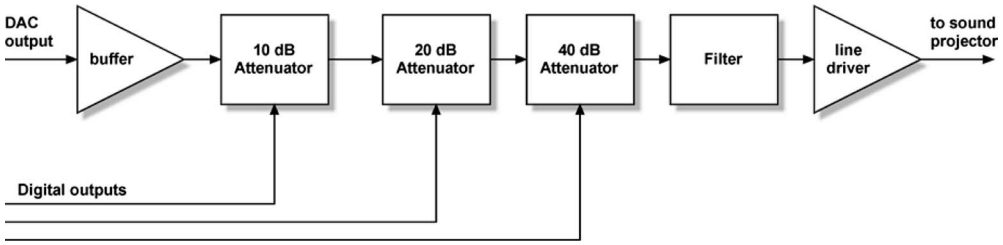


FIG. 3. Block diagram of output electronics for a single output channel.

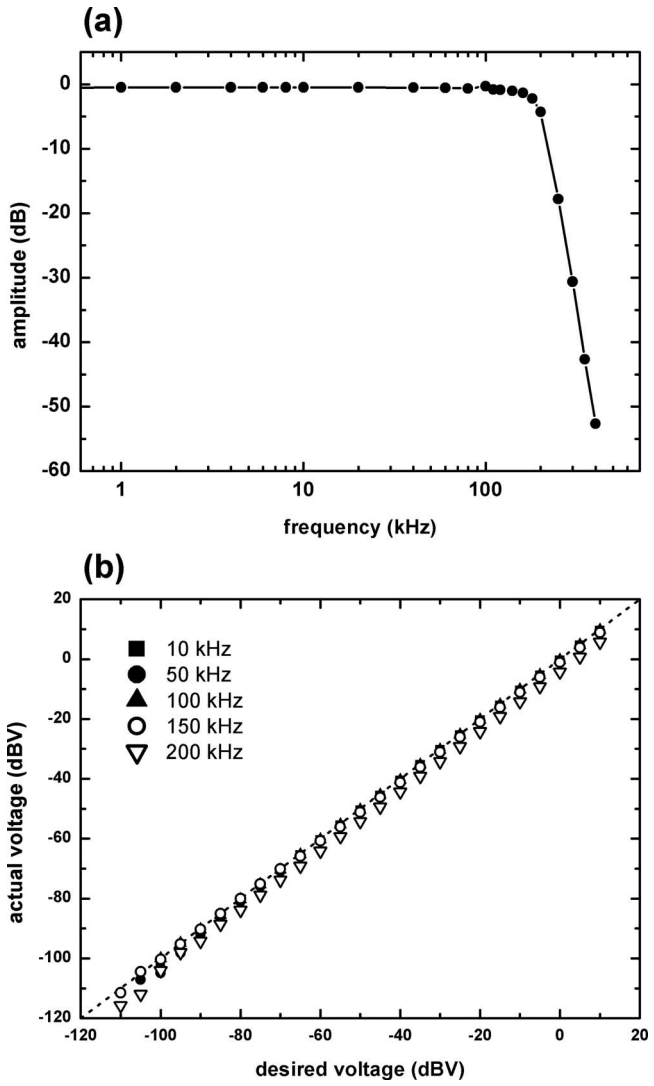


FIG. 4. (a) Frequency response of the output system with a 200 kHz low-pass filter. The low-frequency response remains flat to down to at least 1 Hz. (b) Linearity of output signal at 10–200 kHz.

network, without any of the attenuators engaged (i.e., no desired attenuation), is approximately 0.5 dB. Figure 4(b) shows the linearity and dynamic range of the output signal at 10, 50, 100, 150, and 200 kHz. The dynamic range is approximately 110 dB. The accuracy of the attenuators decreases with increasing frequency; however, the measured attenuations are still within 1 dB of the desired values at frequencies up to 200 kHz.

IV. SOFTWARE APPLICATION

The EVREST software is a stand-alone application, written in the LABVIEW® graphical programming language (National Instruments Corporation, 2008) for the Microsoft Windows operating system. Although designed for use with the EVREST hardware, the software is compatible with any multifunction DAQ card that supports the NI-DAQMX application programming interface. The EVREST front panel uses "tab" structures to group related controls and indicators. The main user interface tab switches between the "Measure" screen, used for DAQ, and the "Analyze" screen, used to view, analyze, and archive previously acquired data.

A. Measure screen

1. Measurement overview

Measurements rely on stimulus generation from one or two channels and simultaneous DAQ. To generate the sound stimuli, digital representations of the stimulus waveforms are converted to analog signals and generated at the specified analog output channels on the DAQ device. As stimuli are generated, the signal at the specified analog input channel is digitized and streamed into a memory buffer. Digitized samples are read from the buffer one sweep at a time, written to the computer hard disk, and used to compute the time average of the waveform. Sweeps whose peak amplitude exceeds a user-defined reject level are excluded from averaging. A number of initial sweeps may also be automatically rejected (typically used with continuous stimuli to ignore the initial latency in the AEP). Both normal and weighted averaging are supported (Elberling and Wahlgreen, 1985). At user-defined intervals, the signal amplitude is computed and, for AEP measurements, ORD calculations are performed. The measurement is concluded when the desired number of sweeps is obtained, when a response is detected, or when the residual background noise is reduced below a user-defined value (Elberling and Don, 1984; Don and Elberling, 1994).

Figure 5 shows an example of the Measure screen after a click-evoked potential was measured in the dolphin subject. The upper right area shows the grand average of the AEP waveform. The lower right area contains a tab structure used to switch between a variety of output displays (described below). The region at the top left contains six tabs (Stimuli, AEP, SPL, Autosave, Automation, and Setup) used to define parameters related to generating sound stimuli, calibrating stimuli, measuring AEPs, and saving the data.

2. Generating stimuli

The sound stimulus for each output channel is defined on the Stimuli tab (visible in Fig. 5). There are ten individual waveform components, each of which can be independently switched on/off and assigned to either output channel (but not to both). The parameters for the waveform components are defined on individual tabs; parameters for all ten components may also be simultaneously viewed or edited using a spreadsheet-style table (Fig. 6). The basic waveform options for each component consist of: a rectangular click, tone-pip (user-specified number of rise/fall and plateau cycles), pure tone, or FM tone (linear up/down sweep or sinusoidal FM). Amplitude modulation, either carrier plus sidebands or carrier-suppressed sidebands, can then be applied to any of the basic waveforms. The duration, rise/fall characteristics, polarity, and starting time of each waveform component can be independently specified. The polarity of the composite waveform for each channel (i.e., the sum of the waveform components assigned to the same channel) can also be flipped on alternate sweeps, a technique often used to eliminate stimulus artifacts from the averaged AEP. As an alternative to defining stimuli from individual waveform compo-

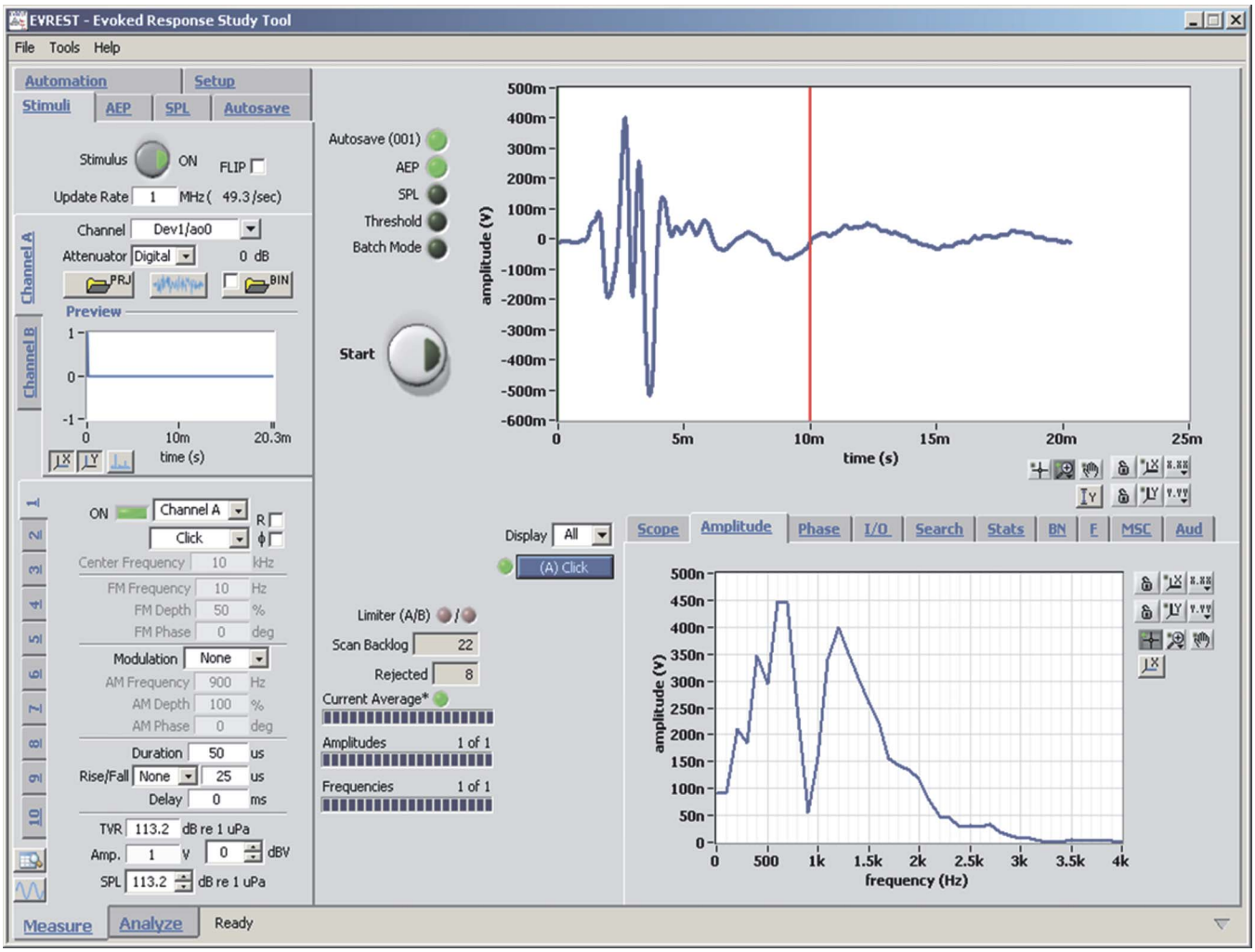


FIG. 5. (Color online) The EVREST software Measure screen, used for generating stimuli, calibrating stimuli, and recording evoked responses.

nents, binary files created with EVREST or third party software can be loaded from disk, allowing the use of any user-defined stimulus waveform.

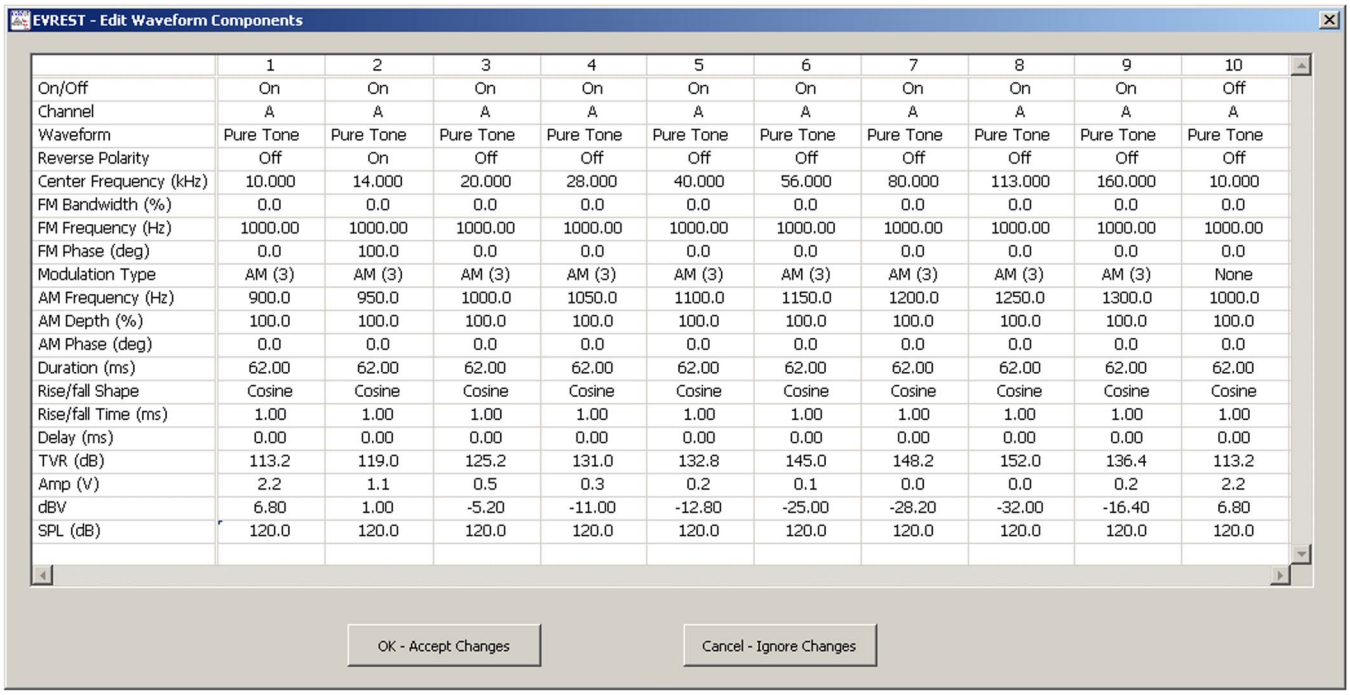
To generate multiple stimuli at high presentation rates, EVREST uses a circular buffer to hold the stimulus waveform for each channel. The values in the buffer are sequentially generated; when the last value is produced, generation continues from the beginning of the buffer. For stimuli with durations shorter than the recording sweep, the buffer is extended to match the length of the recording sweep. The circular buffer allows stimulus presentation rates up to the theoretical maximum (i.e., the reciprocal of the recording sweep). Proper choice of stimulus duration and sweep duration also allows continuous generation of a stimulus with no gaps between successive sweeps.

The stimulus for the click-evoked potential shown in Fig. 5 used only a single waveform component to define a compressive click assigned to output channel A. Figure 6 illustrates a more complex stimulus defined using the tabular view, where each column corresponds to a different waveform component. This stimulus features nine separate sinusoidal AM tone components, with carrier frequencies spaced at 1/2-octave intervals from 10 to 160 kHz.

3. Calibrating stimuli and measuring AEPs

DAQ parameters are defined using either the AEP tab [Fig. 7(a), for evoked potential measurements] or the SPL tab [Fig. 7(b), for stimulus calibration]. Both types of measurements work in a similar fashion, and many of the same DAQ parameters exist on the AEP and SPL tabs. The primary differences between AEP and SPL measurements involve the manner in which the recorded data are analyzed.

Stimulus calibration involves measuring the received SPL at the subject for a known stimulus level and calculating the transmitting voltage response (TVR), defined as the SPL produced from a 1-V peak amplitude signal at the DAQ output. The TVRs are used on the Stimuli tab to provide the conversion between DAQ voltage output and received SPL at the subject; they enable the user to specify stimulus levels in terms of SPL rather than voltage. SPL measurements typically involve calculating spectral or time domain amplitudes from the measured grand average, converting the measured voltage to SPL, and displaying the result. The specific amplitude metric is user-selectable from the following options: peak-peak; root-mean-squared; the spectral peak at the center frequency, AM frequency, FM frequency, or twice any of these values; or a "default" option where the stimulus wave-



	1	2	3	4	5	6	7	8	9	10
On/Off	On	On	On	On	On	On	On	On	On	Off
Channel	A	A	A	A	A	A	A	A	A	A
Waveform	Pure Tone	Pure Tone	Pure Tone	Pure Tone	Pure Tone	Pure Tone	Pure Tone	Pure Tone	Pure Tone	Pure Tone
Reverse Polarity	Off	On	Off	Off	Off	Off	Off	Off	Off	Off
Center Frequency (kHz)	10.000	14.000	20.000	28.000	40.000	56.000	80.000	113.000	160.000	10.000
FM Bandwidth (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FM Frequency (Hz)	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00
FM Phase (deg)	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Modulation Type	AM (3)	AM (3)	AM (3)	AM (3)	AM (3)	AM (3)	AM (3)	AM (3)	AM (3)	None
AM Frequency (Hz)	900.0	950.0	1000.0	1050.0	1100.0	1150.0	1200.0	1250.0	1300.0	1000.0
AM Depth (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
AM Phase (deg)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Duration (ms)	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00
Rise/fall Shape	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine	Cosine
Rise/fall Time (ms)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Delay (ms)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TVR (dB)	113.2	119.0	125.2	131.0	132.8	145.0	148.2	152.0	136.4	113.2
Amp (V)	2.2	1.1	0.5	0.3	0.2	0.1	0.0	0.0	0.2	2.2
dBV	6.80	1.00	-5.20	-11.00	-12.80	-25.00	-28.20	-32.00	-16.40	6.80
SPL (dB)	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0

FIG. 6. (Color online) Spreadsheet-style table used to view and edit parameters for all ten waveform components.

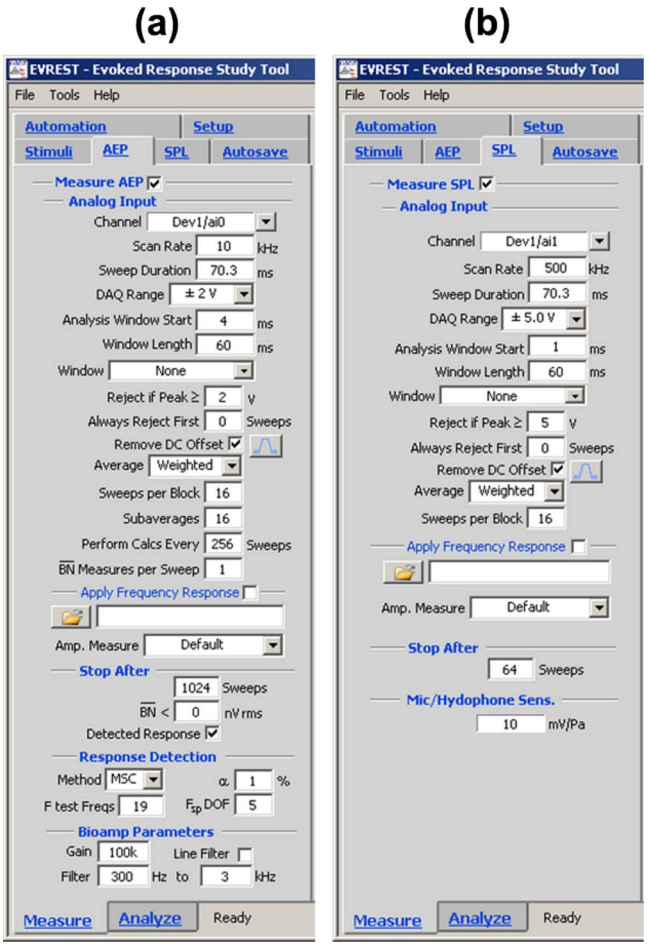


FIG. 7. (Color online) The (a) AEP and (b) SPL tabs from the Measure screen. These tabs are used to define the DAQ and analysis parameters for AEP measurements and stimulus calibrations, respectively.

form determines the amplitude metric. For example, if the stimulus is a sinusoidal AM (carrier plus sidebands) tone, then the default amplitude metric for SPL measurements is the combined band level of the center frequency and the two sidebands. For stimuli consisting of multiple components, SPLs are calculated for each component.

AEP measurements involve not only amplitude and phase measurements, but also estimates of the residual background noise, based on the sweep-to-sweep variance at one or more time latencies (Elberling and Don, 1984; Don and Elberling, 1994), and ORD calculations. The specific AEP amplitude metric is selectable between peak-peak; root-mean-squared; the spectral peak at the center frequency, AM frequency, FM frequency, or twice any of these values; and the default option. For example, if the stimulus is a sinusoidal AM (carrier plus sidebands), the default AEP amplitude is the spectral peak at the AM frequency. Frequency domain ORD techniques include the F test, magnitude-squared coherence, phase coherence, and circular T-squared (Dobie and Wilson, 1989, 1993, 1996); for time domain objective response detection, the modified variance ratio is included (Elberling and Don, 1984; Stürzebecher *et al.*, 2001). For stimuli consisting of multiple components, AEP amplitudes and ORD metrics are independently calculated for each component.

4. Displaying measurement results

The lower right portion of the Measure screen includes a series of tabs used to select between various tables and graphs showing the results of one or more measurements. The Scope tab shows the instantaneous voltage at the DAQ input. This is primarily intended to show the electroencephalogram (EEG) signal before averaging. The Amplitude tab (visible in Fig. 5) and Phase tab show the spectral amplitude

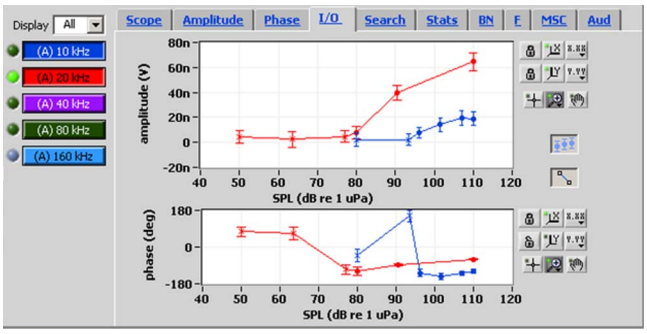


FIG. 8. (Color online) The I/O tab from the Measure screen. This graph shows the measured signal amplitude as a function of the stimulus level. When automation is used, the results of sequential measurements are displayed as I/O functions. The buttons at left are used to toggle the data points on/off for the various waveform components.

and phase, respectively, computed over a user-defined analysis window. These graphs are updated after every 50 sweeps. The signal amplitude is displayed as a function of the stimulus level on the I/O tab (Fig. 8). If automation is used (see below), the results of previous measurements are retained, creating an I/O function, as seen in Fig. 8. A filled symbol represents a detected response and "×" indicates a nondetection (AEP measurements only). If the stimulus contains multiple waveform components, the user selects which particular I/O functions (one or more) to view using a series of buttons. In Fig. 8, I/O functions for the 10 and 20 kHz components are visible.

The BN, F, MSC, and Stats tabs are only active for AEP measurements. The BN tab [Fig. 9(a)] shows the residual background noise for each block and the averaged background noise over the entire measurement (Elberling and Don, 1984; Don and Elberling, 1994). The F tab [Fig. 9(b)] and MSC tab [Fig. 9(c)] provide graphical views of the ORD calculations for the *F* test and magnitude-squared coherence methods, respectively. The F tab provides a visual indication of the relationship between the spectral power at the modulation rate and the critical value of *F* multiplied by the average noise power; if the spectral power exceeds the critical *F* multiplied by the average noise power, the response is considered to be detected. The MSC graph provides a visual representation of the amplitude and phase of the subaverages (plus signs) and grand average (vector starting at the origin), as well as the critical value that must be exceeded for a response to be detected (circle). If the vector extends beyond the circle, the response is detected. If the stimulus contains multiple waveform components, the user selects which particular components (one or more) to view on the F tab and MSC tab. The Stats tab provides a tabular summary of all the ORD calculations.

The Search and Aud (i.e., Audiogram) tabs are only active during AEP threshold measurements (see below). The Search tab shows the stimulus SPL as a function of the measurement number within the automated sequence. The Aud tab shows the estimated thresholds as a function of stimulus center frequency.

5. Saving data

Activating the Autosave feature automatically saves the stimulus and recording parameters, measured amplitude and

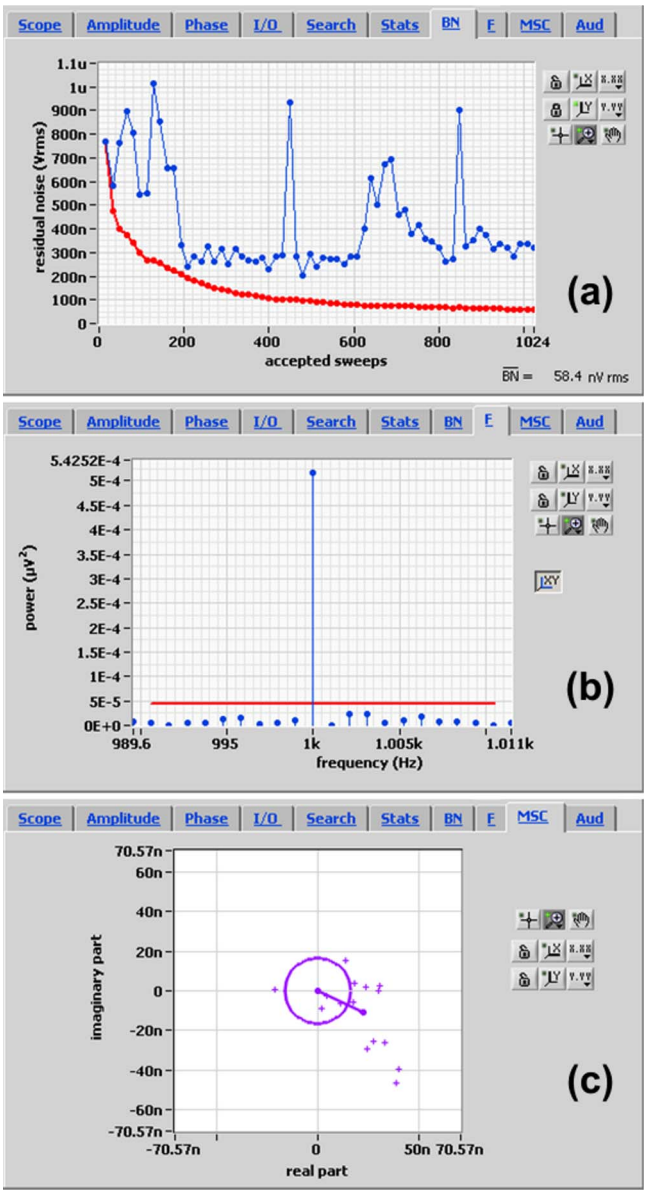


FIG. 9. (Color online) Examples of the (a) BN, (b) F, and (c) MSC tabs from the Measure screen. The BN graph shows the estimated residual background noise for each block as well as the background noise for the averaged signal. The F and MSC tabs display the results of the corresponding ORD test. For the *F* test, the response is detected if the spectral power at the modulation rate exceeds the critical value for *F* multiplied by the estimated noise power (the horizontal line). For the MSC method, the response is detected if the vector starting at the origin extends beyond the circle; this indicates that the power in the grand average at the modulation rate exceeds the critical value for MSC multiplied by the average of the powers in the individual subaverages.

phase, ORD metrics, and the averaged signal waveform to a text file. The text file name consists of a user-defined prefix, stimulus frequencies, and unique number. If desired, the digitized samples from the analog input buffer, before artifact rejection or averaging, may be automatically saved to disk as a binary file. The binary data file represents the raw data stream and may be used during post hoc analyses to recreate the measurement process. This allows one to change analysis settings that require access to the original sweeps, not just the averaged time waveform (e.g., changes to the artifact

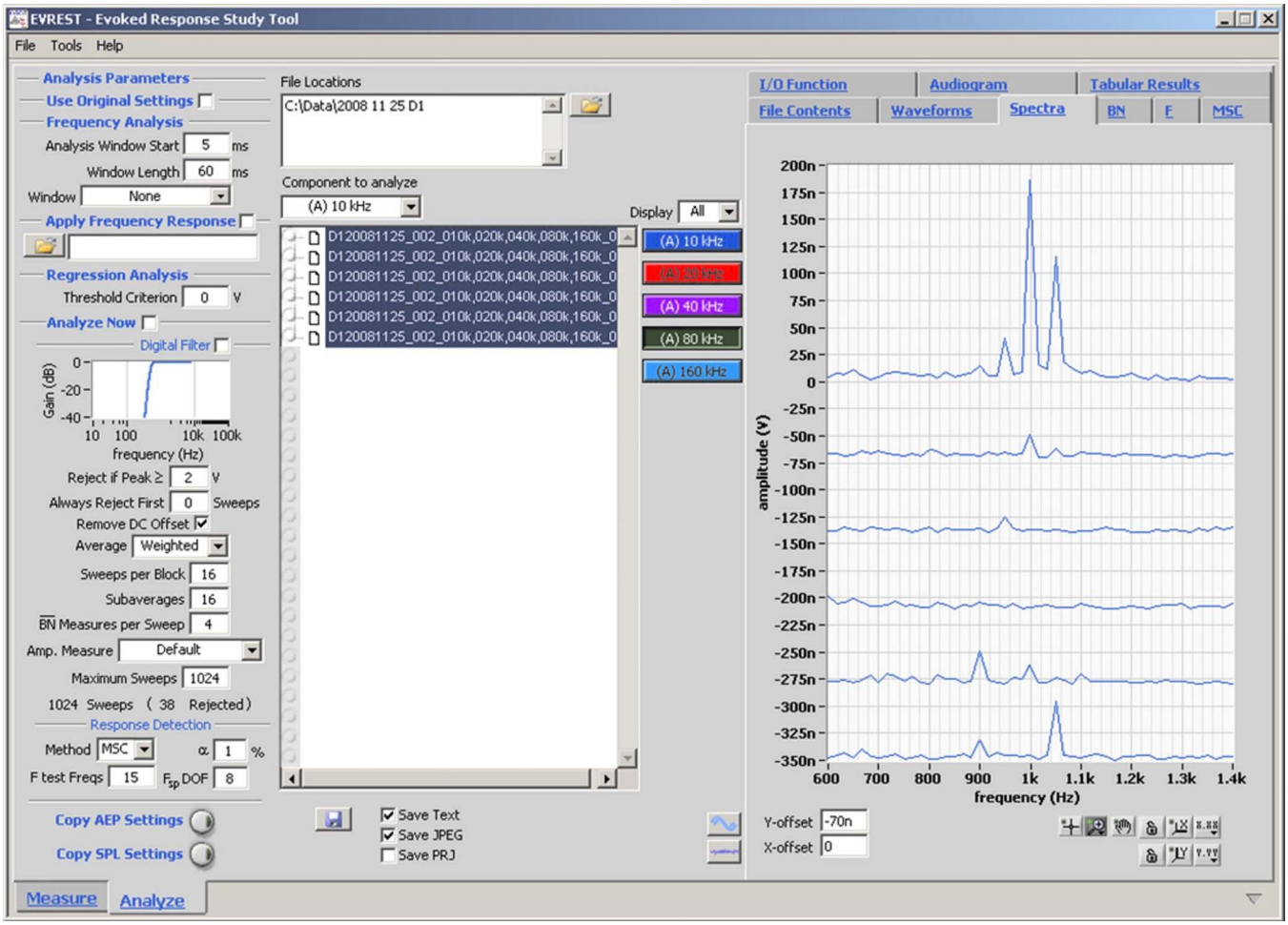


FIG. 10. (Color online) The EVREST software Analyze screen, used for viewing, analyzing, and archiving previously acquired data.

rejection parameters, a change from normal to weighted averaging, and the use of alternate response detection methods).

6. Automating measurements

Measurements can be automated in two ways. The first approach allows the user to define, in tabular format, a number of measurements to be conducted in sequence. Each row of the table specifies a sequence of measurements. The columns in the table indicate the stimulus carrier frequencies, AM frequencies, starting stimulus levels, stimulus amplitude step sizes (the change in amplitude from one measurement to the next), and the total number of measurements. This approach allows the automatic measurement of I/O functions or modulation rate transfer functions. The second automation approach is an adaptive staircase AEP threshold technique, where the stimulus SPL is automatically adjusted from one measurement to the next based on whether an AEP was detected during the previous trial. If a response was detected, the stimulus SPL is reduced. If the response was not detected, the stimulus SPL is increased. At the transitions from detections to non-detections (or vice-versa), the amount (in decibels) the SPL is increased or decreased (the step size) is reduced by user-defined ratios. The measurement is complete when the step size is reduced below a specified value. The threshold is defined as the mean of the SPLs corresponding to the lowest detection and the next highest non-detection. At the conclusion of either type of automated AEP measurement, a screen displays the measured I/O functions for each waveform component. Within this screen, the user can use the mouse to interactively choose additional SPLs to test for each component. The adaptive threshold process, coupled with the use of multiple, simultaneous stimuli, has shown great promise for rapid hearing assessment in dolphins, with auditory testing at nine frequencies from 10 to 160 kHz occurring in as little as 5–9 min (Finneran *et al.*, 2008).

B. Analyze screen

1. Viewing and analyzing data

Figure 10 shows the Analyze screen, which is used to view and/or analyze previously acquired data. The left panel specifies the DAQ/analysis parameters, which are similar to those found on the AEP and SPL tabs, e.g., the artifact reject level, time interval for spectral analysis, tapered window style, and averaging method. The central area contains a list of data files within the currently selected folder. The right side of the screen contains a tab structure with graphs and tabular displays.

The operations on the Analyze screen operate in parallel to those on the Measure screen, so previously acquired data

can be viewed or analyzed while new data are simultaneously acquired. If the data are viewed, the results obtained during the measurement process are displayed. If the data are analyzed, the binary data are read from the disk one sweep at a time and processed through the same analysis sequence that occurred during the measurement. During this "reanalysis," certain parameters may be altered from those values used during the original measurement. The parameters that may be changed include the reject level, averaging method, maximum number of sweeps, time period for spectral analysis, ORD settings, and the amplitude metric. Zero-phase digital filtering may also be performed, if desired.

The results of viewing/analyzing the selected data files are shown on the graphs and tables within the tab structure on the right side of the Analyze screen. Most of the tabs function analogously to the similarly named tabs on the Measure screen (e.g., BN, F, MSC, and Audiogram). An exception is the Tabular Results tab, which contains a table showing the stimulus parameters and measurement results for each waveform component within each file. The Waveforms and Spectra tabs also operate slightly different than those on the Measure screen. Since multiple data files can be selected, the graphs on the Waveforms and Spectra tabs can display multiple waveforms/spectra in waterfall-style plots with adjustable vertical spacing. Waveform and spectral displays also include options to show replicates, where the total number of sweeps is divided in half and two averages are displayed, and the "±average" (Schimmel, 1967), where half of the sweeps are reversed in polarity during averaging (to estimate residual background noise). The Spectra tab in Fig. 10 shows the spectral amplitude from six consecutive AEP measurements in the dolphin subject; each measurement featured a five-component stimulus. The stimulus carrier frequencies were 10, 20, 40, 80, and 160 kHz, with respective modulation rates of 900, 950, 1000, 1050, and 1100 Hz. The SPL of each component was adjusted using the adaptive threshold procedure, so, for example, after the 40 kHz response (at 1000 Hz) was detected during the first (top) measurement, the SPL at 40 kHz was reduced, resulting in a much smaller response at 1 kHz in the spectrum displayed on the next line.

The I/O Function tab displays the signal amplitude as a function of the stimulus level for each waveform component. For AEP measurements with ORD metrics, thresholds are automatically calculated and displayed on the Audiogram tab. If desired, thresholds can also be calculated by performing a linear regression on the I/O function and interpolating/extrapolating to a user-defined threshold AEP voltage. The specific I/O function data points for regression analysis are manually selected by clicking on the points displayed in the I/O function.

2. Saving results

Saving text results from the Analyze screen automatically creates three types of data files: (1) a file containing the time and amplitude values for the grand average waveforms for each selected file, (2) a file containing the frequency, spectral amplitude, and phase angle values for each selected file, and (3) a file containing the text from the Tabular Results tab, which contains the filenames, stimulus parameters, amplitude measurements, and threshold estimates for the selected data files. Bit-mapped images (JPEG format) of the waveforms, spectra, and I/O functions may also be created if desired.

V. CONCLUSIONS

The EVREST system is designed for calibrating sound stimuli and recording and analyzing transient and steady-state AEPs. The EVREST hardware is portable, rugged, battery-powered, and features a bandwidth that encompasses the audible range of echolocating odontocetes, making the system particularly appropriate for field testing of stranded or rehabilitating cetaceans. The EVREST software is available at no charge via collaborative arrangements with the author and agreement that (1) copies will not be distributed to others, (2) it will not be used for profit, and (3) any presentations or publications of findings using it will cite the U.S. Navy Marine Mammal Program as its source. For additional information, please contact the author.

ACKNOWLEDGMENTS

The author would like to thank Dorian Houser for many helpful discussions on the system operating concepts and for helping to test the hardware and software. Carolyn Melka, Randall Dear, and Brian Branstetter also performed a great deal of hardware/software testing and provided helpful suggestions. Financial support was provided by the U.S. Office of Naval Research.

Carder, D. A., and Ridgway, S. H. (**1994**). "A portable system for physiological assessment of hearing in marine animals," J. Acoust. Soc. Am. **96**, 3316.

Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., Houser, D., and Mann, D. A. (**2006**). "Beaked whale auditory evoked potential hearing measurements," J. Comp. Physiol. [A] **192**, 489–495.

Delory, E., d. Rio, J., Castell, J., v. d. Schaar, M., and André, M. (**2007**). "OdiSEA: An autonomous portable auditory screening unit for rapid assessment of hearing in cetaceans," Aquat. Mamm. **33**, 85–92.

Dobie, R. A., and Wilson, A. J. (**1995**). "Objective versus human observer detection of 40-Hz auditory-evoked potentials," J. Acoust. Soc. Am. **97**, 3042–3050.

Dobie, R. A., and Wilson, M. J. (**1989**). "Analysis of auditory evoked potentials by magnitude-squared coherence," Ear Hear. **10**, 2–13.

Dobie, R. A., and Wilson, M. J. (**1993**). "Objective response detection in the frequency domain," Electroencephalogr. Clin. Neurophysiol. **88**, 516–524.

Dobie, R. A., and Wilson, M. J. (**1994**). "Phase weighting: A method to improve objective detection of steady-state evoked potentials," Hear. Res. **79**, 94–98.

Dobie, R. A., and Wilson, M. J. (**1996**). "A comparison of *t* test, *F* test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am. **100**, 2236–2246.

Dolphin, W. F. (**1996**). "Auditory evoked responses to amplitude modulated stimuli consisting of multiple envelope components," J. Comp. Physiol. [A] **179**, 113–121.

Dolphin, W. F. (**1997**). "Electrophysiological measures of auditory processing in odontocetes," Bioacoustics **8**, 79–101.

Dolphin, W. F. (**2000**). "Electrophysiological measures of auditory processing in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 294–329.

Don, M., and Elberling, C. (**1994**). "Evaluating residual background noise in human auditory brain-stem responses," J. Acoust. Soc. Am. **96**, 2746–2757.

Elberling, C., and Don, M. (**1984**). "Quality estimation of averaged auditory

brainstem responses," Scand. Audiol. **13**, 187–197.
Elberling, C., and Wahlgreen, O. (**1985**). "Estimation of auditory brainstem response, ABR, by means of Bayesian inference," Scand. Audiol. **14**, 89–96.
Finneran, J. J. (**2008**). "Evoked response study tool (EVREST) user's guide," SSC San Diego Technical Document 3226, SSC San Diego, San Diego, CA.
Finneran, J. J., and Houser, D. S. (**2004**). "A portable system for marine mammal auditory-evoked potential measurements," J. Acoust. Soc. Am. **115**, 2517.
Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **119**, 3181–3192.
Finneran, J. J., and Houser, D. S. (**2007**). "Bottlenose dolphin (*Tursiops truncatus*) steady-state evoked responses to multiple simultaneous sinusoidal amplitude modulated tones," J. Acoust. Soc. Am. **121**, 1775–1782.
Finneran, J. J., Houser, D. S., Blasko, D., Hicks, C., Hudson, J., and Osborn, M. (**2008**). "Estimating bottlenose dolphin (*Tursiops truncatus*) hearing thresholds from single and multiple simultaneous auditory evoked potentials," J. Acoust. Soc. Am. **123**, 542–551.
Finneran, J. J., Schlundt, C. E., Branstetter, B., and Dear, R. L. (**2007**). "Assessing temporary threshold shift in a bottlenose dolphin (*Tursiops truncatus*) using multiple simultaneous auditory evoked potentials," J. Acoust. Soc. Am. **122**, 1249–1264.
Helweg, D. A., Carder, D. A., and Ridgway, S. H. (**1997**). "A portable virtual instrument for collection of cetacean auditory evoked potentials," J. Acoust. Soc. Am. **102**, 3196.
Houser, D. S., and Finneran, J. J. (**2006a**). "A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods," J. Acoust. Soc. Am. **120**, 1713–1722.
Houser, D. S., and Finneran, J. J. (**2006b**). "Variation in the hearing sensitivity of a dolphin population obtained through the use of evoked potential audiometry," J. Acoust. Soc. Am. **120**, 4090–4099.
Houser, D. S., Gomez-Rubio, A., and Finneran, J. J. (**2008**). "Evoked potential audiometry of 13 Pacific bottlenose dolphins (*Tursiops truncatus gilli*)," Marine Mammal Sci. **24**, 28–41.
John, M. S., Lins, O. G., Boucher, B. L., and Picton, T. W. (**1998**). "Multiple auditory steady-state responses (MASTER): Stimulus and recording parameters," Audiology **37**, 59–82.
Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., and Durieux-Smith, A. (**1995**). "Auditory steady-state responses to tones amplitude-modulated at 80–110 Hz," J. Acoust. Soc. Am. **97**, 3051–3063.
Lins, O. G., and Picton, T. W. (**1995**). "Auditory steady-state responses to multiple simultaneous stimuli," Electroencephalogr. Clin. Neurophysiol. **96**, 420–432.
Mason, S. M., Su, A. P., and Hayes, R. A. (**1977**). "Simple online detector of auditory evoked cortical potentials," Med. Biol. Eng. Comput. **15**, 641–647.
Nachtigall, P. E., Mooney, T. A., Taylor, K. A., Miller, L. A., Rasmussen, M. H., Akamatsu, T., Teilmann, J., Linnenschmidt, M., and Vikingsson, G. A. (**2008**). "Shipboard measurements of the hearing of the white-beaked dolphin *Lagenorhynchus albirostris*," J. Exp. Biol. **211**, 642–647.
Nachtigall, P. E., Supin, A. Y., Amundin, M., Roken, B., Møller, T., Mooney, T. A., Taylor, K. A., and Yuen, M. (**2007**). "Polar bear *Ursus maritimus* hearing measured with auditory evoked potentials," J. Exp. Biol. **210**, 1116–1122.
Nachtigall, P. E., Yuen, M. M. L., Mooney, T. A., and Taylor, K. A. (**2005**). "Hearing measurements from a stranded infant Risso's dolphin, *Grampus griseus*," J. Exp. Biol. **208**, 4181–4188.
National Instruments Corporation (**2008**). LABVIEW Version 8.6, Austin, TX.
Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., and Maiste, A. C. (**1987**). "Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone," J. Acoust. Soc. Am. **82**, 165–178.
Popov, V. V., Supin, A. Y., and Klishin, V. O. (**1997**). "Paradoxical lateral suppression in the dolphin's auditory system: Weak sounds suppress response to strong sounds," Neurosci. Lett. **234**, 51–54.
Popov, V. V., Supin, A. Y., and Klishin, V. O. (**1998**). "Lateral suppression of rhythmic evoked responses in the dolphin's auditory system," Hear. Res. **126**, 126–134.
Popov, V. V., Supin, A. Y., Pletenko, M. G., Tarakanov, M. B., Klishin, V. O., Bulgakova, T. N., and Rosanova, E. I. (**2007**). "Audiogram variability in normal bottlenose dolphins (*Tursiops truncatus*)," Aquat. Mamm. **33**, 24–33.
Popov, V. V., Supin, A. Y., Wang, D., Wank, K., Xiao, J., and Li, S. (**2005**). "Evoked-potential audiogram of the Yangtze finless porpoise *Neophocaena phocaenoides asiaeorientalis* (L)," J. Acoust. Soc. Am. **117**, 2728–2731.
Regan, D., and Regan, M. P. (**1988**). "The transducer characteristic of hair cells in the human ear: A possible objective measure," Brain Res. **438**, 363–365.
Ridgway, S. H., and Carder, D. A. (**1983**). "Audiograms for large cetaceans: A proposed method for field studies," J. Acoust. Soc. Am. **74**, S53.
Schimmel, H. (**1967**). "The (±) reference: Accuracy of estimated mean components in average response studies," Science **157**, 92–94.
Schimmel, H., Rapin, I., and Cohen, M. M. (**1974**). "Improving evoked response audiometry with special reference to the use of machine scoring," Audiology **13**, 33–65.
Stürzebecher, E., Cebulla, M., and Wernecke, K. D. (**2001**). "Objective detection of transiently evoked otoacoustic emissions," Scand. Audiol. **30**, 78–88.
Supin, A. Y., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic, Boston, MA).
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.
Taylor, K. A., Nachtigall, P. E., Mooney, T. A., Supin, A. Y., and Yuen, M. M. L. (**2007**). "A portable system for the evaluation of the auditory capabilities of marine mammals," Aquat. Mamm. **33**, 93–99.
Taylor, K. A., Nachtigall, P. E., Mooney, T. A., Yuen, M. M., and Supin, A. Y. (**2006**). "Evaluation of the auditory capabilities of marine mammals using a portable auditory-evoked potential system," J. Acoust. Soc. Am. **120**, 3326.
Weber, B. A., and Fletcher, G. L. (**1980**). "A computerized scoring procedure for auditory brainstem response audiometry," Ear Hear. **1**, 233–236.

Erratum: "Reliability of estimating the room volume from a single room impulse response" [J. Acoust. Soc. Am. 124, 982–993 (2008)]

Martin Kuster
Laboratory of Acoustic Imaging and Sound Control, Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 10 April 2008; revised 15 April 2009; accepted 15 April 2009)
[DOI: 10.1121/1.3132503]

PACS number(s): 43.55.Gx, 43.60.Lq, 43.55.Br, 43.10.Vx

A typographical error appears in Eq. (16) and Eq. (17). The correct expressions read

$$\overline{p_r^2} = \rho_0 c W\left(\frac{cT_{60}}{6\ln(10)V}\right)[e^{-6\ln(10)r_0/cT_{60}}] \tag{16}$$

and

$$V_{\text{revised}} = \frac{\overline{p_0^2}(r_0)}{\overline{p_r^2}}\frac{4\pi r_0^2 cT_{60}}{6\ln(10)}[e^{-6\ln(10)r_0/cT_{60}}], \tag{17}$$

respectively. The correct equations were used for the numerical results.

 0001-4966/2009/126(1)/501/1/$25.00

ACOUSTICAL NEWS

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 and notices are 2 months prior to publication*

New Fellow of the Acoustical Society of America

Christine H. Shadle—For contributions to the aeroacoustics of speech

Advanced Degree Dissertation in Acoustics

Editor's Note: Abstracts of Doctoral and Master's theses will be welcomed at all times. Please note that they must be limited to 200 words, must include the appropriate PACS classification numbers, and formatted as shown below. If sent by postal mail, note that they must be double spaced. The address for obtaining a copy of the thesis is helpful. Submit abstracts to: Acoustical Society of America, Thesis Abstracts, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, e-mail: asa@aip.org.

Design and evaluation of digital signal processing algorithms for acoustic feedback and echo cancellation [43.38.Tj, 43.60.Dh]—Toon van Waterschoot, *Faculty of Engineering, Katholieke Universiteit Leuven, Leuven, Belgium, March 2009 (Ph.D.).* This thesis deals with several open problems in acoustic echo cancellation and acoustic feedback control. Our main goal has been to develop solutions that provide a high performance and sound quality, and behave in a robust way in realistic conditions. This can be achieved by departing from the traditional ad-hoc methods, and instead deriving theoretically well-founded solutions, based on results from parameter estimation and system identification. In the development of these solutions, the computational efficiency has permanently been taken into account as a design constraint, in that the complexity increase compared to the state-of-the-art solutions should not exceed 50% of the original complexity. [http://hdl.handle.net/1979/2599]

Advisor: Marc Moonen

Calendar of Meetings and Congresses

2009

19–23 July	Leuven, Belgium. **15th International on Photoacoustics and Photothermal Phenomena.** Web: http://www.icppp15.be
12–16 August	Jyväskylä, Finland. **7th Triennual Conference of the European Society for Cognitive Science of Music (ESCOM 2009).** Web: http://www.fyu.fi/hum/laitokset/musikki/en/escom2009
23–28 August	Ottawa, Ont. Canada. **Inter-noise 2009.** Web: http://www.internoise2009.com
23–27 August	Seattle, Washington, USA. **11th International Conference on Music Perception and Cognition.** Web: TBA
6–10 September	Brighton, UK. **InterSpeech 2009 Conference.** Web: http://www.interspeech2009.org
7–11 September	Dresden, Germany. **9th International Conference on Theoretical and Computational Acoustics.** Web: http://ictca2009.com
14–18 September	Kyoto, Japan. **5th Animal Sonar Symposium.** Web: http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15–17 September	Koriyama, Japan. **Autumn Meeting of the Acoustical Society of Japan.** Web: http://www.asj.gr.jp/index-en.html
19–23 September	Rome, Italy. **IEEE 2009 Ultrasonics Symposium.** E-mail: pappalar@uniroma3.it
21–23 September	Beijing, China. **Western Pacific Acoustics Conference (WESPAC).** Web: http://www.wespacx.org
23–25 September	Xi'an, China. **Pacific Rim Underwater Acoustics Conference (PRUAC).** E-mail: lfh@mail.ioa.ac.cn
23–25 September	Cádiz, Spain. **TECNIACUSTICA'09.** Web: http://www.-sea-acustica.es
5–7 October	Tallinn, Estonia. **International Conference on Complexity of Nonlinear Waves.** Web: http://www.ioc.ee/cnw09
18–21 October	New Paltz, NY, USA. **IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009).**
26–28 October	Edinburgh, UK. **Euronoise 2009.** Web: http://www.euronoise2009.org.uk
26–30 October	San Antonio, TX, USA. **158th Meeting of the Acoustical Society of America.** Web: http://asa.aip.org/meetings.html
5–6 November	Dübendorf, Switzerland. **Swiss Acoustical Society Autumn Meeting.** Web: http://www.sga-ssa.ch
23–25 November	Adelaide, Australia. **Australian Acoustics Society National Conference.** Web: http://www.acoustics.asn.au/joomla

2010

15–19	March Dallas, TX, USA. **International Conference on Acoustics, Speech, and Signal Processing.** Web: http://icassp2010.org
19–23 April	Baltimore, MD, USA. Joint Meeting: **158th Meeting of the Acoustical Society of America and Noise Con 2010.** http://asa.aip.org/meetings.html
09–11 June	Aalborg, Denmark. **14th Conference on Low Frequency Noise and Vibration.** Web: http://lowfrequency2010.org
13–16 June	Lisbon, Portugal. **INTERNOISE2010.** Web: http://www.internoise2010.org

 0001-4966/2009/126(1)/502/2/$25.00

23–27 August	Sydney, Australia. **International Congress on Acoustics 2010.** Web: http://www.ica2010sydney.org
15–18 September	Ljubljania, Slovenia. **Alp-Adria-Acoustics Meeting joint with EAA.** E-mail: mirko.cudina@fs.uni-lj.si
26–30 September	Makuhari, Japan. **Interspeech 2010–ICSLP.** Web: http://www.interspeech2010.org
11–14 October	San Diego, California, USA. **IEEE 2010 Ultrasonics Symposium.** E-mail: bpotter@vectron.com
15–19 November	Cancun, Mexico. **2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico.** Web: http://asa.aip.org/meetings.html

2011

27 June–1 July	Aalborg, Denmark. **Forum Acusticum 2011.** Web: http://www.fa2011.org
27–31 August	Florence, Italy. **Interspeech 2011.** Web: http://www.interspeech2011.org
4–7 September	Gdansk, Poland. **International Congress on Ultrasonics.** Web: TBA
4–7 September	Osaka, Japan. **Internoise 2011.** Web: TBA

2013

2–7 June	Montréal, Canada. **21st International Congress on Acoustics (ICA 2013).** Web: http://www.ica2013montreal.org

ACOUSTICAL STANDARDS NEWS

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

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: Schomer@SchomerAndAssociates.com]

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please contact Susan B. Blaeser.
Comments are welcomed on all material in Acoustical Standards News.
This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org.

Standards Meetings Calendar—National

The ASA Committee on Standards (ASACOS) will meet in San Antonio, Texas, during the week of 26–30 October 2009. The specific meeting date and time will be announced soon.

Standards Meetings Calendar—International

1–4 September 2009—Las Vegas, Nevada

• TC/108/SC4, Human exposure to mechanical vibration and shock

9–13 November 2009—Tokyo

• IEC/TC29, Electroacoustics

16–20 November 2009—Seoul

• ISO/TC43, Acoustics

• ISO/TC43/SC1, Noise

• ISO/TC43/SC2, Building acoustics

Recently Published Standards

At the meeting in Miami, Bob Hellweg, Vice Chair of ASACOS, presented a plaque to **Ali T. Herfat** recognizing his accomplishments as Project Leader for the revision of ANSI/ASA S2.9-2008 *American National Standard Parameters for Specifying Damping Properties of Materials and System Damping* (see Fig. 1).

Elliott Berger, Chair of S12/WG 11 "Hearing Protector Attenuation and Performance," who was not present at the Miami meeting, was recognized for the completion of ANSI/ASA S12.6-2008 *American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors*.

At the Portland meeting Craig Champlin, Chair of Accredited Standards Committee S3, Bioacoustics, presented a plaque to **Conrad Wall, III**, with appreciation for his efforts as Chairman of S3/WG 82, Basic Vestibular Function Test Battery (see Fig. 2), which published **ANSI/ASA S3.45-2009** *American National Standard Procedures for Testing Basic Vestibular Function*.

S12 also recognized **Steve Antonides**, who was not present at the Portland meeting. Mr. Antonides is Chair of S12/WG 40, "Measurement of Noise Aboard Ships," which completed ANSI/ASA S12.67-2008 *American National Standard Pre-installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment*.

Also at Portland, Accredited Standards Committee S2, Mechanical Vibration and Shock, commended **Maryon "Skip" Williams** who served as project leader for the revision on ANSI/ASA S2.28-2009, *American National Standard Guide for the Measurement and Evaluation of Broadband Vibration of Surface Ship Auxiliary Rotating Machinery*. Mr. Williams was not present to receive his plaque.

FIG. 1. (Color online) R. D. Hellweg (right) presents a plaque to Ali T. Herfat.

FIG. 2. (Color online) Craig Champlin (right) presents a plaque to Conrad Wall, III.

 0001-4966/2009/126(1)/504/10/$25.00

Accredited Standards Committee on Acoustics, S1

(P. Battenberg, Chair; R.J. Peppin, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

S1 WORKING GROUPS

S1/Advisory—Advisory Planning Committee to S1 (P. Battenberg)

S1/WG01—Standard Microphones and their Calibration (V. Nedzelnitsky)

S1/WG04—Measurement of Sound Pressure Levels in Air (VACANT, Chair; E. Dunens, Vice Chair)

S1/WG05—Band Filter Sets (A.H. Marsh)

S1/WG09—Calibration of Underwater Electroacoustic Transducers (R.M. Drake)

S1/WG17—Sound Level Meters and Integrating Sound Level Meters (G.R. Stephany)

S1/WG19—Insertion Loss of Windscreens (A.J. Campanella)

S1/WG20—Ground Impedance (Measurement of Ground Impedance and Attenuation of Sound due to the Ground) (K. Attenborough, Chair; J. Sabatier, Vice Chair)

S1/WG22—Bubble Detection and Cavitation Monitoring (VACANT)

S1/WG26—High Frequency Calibration of the Pressure Sensitivity of Microphones (VACANT)

S1/WG27—Acoustical Terminology (J.S. Vipperman)

S1 STANDARDS ON ACOUSTICS

ANSI S1.1-1994 (R 2004) American National Standard Acoustical Terminology.

ANSI S1.4-1983 (R 2006) American National Standard Specification for Sound Level Meters. This Standard includes ANSI S1.4A-1985 (R 2006) Amendment to ANSI S1.4-1983.

ANSI S1.6-1984 (R 2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements.

ANSI S1.8-1989 (R 2006) American National Standard Reference Quantities for Acoustical Levels.

ANSI S1.9-1996 (R 2006) American National Standard Instruments for the Measurement of Sound Intensity.

ANSI S1.11-2004 American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters.

ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air.

ANSI/ASA S1.14-1998 (R 2008) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz.

ANSI S1.15/Part 1-1997 (R 2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones.

ANSI S1.15/Part 2-2005 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique.

ANSI S1.16-2000 (R 2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones.

ANSI S1.17/Part 1-2004 American National Standard Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air.

ANSI S1.18-1999 (R 2004) American National Standard Template Method for Ground Impedance.

ANSI S1.20-1988 (R 2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers.

ANSI S1.22-1992 (R 2007) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics.

ANSI S1.24 TR-2002 (R 2007) ANSI Technical Report Bubble Detection and Cavitation Monitoring.

ANSI S1.25-1991 (R 2007) American National Standard Specification for Personal Noise Dosimeters.

ANSI S1.26-1995 (R 2004) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere.

ANSI S1.40-2006 American National Standard Specifications and Verification Procedures for Sound Calibrators.

ANSI S1.42-2001 (R 2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements.

ANSI S1.43-1997 (R 2007) American National Standard Specifications for Integrating-Averaging Sound Level Meters.

Accredited Standards Committee on Mechanical Vibration and Shock, S2

(A.T. Herfat, Chair; C.F. Gaumond, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

S2 WORKING GROUPS

S2/WG01—S2 Advisory Planning Committee (A.T. Herfat, Chair; C.F. Gaumond, Vice Chair)

S2/WG02—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D.J. Evans)

S2/WG03—Signal Processing Methods (T.S. Edwards)

S2/WG04—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W. Madigosky, Chair, J. Niemiec, Vice Chair)

S2/WG05—Use and Calibration of Vibration and Shock Measuring Instruments (D.J. Evans, Chair; B.E. Douglas, Vice Chair)

S2/WG06—Vibration and Shock Actuators (G.B. Booth)

S2/WG07—Acquisition of Mechanical Vibration and Shock Measurement Data (B.E. Douglas)

S2/WG08—Analysis Methods of Structural Dynamics (M. Mezache)

S2/WG09—Training and Accreditation (R.L. Eshleman, Chair; D. Corelli, Vice Chair)

S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R.L. Eshleman, Chair; H.C. Pusey, Vice Chair)

S2/WG10/Panel 1—Balancing (R.L. Eshleman)

S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation (R. Bankert)

S2/WG10/Panel 3—Machinery Testing (R.L. Eshleman)

S2/WG10/Panel 4—Prognosis (A.J. Hess)

S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever)

S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (VACANT)

S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (M. Mezache)

S2/WG13—Shock Test Requirements for Shelf-mounted and Other Commercial Electronics Systems (B. Lang)

S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock (D.D. Reynolds, Chair; R. Dong, Vice Chair)

S2 STANDARDS ON MECHANICAL VIBRATION AND SHOCK

ANSI S2.1-2000 / ISO 2041:1990 American National Standard Vibration and Shock—Vocabulary (a Nationally Adopted International Standard).

ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups.

ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements.

ANSI S2.8-2007 American National Standard Technical Information Used for Resilient Mounting Applications.

ANSI/ASA S2.9-2008 American National Standard Parameters for Specifying Damping Properties of Materials and System Damping.

ANSI S2.16-1997 (R 2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment.

ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications.

ANSI S2.20-1983 (R 2006) American National Standard Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects.

ANSI S2.21-1998 (R 2007) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements.

ANSI S2.22-1998 (R 2007) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials.

ANSI S2.23-1998 (R 2007) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials.

ANSI S2.24-2001 (R 2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials.

ANSI S2.25-2004 American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships.

ANSI S2.26-2001 (R 2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment.

ANSI S2.27-2002 (R 2007) American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery.

ANSI/ASA S2.28-2009 American National Standard Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery.

ANSI/ASA S2.29-2003 (R 2008) American National Standard Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery.

ANSI S2.31-1979 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers.

ANSI S2.32-1982 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation.

ANSI S2.34-1984 (R 2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix.

ANSI S2.42-1982 (R 2004) American National Standard Procedures for Balancing Flexible Rotors.

ANSI S2.43-1984 (R 2005) American National Standard Criteria for Evaluating Flexible Rotor Balance.

ANSI S2.46-1989 (R 2005) American National Standard Characteristics to be Specified for Seismic Transducers.

ANSI S2.48-1993 (R 2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics.

ANSI S2.60-1987 (R 2005) American National Standard Balancing Machines—Enclosures and Other Safety Measures.

ANSI S2.61-1989 (R 2005) American National Standard Guide to the Mechanical Mounting of Accelerometers.

ANSI S2.70-2006 American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand (Revision of ANSI S3.34-1986).

ANSI S2.71-1983 (R 2006) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings (Reaffirmation and redesignation of ANSI S3.29-1983).

ANSI S2.72/Part 1-2002 (R 2007) / ISO 2631-1:1997 (Redesignation of ANSI S3.18/Part 1-2002/ISO 2631-1:1997) American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 1: General requirements (a Nationally Adopted International Standard).

ANSI S2.72/Part 4-2003 (R 2007)/ISO 2631-4:2001 (Redesignation of ANSI S3.18/Part 4-2003/ISO 2631-4:2001) American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (a Nationally Adopted International Standard).

ANSI S2.73-2002 (R 2007)/ISO 10819:1996 (Redesignation of ANSI S3.40-2002 / ISO 10819:1996) American National Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (a Nationally Adopted International Standard).

Accredited Standards Committee on Bioacoustics, S3

(C.A. Champlin, Chair; D.A. Preves, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance, and comfort.

S3 WORKING GROUPS

S3/Advisory—Advisory Planning Committee to S3 (C. Champlin, Chair; D.A. Preves, Vice Chair)

S3/WG35—Audiometric Equipment (R.L. Grason)

S3/WG36—Speech Intelligibility (R.S. Schlauch)

S3/WG37—Coupler Calibration of Earphones (C.J. Struck)

S3/WG39—Human Exposure to Mechanical Vibration and Shock (parallel to ISO/TC 108/SC 4) (D.D. Reynolds, Chair; R. Dong, Vice Chair)

S3/WG43—Method for Calibration of Bone Conduction Vibrators (J.D. Durrant)

S3/WG48—Hearing Aids (D.A. Preves)

S3/WG51—Auditory Magnitudes (R.P. Hellman)

S3/WG56—Criteria for Background Noise for Audiometric Testing (J. Franks)

S3/WG59—Measurement of Speech Levels (M.C. Killion and L.A. Wilber, Co-Chairs)

S3/WG62—Impulse Noise with Respect to Hearing Hazard (G.R. Price)

S3/WG67—Manikins (M.D. Burkhard)

S3/WG72—Measurement of Auditory Evoked Potentials (R.F. Burkard)

S3/WG76—Computerized Audiometry (A.J. Miltich)

S3/WG79—Methods for Calculation of the Speech Intelligibility Index (C.V. Pavlovic)

S3/WG80—Probe-tube Measurements of Hearing Aid Performance (W.A. Cole)

S3/WG81—Hearing Assistance Technologies (L. Thibodeau and L.A. Wilber, Co-Chairs)

S3/WG82—Basic Vestibular Function Test Battery (C. Wall)

S3/WG83—Sound Field Audiometry (T.R. Letowski)

S3/WG84—Otoacoustic Emissions (VACANT)

S3/WG88—Standard Audible Emergency Evacuation and Other Signals (R. Boyer)

S3/WG89—Spatial Audiometry in Real and Virtual Environments (J. Besing)

S3/WG91—Text-to-Speech Synthesis Systems (C. Bickley and A. Syrdal, Co-Chairs)

S3 Liaison Group

S3/L1 S3 U. S. TAG Liaison to IEC/TC 87 Ultrasonics (W.L. Nyborg)

S3 STANDARDS ON BIOACOUSTICS

ANSI/ASA S3.1-1999 (R 2008) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms.

ANSI S3.2-2009 American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems.

ANSI S3.4-2007 American National Standard Procedure for the Computation of Loudness of Steady Sounds.

ANSI S3.5-1997 (R 2007) American National Standard Methods for Calculation of the Speech Intelligibility Index.

ANSI S3.6-2004 American National Standard Specification for Audiometers.

ANSI/ASA S3.7-1995 (R 2008) American National Standard Method for Coupler Calibration of Earphones.

ANSI S3.13-1987 (R 2007) American National Standard Mechanical Coupler for Measurement of Bone Vibrators.

ANSI/ASA S3.20-1995 (R 2008) American National Standard Bioacoustical Terminology.

ANSI S3.21-2004 American National Standard Methods for Manual Pure-Tone Threshold Audiometry.

ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics.

ANSI S3.25-1989 (R 2003) American National Standard for an Occluded Ear Simulator.

ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions.

ANSI S3.36-1985 (R 2006) American National Standard Specification for a Manikin for Simulated in situ Airborne Acoustic Measurements.

ANSI S3.37-1987 (R 2007) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids.

ANSI S3.39-1987 (R 2007) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance).

ANSI S3.41-1990 (R 2001) American National Standard Audible Emergency Evacuation Signal.

ANSI S3.42-1992 (R 2007) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal.

ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment.

ANSI/ASA S3.45-2009 American National Standard Procedures for Testing Basic Vestibular Function.

ANSI S3.46-1997 (R 2007) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids.

Animal Bioacoustics Subcommittee, S3/SC 1

(D.K. Delaney, Chair; M.C. Hastings, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations. Animals to be covered may potentially include commercially grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals.

S3/SC 1 WORKING GROUPS

S3/SC 1/WG01—Animal Bioacoustics Terminology (A.E. Bowles)

S3/SC 1/WG02—Effects of Sound on Fish and Turtles (R.R. Fay and A.N. Popper, Co-Chairs)

S3/SC 1/WG03—Underwater Passive Acoustic Monitoring for Bioacoustic Applications (A.M. Thode)

S3/SC 1/WG04—Description and Measurement of the Ambient Sound in Parks, Wilderness Areas, and Other Quiet and/or Pristine Areas (K. Fristrup and G.R. Stanley, Co-Chairs)

Accredited Standards Committee on Noise, S12

(W.J. Murphy, Chair; R.D. Hellweg, Vice Chair)

Scope: Standards, specifications, and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation, and control, including biological safety, tolerance and comfort, and physical acoustics as related to environmental and occupational noise.

S12 WORKING GROUPS

S12/Advisory—Advisory Planning Committee to S12 (W.J. Murphy)

S12/WG03—Measurement of Noise from Information Technology and Telecommunications Equipment (K.X.C. Man)

S12/WG11—Hearing Protector Attenuation and Performance (E.H. Berger)

S12/WG13—Method for the Selection of Hearing Protectors that Optimize the Ability to Communicate (D. Byrne)

S12/WG14—Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (W.J. Murphy)

S12/WG15—Measurement and Evaluation of Outdoor Community Noise (P.D. Schomer)

S12/WG23—Determination of Sound Power (B.M. Brooks and J. Schmitt, Co-Chairs)

S12/WG31—Predicting Sound Pressure Levels Outdoors (L. Pater)

S12/WG32—Revision of ANSI S12.7-1986 Methods for Measurement of Impulse Noise (W. Ahroon)

S12/WG36—Development of Methods for Using Sound Quality (P. Davies and G.L. Ebbit, Co-Chairs)

S12/WG38—Noise Labeling in Products (R.D. Hellweg)

S12/WG40—Measurement of the Noise Aboard Ships (S.P. Antonides, Chair; S.A. Fisher, Vice Chair)

S12/WG41—Model Community Noise Ordinances (L.S. Finegold, Chair; B.M. Brooks, Vice Chair)

S12/WG44—Speech Privacy (G.C. Tocci, Chair; D. Sykes, Vice Chair)

S12/WG45—Measurement of Occupational Noise Exposure from Telephone Equipment (K.A. Woo, Chair; L.A. Wilber, Vice Chair)

S12/WG46—Acoustical Performance Criteria for Relocatable Classrooms (T. Hardiman and P.D. Schomer, Co-Chairs)

S12/WG47—Underwater Noise Measurements of Ships (M. Bahtiarian, Chair; D.J. Vendittis, Vice Chair)

S12/WG48—Railroad Horn Sound Emission Testing (J. Erdreich, Chair; J.J. Earshen, Vice Chair)

S12/WG49—Noise from Hand-operated Power Tools, Excluding Pneumatic Tools (C. Hayden, Chair; B.M. Brooks, Vice Chair)

S12/WG50—Information Technology (IT) Equipment in Classrooms (R.D. Hellweg)

S12/WG51—Procedure for Measuring the Ambient Noise Level in a Room (J.G. Lilly)

S12/WG52—Revision of ANSI S12.60-2002 (P.D. Schomer)

S12 LIAISON GROUPS

S12/L1 IEEE 85 Committee for TAG Liaison—Noise Emitted by Rotating Electrical Machines (Parallel to ISO/TC 43/SC 1/WG 13) (R.G. Bartheld)

S12/L2 Measurement of Noise from Pneumatic Compressors, Tools and Machines (Parallel to ISO/TC 43/SC 1/WG 9) (VACANT)

S12/L3 SAE Committee for TAG Liaison on Measurement and Evaluation of Motor Vehicle Noise (parallel to ISO/TC 43/SC 1/WG 8) (R.F. Schumacher)

S12/L4 SAE Committee A-21 for TAG Liaison on Measurement and Evaluation of Aircraft Noise (J.D. Brooks)

S12/L5 ASTM E-33 on Environmental Acoustics (to include activities of ASTM E33.06 on Building Acoustics, parallel to ISO/TC 43/SC 2 and ASTM E33.09 on Community Noise) (K.P. Roy)

S12/L6 SAE Construction-Agricultural Sound Level Committee (I. Douell)

S12/L7 SAE Specialized Vehicle and Equipment Sound Level Committee (T.M. Disch)

S12/L8 ASME PTC 36 Measurement of Industrial Sound (R.A. Putnam, Chair; B.M. Brooks, Vice Chair)

S12 STANDARDS ON NOISE

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources.

ANSI/ASA S12.2-2008 American National Standard Criteria for Evaluating Room Noise.

ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment.

ANSI S12.5-2006/ISO 6926:1999 American National Standard Acoustics—Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (a Nationally Adopted International Standard).

ANSI/ASA S12.6-2008 American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors.

ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise.

ANSI/ASA S12.8-1998x (R 2008) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers.

ANSI S12.9/Part 1-1988 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1.

ANSI/ASA S12.9/Part 2-1992 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound.

ANSI/ASA S12.9/Part 3-1993 (R 2008) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present.

ANSI S12.9/Part 4-2005 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-Term Community Response.

ANSI/ASA S12.9/Part 5-2007 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use.

ANSI/ASA S12.9/Part 6-2008 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimation of Awakenings Associated with Outdoor Noise Events Heard in Homes.

ANSI/ASA S12.10-2002 (R 2007)/ISO 7779:1999 American National Standard Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment (a Nationally Adopted International Standard).

ANSI/ASA S12.11/Part 1-2003 (R 2008)/ISO 10302:1996 (MOD) American National Standard Acoustics—Measurement of noise and vibration of small air-moving devices—Part 1: Airborne noise emission (a Modified Nationally Adopted International Standard).

ANSI/ASA S12.11/Part 2-2003 (R 2008) American National Standard Acoustics—Measurement of Noise and Vibration of Small Air-Moving Devices—Part 2: Structure-Borne Vibration.

ANSI/ASA S12.12-1992 (R 2007) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity.

ANSI S12.13 TR-2002 ANSI Technical Report Evaluating the Effectiveness of Hearing Conservation Programs through Audiometric Data Base Analysis.

ANSI/ASA S12.14-1992 (R 2007) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors.

ANSI/ASA S12.15-1992 (R 2007) American National Standard For Acoustics—Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted.

ANSI/ASA S12.16-1992 (R 2007) American National Standard Guidelines for the Specification of Noise of New Machinery.

ANSI S12.17-1996 (R 2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment.

ANSI S12.18-1994 (R 2004) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level.

ANSI S12.19-1996 (R 2006) American National Standard Measurement of Occupational Noise Exposure.

ANSI S12.23-1989 (R 2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment.

ANSI S12.42-1995 (R 2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices.

ANSI/ASA S12.43-1997 (R 2007) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions.

ANSI/ASA S12.44-1997 (R 2007) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and Other Specified Positions from Sound Power Level.

ANSI/ASAS12.50-2002 (R 2007) / ISO 3740:2000 American National Standard Acoustics -Determination of sound power levels of noise sources—Guidelines for the use of basic standards (a Nationally Adopted International Standard).

ANSI/ASA S12.51-2002 (R 2007) / ISO 3741:1999 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms (a Nationally Adopted International Standard). This Standard includes Technical Corrigendum 1-2001. This standard replaces ANSI S12.31-1990 and ANSI S12.32-1990.

ANSI S12.53/Part 1-1999 (R 2004)/ISO 3743-1:1994 American National Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms (a Nationally Adopted International Standard). This standard, along with ANSI S12.53/Part 2-1999, replaces ANSI S12.33-1990.

ANSI S12.53/Part 2-1999 (R 2004)/ISO 3743-2:1994 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (a Nationally Adopted International Standard). This standard, along with ANSI S12.53/Part 1-1999 replaces ANSI S12.33-1990.

ANSI S12.54-1999 (R 2004)/ISO 3744:1994 American National Standard Acoustics -Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane (a Nationally Adopted International Standard). This standard replaces ANSI S12.34-1988.

ANSI S12.55-2006/ISO 3745:2003 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms (a Nationally Adopted International Standard). This standard replaces ANSI S12.35-1990.

ANSI S12.56-1999 (R 2004)/ISO 3746:1995 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane (a Nationally Adopted International Standard). This standard replaces ANSI S12.36-1990.

ANSI/ASA S12.57-2002 (R 2007)/ISO 3747:2000 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method in situ (a Nationally Adopted International Standard).

ANSI S12.60-2002 (R 2009) American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools.

ANSI S12.65-2006 American National Standard for Rating Noise with Respect to Speech Interference (Revision of ANSI S3.14-1977).

ANSI/ASA S12.67-2008 American National Standard Pre-Installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment.

ANSI/ASA S12.68-2007 American National Standard Methods of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn.

ASA Committee on Standards (ASACOS)

ASACOS (P. D. Schomer, Chair and ASA Standards Director)

U. S. Technical Advisory Groups for International Standards Committees

ISO/TC 43 Acoustics, **ISO/TC 43/SC 1** Noise (P. D. Schomer, U.S. TAG Chair)

ISO/TC 108 Mechanical Vibration, Shock and Condition Monitoring (D. J. Evans, U.S. TAG Chair)

ISO/TC 108/SC2 Measurement and Evaluation of Mechanical Vibration and Shock as Applied to Machines, Vehicles and Structures (A. F. Kilcullen, and R. F. Taddeo, U.S. TAG Co-Chairs)

ISO/TC 108/SC3 Use and Calibration of Vibration and Shock Measuring Instruments (D. J. Evans, U.S. TAG Chair)

ISO/TC 108/SC4 Human Exposure to Mechanical Vibration and Shock (D. D. Reynolds, U.S. TAG Chair)

ISO/TC 108/SC5 Condition Monitoring and Diagnostic Machines (D. J. Vendittis, U.S. TAG Chair; R. Taddeo, U.S. TAG Vice-Chair)

ISO/TC 108/SC6 Vibration and Shock Generating Systems (C. Peterson, U.S. TAG Chair)

IEC/TC 29 Electroacoustics (V. Nedzelnitsky, U.S. Technical Advisor)

STANDARDS NEWS FROM THE UNITED STATES

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

American National Standards Call for Comment on Proposals Listed

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

ASA (ASC S1) (Acoustical Society of America)

Revisions

BSR/ASA S1.18-200x, Method for Determining Ground Impedance (revision of ANSI S1.18-1999 (R2004))

Describes procedures for obtaining ground impedance from in-situ measurements of sound pressure spectra based on measurements of the magnitude and phase of the spectra of the difference in sound pressures measured by two vertically separated microphones using specified geometries. This standard extends and revises the template method in ANSI S1.18-1999 to enable the user to obtain impedance spectra that result entirely from measurements and are independent of any model for ground impedance. (May 4, 2009)

ASA (ASC S2) (Acoustical Society of America)

Reaffirmations

BSR/ASA S2.25-2004 (R200x), Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships (reaffirmation and redesignation of ANSI S2.25-2004)

Contains guidelines for limiting the hull and superstructure vibration of ships for the purposes of habitability and mechanical suitability. The mechanical suitability guidelines result in a suitable environment for installed equipment and preclude many major vibration problems, such as unbalance, misalignment, and other damage to the propulsion system. To obtain data to compare with the guidelines, this standard also specifies data acquisition and processing procedures. (May 11, 2009)

ASA (ASC S3) (Acoustical Society of America)

Revisions

BSR/ASA S3.2-200x, Method for Measuring the Intelligibility of Speech over Communication Systems (revision and redesignation of ANSI S3.2-1989 (R1999))

Includes measurement of speech intelligibility over entire communication systems, evaluation of the contributions of elements of speech communication systems, and evaluation of factors that affect the intelligibility of speech. Speech intelligibility over a communication system is measured by comparing the monosyllabic words trained listeners receive, and identify with the words trained talkers speak into a communication system that connects the talkers with the listeners. (May 11, 2009)

ASA (ASC S12) (Acoustical Society of America)

New Standards

BSR/ASA S12.60/Part 2-200x, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools—Part 2: Relocatable Classroom Factors (new standard)

Provides a relocatable-classroom-specific supplemental version of ANSI S12.60. Includes siting requirements, acoustical performance criteria and design requirements for relocatable classrooms. This standard seeks to provide design flexibility without compromising goal of obtaining adequate speech intelligibility for students and teachers in learning spaces within the standard's scope. (April 27, 2009)

BSR/ASA S12.64-200x, Quantities and Procedures for Description and Measurement of Underwater Sound from Ships—Part 1: General Guidelines (new standard)

Describes the measurement systems, procedures, and methodologies used for the beam aspect measurement of underwater sound pressure levels from ships at given operating conditions. Resulting quantities are nominal source level values. Does not require use of specific ocean location, but provides requirements for an ocean test site. Underwater SPL measurements are performed in the far-field & then corrected to a reference distance of 1 m. Applicable to all surface vessels manned or unmanned. (May 11, 2009)

Reaffirmations

BSR/ASA S12.18-1994 (R200x), Procedures for Outdoor Measurement of Sound Pressure Level (reaffirmation and redesignation of ANSI S12.18-1994 (R2004))

Describes two methods for the measurement of sound pressure levels (SPL) in the outdoor environment, considering the effects of the ground, the effects of refraction due to wind and temperature gradients and the effects due to turbulence. This standard focuses on measurement of SPL produced by specific sources outdoors. The measured SPL can be used to calculate SPL at other distances from the source or to extrapolate to other environmental conditions, or assess compliance with regulation.

ASTM (ASTM International)

Revisions

BSR/ASTM C769-200x, Test Method for Sonic Velocity in Manufactured Carbon and Graphite Materials for Use in Obtaining an Approximate Young S Modulus (revision of ANSI/ASTM C769-2005)

http://www.astm.org/DATABASE.CART/WORKITEMS/WK17523.htm (June 22, 2009)

IEEE (Institute of Electrical and Electronics Engineers)

New Standards

BSR C63.2-200x, Electromagnetic Noise and Field Strength Instrumentation, 10 Hz–40 GHz Specifications (new standard)

Frequency range is 10 Hz to 40 GHz. C63.2 now harmonizes the parameters of the quasi-peak detector with the requirements of CISPR 16-1. An optional discharge time constant is also specified. (June 8, 2009)

BSR C63.10-200x, Standard for Testing Unlicensed Wireless Devices (new standard)

Specifies methods, instrumentation, and facilities requirements for measurement of radio-frequency (RF) signals and RF noise emitted from unlicensed wireless devices. (June 8, 2009)

Reaffirmations

BSR/IEEE 488.1-2003 (R200x), Standard for Higher Performance Protocol for the Standard Digital Interface for Programmable Instrumentation (reaffirmation of ANSI/IEEE 488.1-2003)

Applies to interface systems used to interconnect both programmable and nonprogrammable electronic measuring apparatus with other apparatus and accessories necessary to assemble instrumentation systems. The basic functional specifications of this standard may be used in digital interface applications that require longer distances, more devices, increased noise immunity, or combinations of these. (June 9, 2009)

Revisions

BSR C63.4-200x, Methods of Measurement of Radio-Noise Emissions from Low-Voltage Electrical and Electronic Equipment in the Range of 9 kHz to 40 GHz (revision of ANSI C63.4-2003)

Specifies U.S. consensus standard methods, instrumentation, and facilities for measurement of radio-frequency (RF) signals and noise emitted from electrical and electronic devices in the frequency range 9 kHz to 40 GHz. It does not include generic nor product-specific emission limits. Where possible, the specifications in this standard are harmonized with other national and international standards used for similar purposes. (June 8, 2009)

SCTE (Society of Cable Telecommunications Engineers)

New Standards

BSR/SCTE 158-200x, Recommended Environmental Condition Ranges for Broadband Communications Equipment (new standard)

Specifies the recommended environmental conditions (temperature, humidity, altitude, and vibration) for the operation, storage and shipment of broadband communications equipment. (May 11, 2009)

TIA (Telecommunication Industry Association)

Revisions

BSR/TIA 470.120-C-200x, Telecommunications—Telephone Terminal Equipment—Transmission Requirements for Analog Speakerphones (revision and redesignation of ANSI/TIA 470-B-2006)

Provides speakerphone acoustic performance requirements for Customer Premises Equipment (CPE) intended for analog connection to the Public Switched Telephone Network (PSTN). These requirements should ensure compatibility and satisfactory performance to the user in a high percentage of installations. Test measurement methods reference procedures in IEEE Std 1329 where applicable.

BSR/TIA 1083-A-200x, Telephone Terminal Equipment—Handset—Magnetic Measurement Procedures and Performance Requirements (revision and redesignation of ANSI/TIA 1083-2007)

Defines measurement procedures and performance requirements for the handset-generated audio band magnetic noise of wireline telephones. A telephone complies with this standard if it meets the requirements in this standard when manufactured and can be expected to continue to meet these requirements when properly used and maintained. (June 22, 2009)

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards.

The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

ABYC (American Boat and Yacht Council)

BSR/ABYC A-23-200x, Sound Signal Appliances (new standard)

Provides a guide for the design, construction, performance, and installation of sound signal appliances for vessels operating in international waters and vessels operating in inland waters. Project Need: To identify safety issues with sound signal appliances. Stakeholders: Boat manufacturers, insurance personnel, surveyors, trade organizations, and consumers.

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.35-200x, Methods of Measurement of Performance Characteristics of Hearing Aids under Simulated in situ Working Conditions (revision and redesignation of ANSI S3.35-2004)

Describes methods to measure the acoustical effects of a simulated median adult wearer on the performance of a hearing aid using: direct simulated real-ear aided measurements (sound pressure developed by a hearing aid in an ear simulator for a given free-field input sound pressure), and insertion measurements (the difference between the sound pressures developed in the ear simulator with and without a hearing aid in place). These test methods are not intended for quality control. Project Need: To add definitions for first order- and second order-directional microphone systems in Annex B. Stakeholders: Hearing aid manufacturers, hearing aid dispensers.

Final actions on American National Standards

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

ASA (ASC S2) (Acoustical Society of America)

Revisions

ANSI/ASA S2.28-2009, Guide for the Measurement and Evaluation of Vibration of Shipboard Machinery (revision and redesignation of ANSI S2.28-2003)

ASA (ASC S3) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S3.21-2004 (R 2009), Methods for Manual Pure-Tone Threshold Audiometry (reaffirmation and redesignation of ANSI S3.21-2004)

ASA (ASC S12) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S12.60-2002 (R 2009) American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools

IEEE (Institute of Electrical and Electronics Engineers)

New Standards

ANSI/IEEE 1652-2008, Standard for the Application of Free Field Acoustic Reference to Telephony Measurements (new standard)

MEETING NOTICE

Technical Committee on Sound (TCoS)

Sponsor: Technical Committee on Sound (TCoS)

Purpose: **AHRI Standard 370 revision (Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment)**

Date: July 9, 2009

Time: 10:00 a.m. EDT

Location of Meeting: Web Meeting

Contact: Michael Woodford, 703-600-0344; E-mail: mwoodford@ahrinet.org

STANDARDS NEWS FROM ABROAD

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

Newly Published ISO and IEC Standards

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

ISO Standards

ACOUSTICS (TC 43)

ISO 362-1/Cor1:2009, Measurement of noise emitted by accelerating road vehicles—Engineering method—Part 1: M and N categories—Corrigendum

ISO 9612:2009, Acoustics—Determination of occupational noise exposure—Engineering method

ISO 3382-2/Cor1:2009, Acoustics—Measurement of room acoustic parameters—Part 2: Reverberation time in ordinary rooms—Corrigendum

ISO 10843/Cor1:2009, Acoustics—Methods for the description and physical measurement of single impulses or series of impulses—Corrigendum

ISO 13473-5:2009, Characterization of pavement texture by use of surface profiles—Part 5: Determination of megatexture

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 22266-1:2009, Mechanical vibration—Torsional vibration of rotating machinery—Part 1: Land-based steam and gas turbine generator sets in excess of 50 MW

TEXTILE MACHINERY AND ALLIED MACHINERY AND ACCESSORIES (TC 72)

ISO 9902-1/Amd1:2009, Textile machinery—Noise test code—Part 1: Common requirements—Amendment 1

ISO 9902-2/Amd1:2009, Textile machinery—Noise test code—Part 2: Spinning preparatory and spinning machinery—Amendment 1

ISO 9902-3/Amd1:2009, Textile machinery—Noise test code—Part 3: Nonwoven machinery—Amendment 1

ISO 9902-4/Amd1:2009, Textile machinery—Noise test code—Part 4: Yarn processing, cordage and rope manufacturing machinery—Amendment 1

ISO 9902-5/Amd1:2009, Textile machinery—Noise test code—Part 5: Weaving and knitting preparatory machinery—Amendment 1

ISO 9902-6/Amd1:2009, Textile machinery—Noise test code—Part 6: Fabric manufacturing machinery—Amendment 1

ISO 9902-7/Amd1:2009, Textile machinery—Noise test code—Part 7: Dyeing and finishing machinery—Amendment 1

ISO Technical Specifications

ACOUSTICS (TC 43)

ISO/TS7849-1:2009, Acoustics—Determination of airborne sound power levels emitted by machinery using vibration measurement—Part 1: Survey method using a fixed radiation factor

ISO/TS7849-2:2009, Acoustics—Determination of airborne sound power levels emitted by machinery using vibration measurement—Part 2: Engineering method including determination of the adequate radiation factor

IEC Standards

ELECTROACOUSTICS (TC 29)

IEC 60645-6 Ed. 1.0 b:2009, Electroacoustics—Audiometric equipment—Part 6: Instruments for the measurement of otoacoustic emissions

IEC 60645-7 Ed. 1.0 b:2009, Electroacoustics—Audiometric equipment—Part 7: Instruments for the measurement of auditory brainstem responses

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*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,457,429

43.38.Dv LAMINATED MOTOR STRUCTURE FOR ELECTROMAGNETIC TRANSDUCER

Enrique M. Stiles, assignor to STEP Technologies Incorporated
25 November 2008 (Class 381/414); filed 31 December 2003

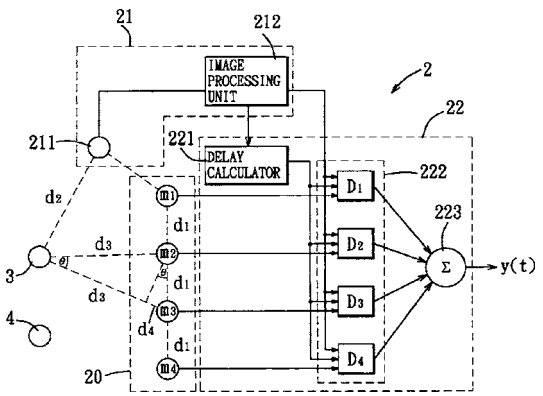
This patent does get around to the device described in the title, the essence of which borrows from ac transformer and motor design, but only after most of the prose and many of the figures go on about various and sundry yoke, pole piece, top plate, and enclosure variations, the purpose of which is to keep the voice coil cool, and in which the motor structures (probably) use the laminated motor structure mentioned in the title. The laminated motor structure 310 uses segments 312a-n and 316a-n to construct the top plate, pole piece, and back plate, in order to eliminate (or at least greatly reduce) the eddy currents that are induced in same by the changing voice coil magnetic field. The static magnetic field in gap 28 is still created by magnets 24.—NAS

7,313,243

43.38.Hz SOUND PICKUP METHOD AND SYSTEM WITH SOUND SOURCE TRACKING

Tien-Ming Hsu, assignor to Acer Incorporated
25 December 2007 (Class 381/92); filed in Taiwan 20 November 2003

Simply put, somehow the image processing unit 212 takes input from the camera 211 and calculates the separation distance of the microphones m1–m4. The delay calculator sets the delays (D1–D4) appropriately. Naturally, the conversion from image to delay time is taken for granted.—MK



7,471,804

43.38.Hz FLAT PANEL MONITOR FRAME WITH INTEGRAL SPEAKERS

Noel Lee, assignor to Monster Cable Products, Incorporated
30 December 2008 (Class 381/388); filed 6 January 2004

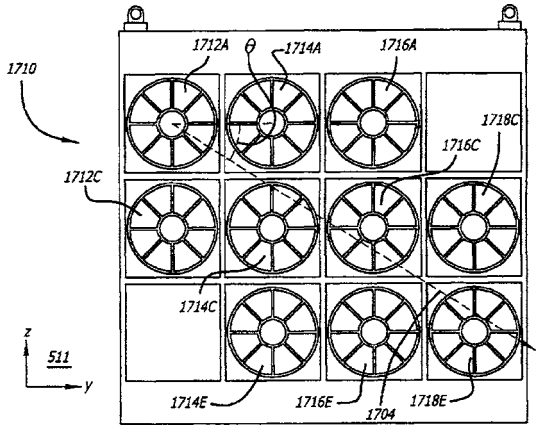
This is a design patent masquerading as a utility patent. It covers the placement of loudspeaker drivers in a frame without any justification.—MK

7,480,389

43.38.Hz SOUND DIRECTION SYSTEM

Bradley Joel Ricks and Andrew Rutkin, assignors to Harman International Industries, Incorporated
20 January 2009 (Class 381/335); filed 7 March 2002

This patent describes a steerable loudspeaker array. The illustration shows one side of a thin rectangular cabinet housing 20 loudspeakers; the remaining speakers are mounted on the opposite side to create a two-dimensional array of ten pairs. By delaying the signals to individual pairs, a



 0001-4966/2009/126(1)/514/11/$25.00

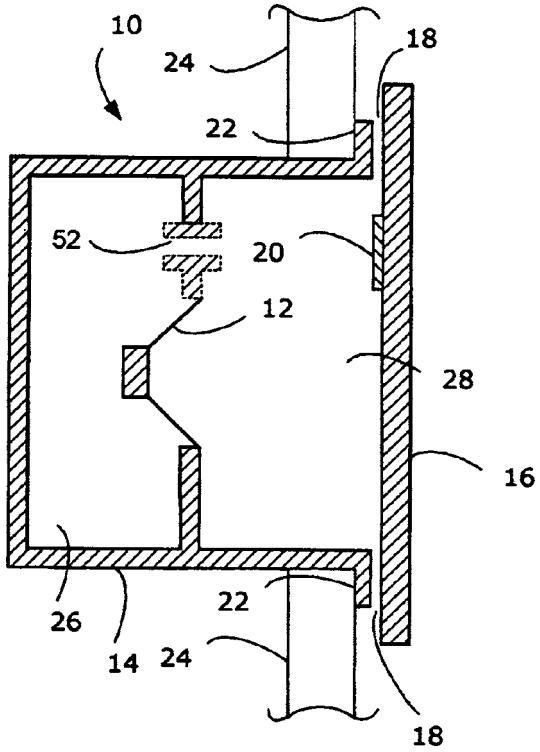
"fat" endfire array is created that can be steered vertically through a full 360 deg. (The dashed arrow indicates a steering angle of about 30 deg downward.) Instead of a symmetrical mushroom-shaped coverage pattern, coverage is typically wider horizontally than vertically—a desirable by-product. Other practical embodiments are described in the patent.—GLA

7,463,746

43.38.Ja NARROW OPENING ELECTROACOUSTICAL TRANSDUCING

Gerald F. Caron *et al.*, assignors to Bose Corporation
9 December 2008 (Class 381/345); filed 31 March 2003

Piezoelectric driver 20 is mounted to panel 16. Gap 18 allows the sound from cone type electrodynamic transducer 12 and port 52, as well as the sound from driver 20, to escape to a listening space (to the right of wall 24). Panel 18 can also be, and probably is, excited by the driver. Panel 18 is sized to obscure housing 14 and interior volume 20, and the baffle in which



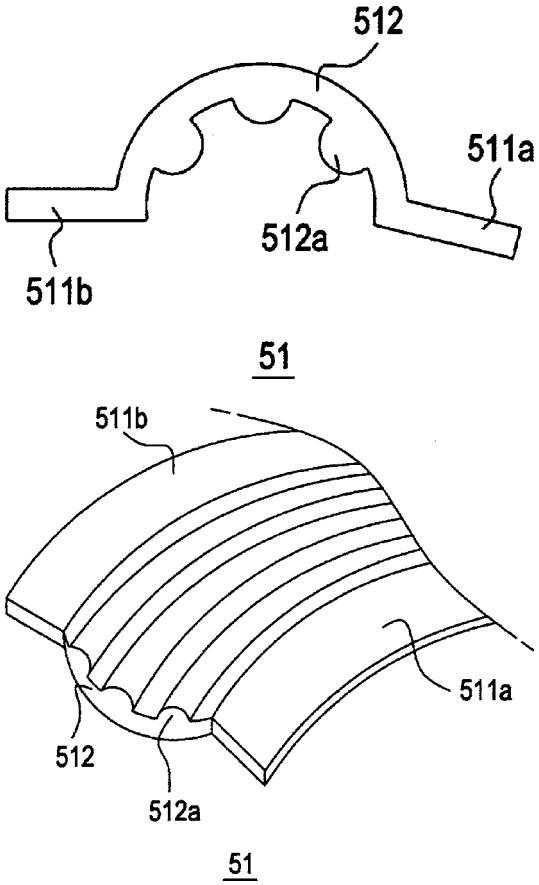
the cone speaker and port are mounted, and maybe for other reasons. Assembly 10 mounts in wallboard 24. The device allows for a two-way wall-mounted loudspeaker if rectangular wall features are acceptable. The combination of volume 26, port 52, driver 12, volume 20, and port (or rather, gap) 18 allows for extensive tuning of the system.—NAS

7,463,749

43.38.Ja DIAPHRAGM EDGE OF SPEAKER

Jong-pyo Lee, Ansan, Republic of Korea
9 December 2008 (Class 381/398); filed in Republic of Korea 8 March 2003

Ribs 512a are embossed in speaker surround 51, using the up-roll shown as well as down-roll, M-roll, W-roll, or any similar topology, to reduce the dip in frequency response known as compliance "suck-out" that



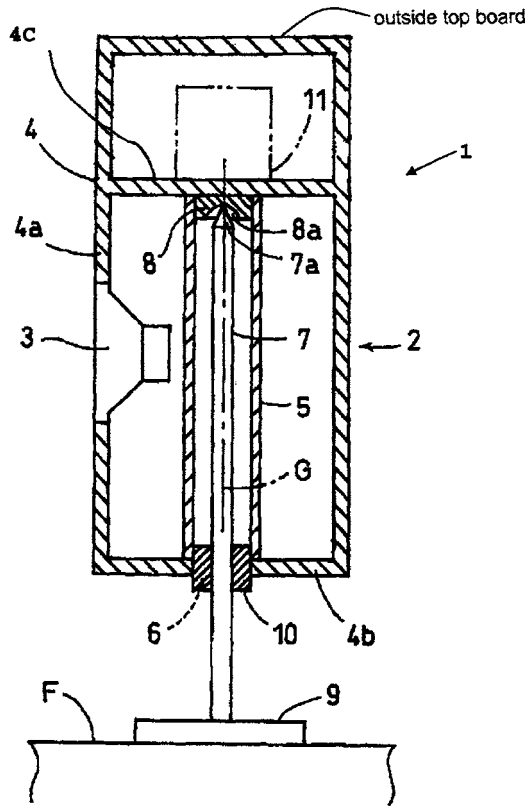
typically occurs around 1 kHz in mid-size cone type electrodynamic transducers. The preferred material for the surround appears to be compressed silicon rubber with a viscose rayon filler, that is "powdered" to have a length from 0.1 to 3.0 mm, and a rubber-to-filler weight ratio of 100:3.—NAS

7,478,703

43.38.Ja SPEAKER CABINET AND SPEAKER DEVICE

Takeshi Nakamura, assignor to Murata Manufacturing Company, Limited
20 January 2009 (Class 181/199); filed in Japan 9 May 2003

The number of really strange loudspeaker designs has declined in recent years, but they still appear from time to time. This loudspeaker cabinet is suspended from rod 7 that extends upward through tube 5. Thus, the entire



speaker system becomes a pendulum whose oscillations are restrained by damper 10. The arrangement is intended to minimize undulations of top board 4C and thereby provide a stable mounting surface for tweeter 11. —GLA

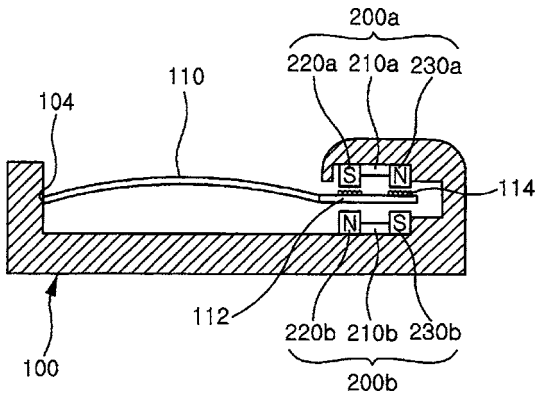
7,480,392

43.38.Ja PLATE TYPE SPEAKER USING HORIZONTAL VIBRATION VOICE COIL

Joung-Youl Shin, 414, Naeson-Dong, Uiwang 437-804 and Byung-Wan Han, 1478-4 Seocho3-Dong, Seocho-Gu, Seoul 137-888, both of Republic of Korea

20 January 2009 (Class 381/419); filed in Republic of Korea 5 December 2003

This design might be described as one stave of a barrel stave loudspeaker. Curved, flexible plate 10 is edge-driven by a flat voice coil 114. The goal is to create a very shallow loudspeaker for use in cellular phones and other restricted spaces.—GLA



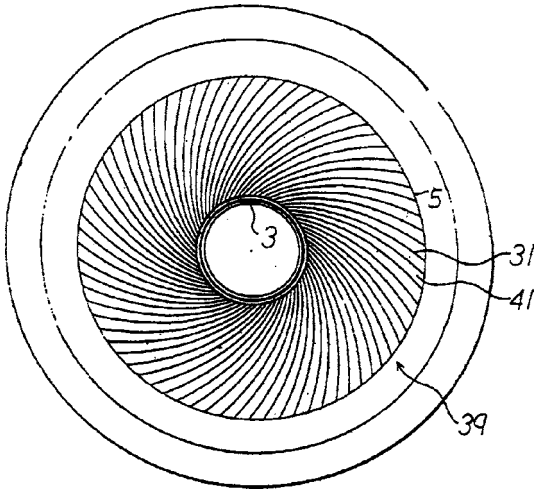
7,483,545

43.38.Ja ACOUSTIC DIAPHRAGM

Tadashi Nagaoka, Nishinomiya 668-8107, Japan

27 January 2009 (Class 381/423); filed 19 January 2005

To those involved with loudspeaker design the illustration will seem familiar. In this case, however, what appear to be folds or ridges are actually individual "elements." In a preferred two-layer embodiment the geometry is reversed in the second layer. According to the patent, "Thus, an acoustic diaphragm having the advantageous characteristics of a human eardrum and of a feather…" has finally been achieved.—GLA



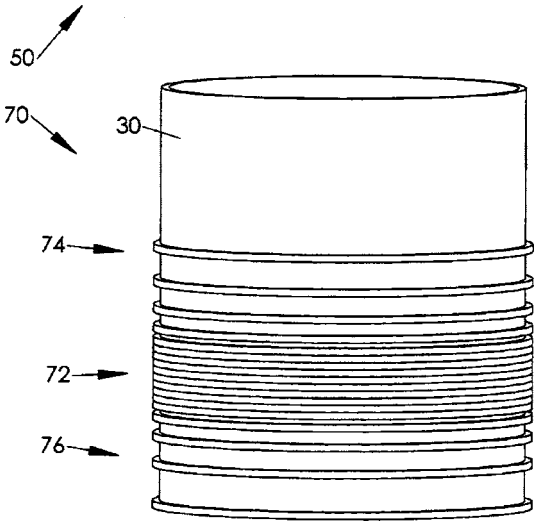
7,492,918

43.38.Ja AUDIO SPEAKER WITH GRADUATED VOICE COIL WINDINGS

Enrique M. Stiles and Richard C. Calderwood, assignor to Step Technologies Incorporated

17 February 2009 (Class 381/409); filed 17 August 2004

This patent document includes more than a dozen carefully prepared, beautifully drawn diagrams. The invention itself is less impressive. Maintaining linear travel over long cone excursions in a moving coil loudspeaker is difficult and involves numerous tradeoffs. The inventors argue that a



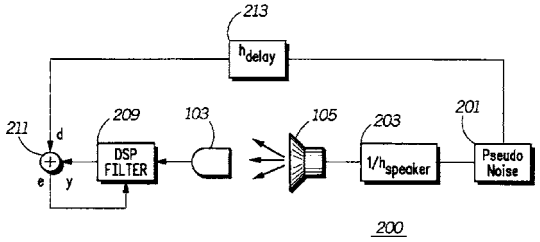
smooth, gradual non-linear characteristic is actually preferable. Such a characteristic can be achieved by graduating the spacing of voice coil turns as shown or by using wire of varying diameters. There is no indication that a working model has been built or tested.—GLA

7,492,909

43.38.Lc METHOD FOR ACOUSTIC TRANSDUCER CALIBRATION

Charles H. Carter, Jr., assignor to Motorola, Incorporated
17 February 2009 (Class 381/58); filed 5 April 2001

Using pseudo-random noise to equalize loudspeakers is a common procedure. Similarly, if one has a loudspeaker whose response is known, it can be used to calibrate a microphone. Might these techniques be used to equalize the internal microphone and loudspeaker of a cellular telephone? The answer is yes and the idea can even be patented.—GLA

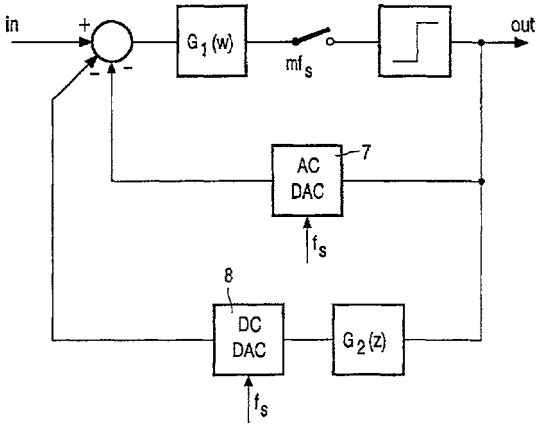


7,489,791

43.38.Md A/D CONVERTER WITH INTEGRATED BIASING FOR A MICROPHONE

Henricus J. Kunnen, legal representative and Eise Carel Dijkmans, assignors to Koninklijke Philips Electronics N.V.
10 February 2009 (Class 381/113); filed in the European Patent Office 5 July 2000

This nicely written patent addresses the issue of connecting a delta-sigma converter directly to an electret microphone. The required bias resistor introduces noise. However, by appropriate design of the dc feedback path (through filter G2z) and the digital-to-analog converter 8, this can be eliminated.—MK



7,492,293

43.38.Md VARIABLE RATE ANALOG-TO-DIGITAL CONVERTER

Robert Townsend Short *et al.*, assignors to Olympus Communication Technology of America, Incorporated
17 February 2009 (Class 341/123); filed 28 March 2007

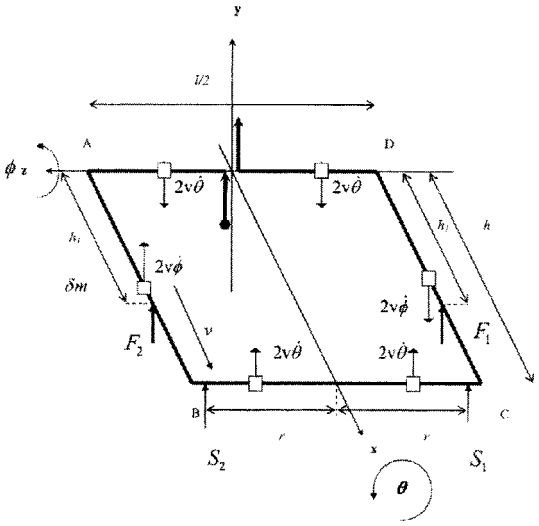
Something is terribly wrong when an analog-to-digital converter patent has no citations prior to 1997. In a low power system, clock rate is an issue. Therefore, if one can vary the clock rate, then one can conserve power. The inventors propose a number of very standard methods for changing the sampling rate.—MK

7,480,576

43.38.Pf COMPENSATING FOR FREQUENCY CHANGE IN FLOWMETERS

Michael S. Tombs, assignor to Invensys Systems, Incorporated
20 January 2009 (Class 702/45); filed 13 February 2007

This patent discloses the theory of operation of the Coriolis flowmeter, a device often used to measure the flow rate of liquids in industrial processes. The idealized configuration of flow is shown in the figure here, with liquid inlet and outlet axes closely aligned. A torque is applied along two orthogonal axes (here the x and z axes) using shakers located at F1 and F2. By running these out of phase, a combination of motions about the x and z axes will be imparted to fluid flowing through the square loop of pipe in the direction of ABCD. The velocity of motion at sensors S1 and S2 is measured, and the phase of relative motion as well. It is shown in the analysis



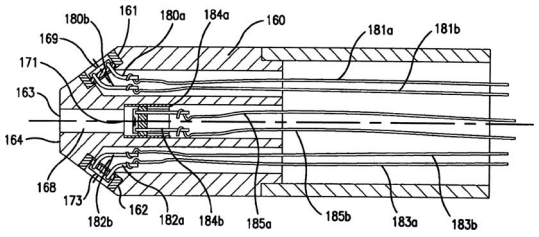
(which is in the patent) that the phase shift between S1 and S2 is a measure of the mass flow rate. The authors go further to say that by conducting the measurement at two separate frequencies, the density of the fluid can be determined unambiguously. Thus, both the density and the velocity of the fluid can be determined by measurements at the exterior of the pipe. The discussion of the phenomenon is thorough but sometimes confusing in its notation. It is claimed that the ability to separate density and velocity components of the mass flow rate is novel, and so is the technique of two-frequency excitation for accomplishing this.—JAH

7,484,418

43.38.Pf ULTRA MINIATURE MULTI-HOLE PROBES HAVING HIGH FREQUENCY RESPONSE

Anthony D. Kurtz, assignor to Kulite Semiconductor Products, Incorporated
3 February 2009 (Class 73/754); filed 6 November 2007

The authors disclose a pressure probe (it looks like a Pitot tube, but is not called that) that is "very small." Dimensions are not given, but it is claimed to be "less than 100 mils" (2.5 mm) in diameter. The overall design



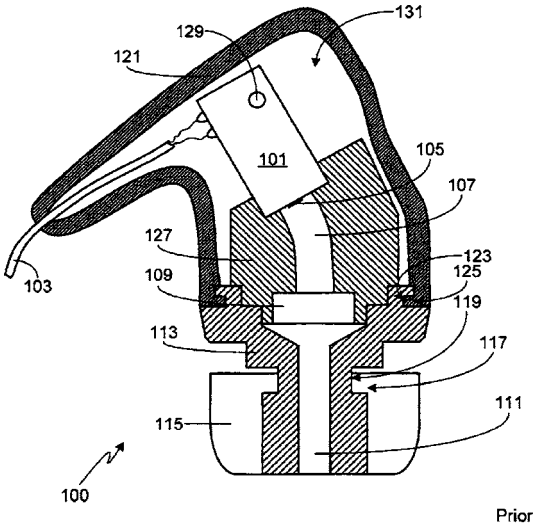
is driven by a desire to make the silicone micro-electronic mechanical pressure sensors inside be capable of operating at high temperatures ($>500\,^{\circ}\mathrm{F}$) and to be able to derive not only stagnation pressure from the sensor outputs but flow angle and velocity. The devices will apparently be commercially available soon, if not now.—JAH

7,489,794

43.38.Si EARPIECE WITH ACOUSTIC VENT FOR DRIVER RESPONSE OPTIMIZATION

Jerry J. Harvey, assignor to Ultimate Ears, LLC
10 February 2009 (Class 381/380); filed 17 July 2006

A conventional insertable earphone houses a driving transducer 101 whose forward output is conducted through channels 107 and 111 to the wearer's inner ear. The transducer's rear chamber is vented through port 129 to sealed chamber 131. The patent explains that although the volume of the chamber can be chosen to equalize low frequency response, optimum chamber volume is difficult to control in production. Instead, a much smaller port is used to provide a resistive rear load that is largely independent of chamber volume. What is set forth in the patent claims, however, is not the idea of a resistive load but rather a trial-and-error method of finding the optimum port diameter for a particular earpiece design.—GLA



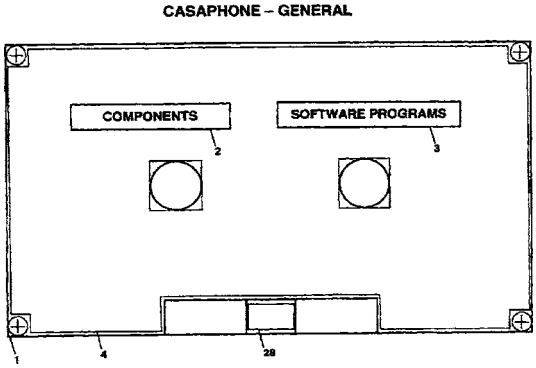
Prior Art

7,489,946

43.38.Si CELLULAR PHONE IN FORM FACTOR OF A CONVENTIONAL AUDIO CASSETTE

Sudharshan Srinivasan *et al.*, Fremont, California
10 February 2009 (Class 455/556.1); filed 12 April 2006

A great idea that just missed the boat: Why not make a hands-free cellular phone that plugs into the slot of your tape cassette player?—GLA

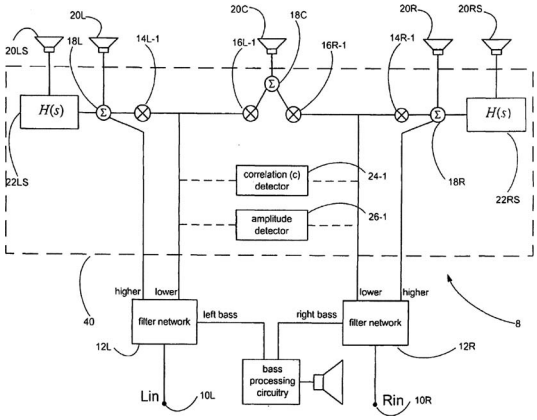


7,490,044

43.38.Vk AUDIO SIGNAL PROCESSING

Abhijit Kulkarni, assignor to Bose Corporation
10 February 2009 (Class 704/501); filed 8 June 2004

Numerous patents have been issued for circuits that generate synthetic surround sound signals from conventional two-channel program material. This patent teaches that such prior art circuits can produce unnatural effects, especially if the original stereo program has been compressed by algorithms such as MP3. An improved method is disclosed that includes spectral filtering and steering circuitry.—GLA

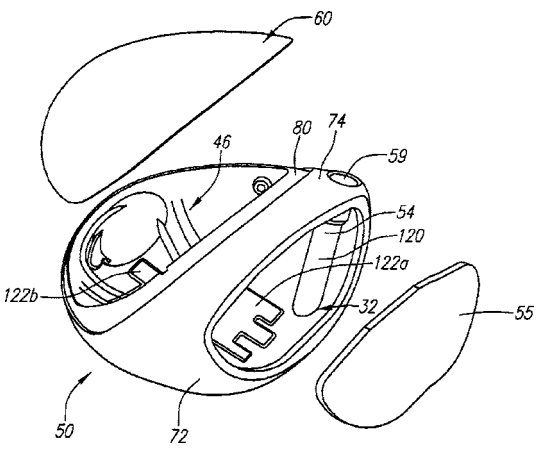


7,452,287

43.40.At MULTIPLE MATERIAL GOLF CLUB HEAD

Matthew J. Erickson *et al.*, assignors to Callaway Golf Company
18 November 2008 (Class 473/345); filed 30 October 2007

"Although the prior art has disclosed many variations of multiple material club heads, the prior art has failed to provide a multiple material, large volume club head with sufficient volume and an appealing sound during impact with a golf ball." Stiffening members on the sole section of major body 50 "enhance the tone of the golf ball head during impact with a golf ball" and "an epoxy based composition is positioned at a face-sole junction in order to reduce the amplitude of the sound generated by the golf club during impact with a golf ball." A good read about some prior art and some of the official rules concerning certain aspects of a driver golf club head. A detailed description of proprietary materials with certain physical parameters, which appears to be a requirement of this class of patents, is included.—NAS

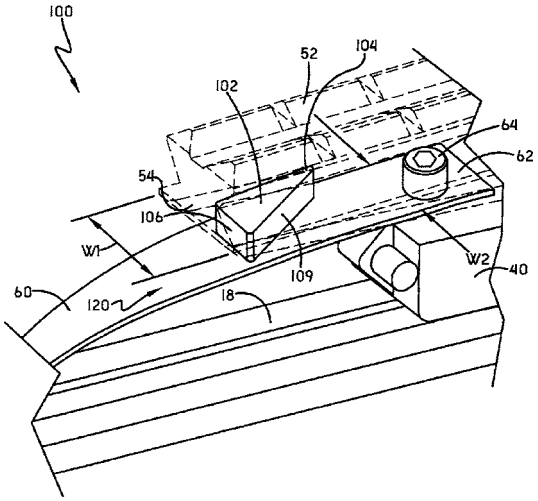


7,455,059

43.40.Tm VIBRATION DAMPENING ARROW RETENTION SPRING

Michael Jay Shaffer, Mogadore, Ohio
25 November 2008 (Class 124/25); filed 29 December 2004

The sound, which can be characterized as a "ping," that emanates from arrow retention spring 60 as an arrow is fired from a crossbow is reduced by vibration dampener 102. The dampener is made from a highly pliable compound such as flexible polyurethane, which is placed between scope mount 54 and the spring.—NAS

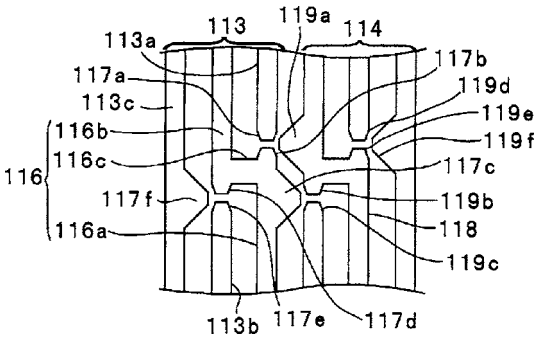


7,479,855

43.58.Kr LONGITUDINALLY-COUPLED-RESONATOR-TYPE ELASTIC WAVE FILTER DEVICE

Masaru Yata, assignor to Murata Manufacturing Company, Limited
20 January 2009 (Class 333/196); filed in Japan 22 September 2006

This patent discloses an improvement on existing construction methods for a surface acoustic wave (SAW) resonator filter device. The authors describe a new electrode configuration that allows them to reduce ringing at the edges of the pass band of a SAW filter by introducing a form of apodization or weighting of the electrodes. One of several ways in which they propose to accomplish this is illustrated in the figure. Essentially, metallized "crossover regions" are used to create a form of apodization that is easier to manufacture than traditional approaches of adding material in these areas. A few examples are given of alternate approaches to this, but that is about all there is.—JAH

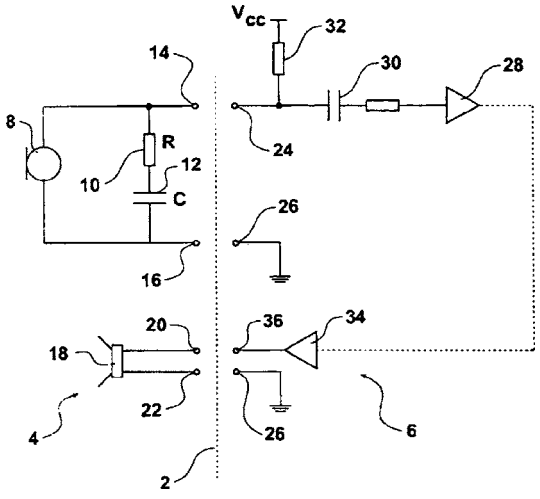


7,489,785

43.50.Ki NOISE CANCELLATION SYSTEM AND HEADPHONE THEREFOR

Mark Donaldson, Parnell, Auckland and Graeme Colin Fuller, Mt. Wellington, Auckland 1001, both of New Zealand
10 February 2009 (Class 381/71.6); filed in New Zealand 28 June 2002

Rather than include noise-canceling circuitry inside a headphone, this system separates the headphone and the associated electronics so that different headsets can be used. A filter is provided to configure the feedback signal for various types of headsets. The microphone output is also "normalized" to the requirements of noise cancellation, allowing the adjustable feedback filter to be a simple resistor-capacitor circuit.—GLA

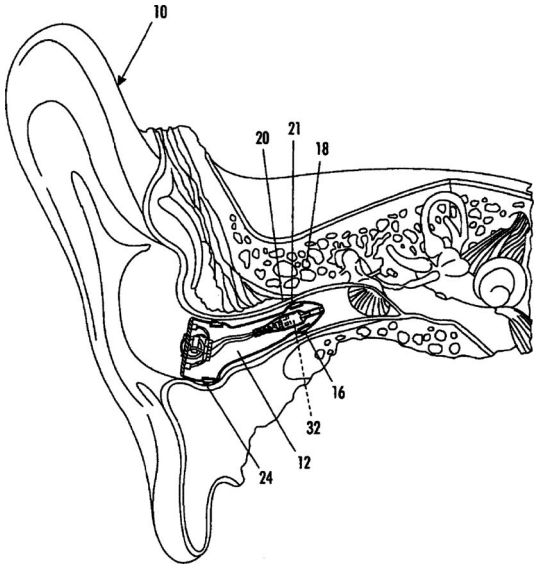


7,480,387

43.66.Ts IN THE EAR HEARING AID UTILIZING ANNULAR ACOUSTIC SEALS

John A. Meyer, Rochester and Dean Thomas Penman, Clarence, both of New York
20 January 2009 (Class 381/322); filed 10 August 2006

Two channels around the circumference of the shell of an in-the-canal hearing aid contain inset, non-circular soft rings that protrude beyond the shell perimeter to improve comfort and acoustic seal.—DAP

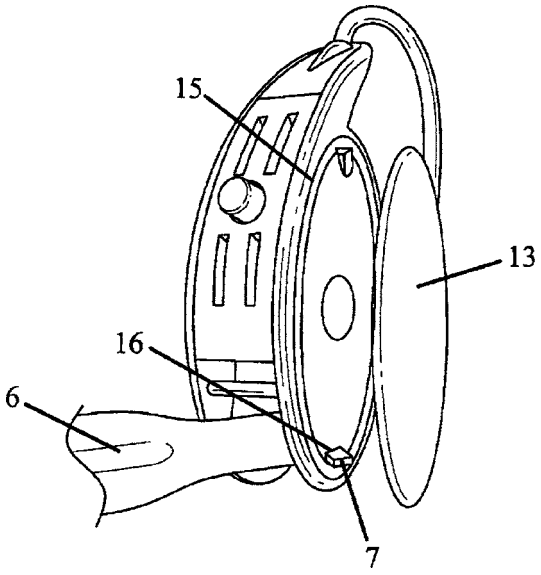


7,480,388

43.66.Ts BEHIND THE EAR HEARING AID PART WITH EXCHANGEABLE COVER

Ben Crook, assignor to Bernafon AG
20 January 2009 (Class 381/322); filed in the European Patent Office 7 August 2007

A detachable ornamental cover plate snaps onto a side wall of a behind-the-ear hearing aid. The coverplate is used to customize the exterior appearance of the hearing aid in order to match the color of the wearer's skin or clothing.—DAP

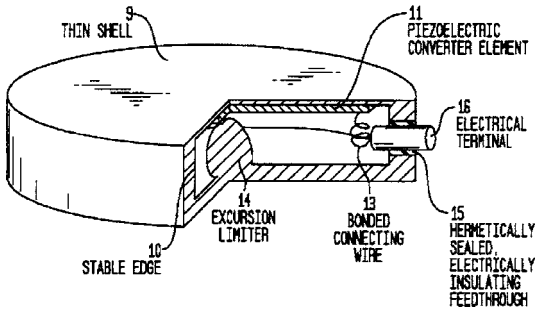


7,481,761

43.66.Ts IMPLANTABLE CONVERTER FOR COCHLEA IMPLANTS AND IMPLANTABLE HEARING AIDS

Matthias Blau *et al.*, assignors to Med-El Elektromedizinische Geräte Ges.m.b.H.
27 January 2009 (Class 600/25); filed in Germany 15 January 2003

An implantable, hermetically sealed piezoelectric converter senses vibrations from an ear ossicle and converts them to electric energy for use as an input to an implantable hearing device. An electronics module, also packaged with the converter in the thin-shell, biocompatible, hollow housing, further conditions the signal.—DAP

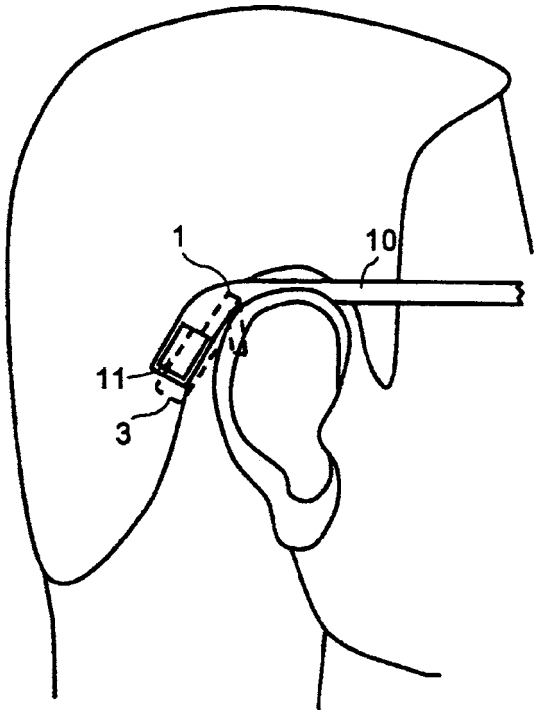


7,488,284

43.66.Ts IMPLANTABLE PROSTHESIS WITH DIRECT MECHANICAL STIMULATION OF THE INNER EAR

Jean-Noël Hanson *et al.*, assignors to MXM
10 February 2009 (Class 600/25); filed in France 29 January 2004

A two-piece implantable hearing device may comprise microphone(s), signal processor, battery, vibrator, and retaining magnet, all packaged in an external housing, which couples vibratory energy through the skin to an implanted plate. The plate, which may have a magnet to hold the external device in place, is connected to a pivotably-mounted rod that is also in contact with a semicircular canal of the patient's inner ear. Thus, externally produced vibrations are coupled to the inner ear via the rod.—DAP



7,489,789

43.66.Ts METHOD FOR NOISE REDUCTION IN AN AUDIO DEVICE AND HEARING AID WITH MEANS FOR REDUCING NOISE

Thomas Kaulberg, assignor to Oticon A/S
10 February 2009 (Class 381/94.3); filed in Denmark 2 March 2004

After a first detector identifies whether speech is present in the input signal, a second detector assesses the temporal modulation depth in multiple frequency bands by measuring peak and noise floor levels. One of two possible amounts of gain attenuation is assigned to each frequency band. Transitions occur by fading between a high gain attenuation (when speech is not detected) and a low gain attenuation (when speech is detected).—DAP

7,492,916

43.66.Ts METHOD FOR MANUFACTURING ACOUSTICAL DEVICES AND FOR REDUCING ESPECIALLY WIND DISTURBANCES

Hans-Ueli Roeck *et al.*, assignors to Phonak AG
17 February 2009 (Class 381/312); filed 2 October 2006

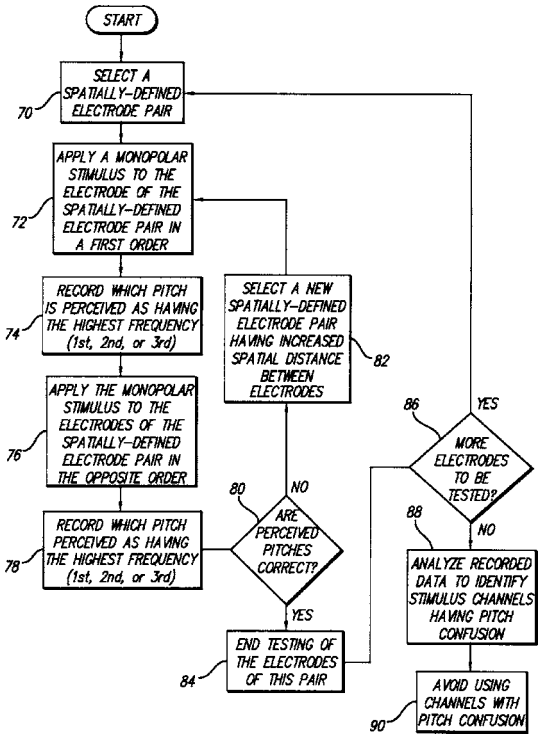
An adjustable high pass frequency response characteristic is applied to the hearing device output when a statistical analysis of the low pass filtered input concludes that wind energy may be present.—DAP

7,493,170

43.66.Ts ADAPTIVE PLACE-PITCH RANKING PROCEDURE FOR OPTIMIZING PERFORMANCE OF A MULTI-CHANNEL NEURAL STIMULATOR

Philip A. Segel and Tracey L. Kruger, assignors to Advanced Bionics, LLC
17 February 2009 (Class 607/57); filed 1 September 2006

Fine-tuning of cochlear implant performance is accomplished by wearers providing pitch ranking judgments (higher, lower, or the same) for sequential presentations of monopolar stimulation pulses applied to selected, variably-spaced, target and competing channel electrode pairs. If pitch judgment is correct, no further testing is done for that electrode pair. If judgment errors occur, place errors are indicated, and a spread of confusion search is conducted by separating the target and competing channels by one electrode contact at a time until channels are utilized that produce no further pitch judgment errors.—DAP



7,487,086

43.72.Ar TRANSCRIPT ALIGNMENT

Kenneth King Griggs, assignor to Nexidia Incorporated
3 February 2009 (Class 704/235); filed 5 June 2007

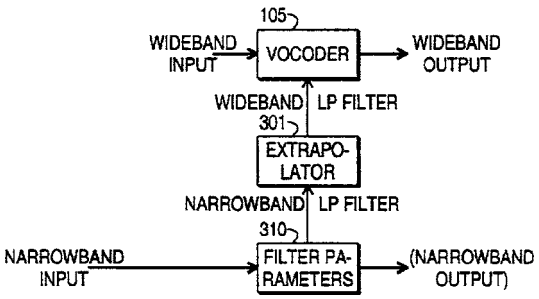
This patent describes a system for jointly processing an audio recording and a corresponding transcription text and determining marks in both files which represent corresponding points in time. That is, the text will be time aligned to match the audio recording. In particular, the system is said to involve methods to deal with transcription errors and audio noise and to be able to perform the task without manual intervention. Two strategies are described. In one, text sequences are extracted which appear to have possible corresponding distinctive audio events. The audio data are then searched for possible candidate regions matching the distinctive events. In another strategy, possible search terms are entered by the user, apparently both as text and speech. In either strategy, the time points resulting from both text and speech candidate regions in the data files are collected, and a scoring algorithm computes the most likely match between text and audio time marks.—DLR

7,483,830

43.72.Gy SPEECH DECODER AND A METHOD FOR DECODING SPEECH

Jani Rotola-Pukkila *et al.*, assignors to Nokia Corporation
27 January 2009 (Class 704/219); filed in Finland 7 March 2000

A narrowband encoded speech input signal, received from a digital telephone system, is converted using two linear prediction (LP) filters and a vocoder into a wideband decoded sample stream with a high sampling rate. Information is extracted about regularities in or differences between frequency domain coefficients of a first LP filter associated with the narrow input frequency band. Coefficients for a second LP filter, which are derived by extrapolating from the regularities of the first LP filter coefficients, are used by a vocoder to convert the narrowband input signal into an output signal with a wider frequency band.—DAP

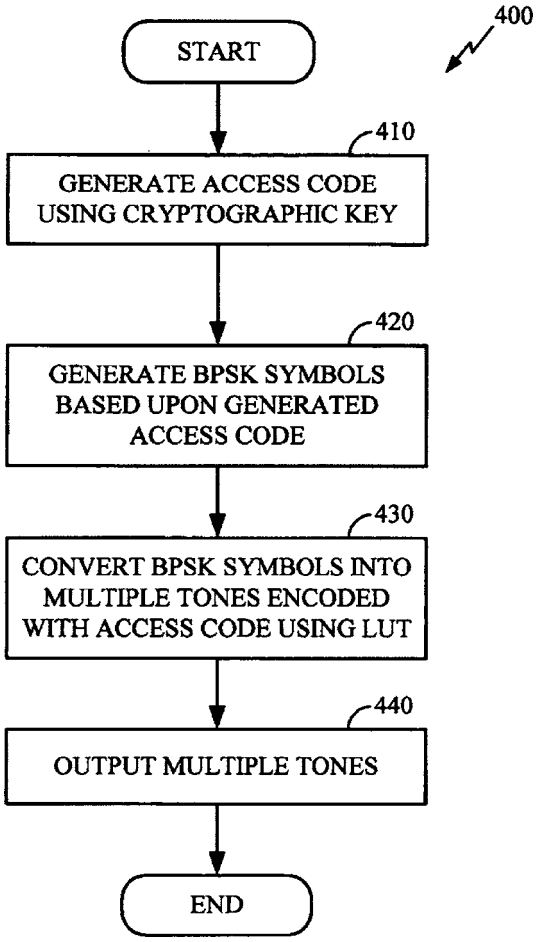


7,487,362

43.72.Gy DIGITAL AUTHENTICATION OVER ACOUSTIC CHANNEL

Jack Steenstra *et al.*, assignors to Qualcomm, Incorporated
3 February 2009 (Class 713/186); filed 23 February 2004

An apparatus for generating an access code which provides improved security in a data processing system includes memory for storing and retrieving a cryptographic key, and a processor for generating an access code using the cryptographic key. A converter module generates multiple parallel binary phase shift keyed (BPSK) symbols that are encoded with the access code. Using a look up table, a second processor in the converter module converts the BPSK symbols to multiple tones that are output as audio.—DAP

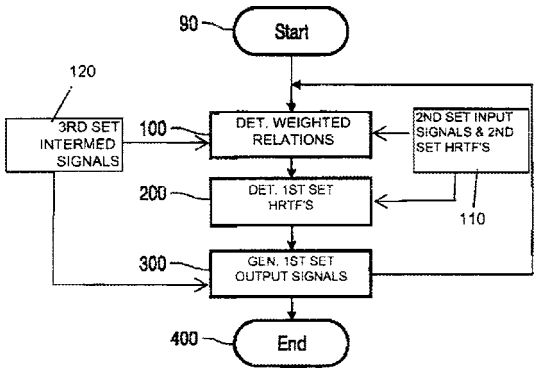


7,489,792

43.72.Gy GENERATION OF A SOUND SIGNAL

Ronaldus Maria Aarts *et al.*, assignors to Koninklijke Philips Electronics N.V.
10 February 2009 (Class 381/309); filed in the European Patent Office 23 September 2002

This patent describes a method to reduce the number of head related transfer functions (HRTFs) and processing complexity required for generating a stereo sound reproduction signal for headphones. For each input signal, a weighted relation is determined comprising intermediate sound signals and at least one weight. A new first set of HRTFs is then determined based on the input sound signals, a second set of HRTFs and the weighted relation. The intermediate sound signals are transferred with the first set of HRTFs to produce output signals.—DAP



7,328,154

43.72.Ne BUBBLE SPLITTING FOR COMPACT ACOUSTIC MODELING

Ambroise Mutel *et al.*, assignors to Matsushita Electrical Industrial Company, Limited
5 February 2008 (Class 704/245); filed 13 August 2003

This patent treats the usual problems with speaker-independent large-vocabulary speech recognition systems. The acoustic models developed from numerous speakers do not perform well in recognition because the high variance of each acoustic unit model, caused by interspeaker variability, induces too much overlap. A method is disclosed for creating more compact acoustic "bubble" models, which involves splitting the training set into more homogeneous speaker groups and then training a bubble model for each group. The speech criterion preferred for splitting is vocal tract length, and it looks like typical approaches to vocal tract length normalization are leveraged for the purpose here.—SAF

7,328,159

43.72.Ne INTERACTIVE SPEECH RECOGNITION APPARATUS AND METHOD WITH CONDITIONED VOICE PROMPTS

Chienchung Chang and Narendranath Malayath, assignors to Qualcomm Incorporated
5 February 2008 (Class 704/275); filed 15 January 2002

This patent is chiefly concerned with voice barge-in from a user operating an interactive voice recognition system that speaks prompts. If the system is to recognize the user's speech while a prompt is playing, a way must be found to separate the two. The method put forth here seems simple at first, but the details are not explained very well. The idea seems to involve selective filtering of the system prompts and the input speech after user speech is detected, to somehow place the user speech and system prompt speech into "conjugate frequency bands." It sounds like an ultra-cheap means of facilitating spectral subtraction.—SAF

7,428,491

43.72.Ne METHOD AND SYSTEM FOR OBTAINING PERSONAL ALIASES THROUGH VOICE RECOGNITION

Kuansan Wang *et al.*, assignors to Microsoft Corporation
23 September 2008 (Class 704/244); filed 10 December 2004

This patent treats the problem of "aliases" in speech recognition, by which term is meant a typical username for a person such as "safulop" for Sean Fulop. Aliases pose a problem for speech recognition because a typical recognition engine is not able to deal with a person saying "safulop" in the

typical fashion, which would probably involve saying the first two letters "ess ay" and then the pronounced surname. A number of methods for assisting speech recognition with aliases are disclosed, e.g., displaying a list of alias variations based on a recognition of a name from within the spoken alias. The user would then have to select the correct alias from the generated list. The patent offers this and other no less cumbersome solutions.—SAF

7,437,286

43.72.Ne VOICE BARGE-IN IN TELEPHONY SPEECH RECOGNITION

Xiaobo Pi and Ying Jia, assignors to Intel Corporation
14 October 2008 (Class 704/233); filed 27 December 2000

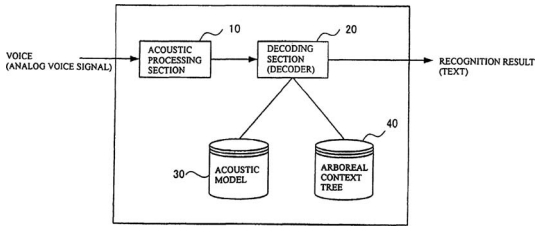
"Voice barge-in" refers to the capability of a telephone-based speech recognition system to accept user speech while the system is in the middle of a spoken prompt. The patent describes some basic methods of enabling barge-in which involve first detecting whether a user is speaking, and second, subtracting the spoken prompt echo spectrum from the user speech spectrum. There are a few more details presented, but it pretty much comes down to this simplified summary.—SAF

7,480,612

43.72.Ne WORD PREDICTING METHOD, VOICE RECOGNITION METHOD, AND VOICE RECOGNITION APPARATUS AND PROGRAM USING THE SAME METHODS

Shinsuke Mori *et al.*, assignors to International Business Machines Corporation
20 January 2009 (Class 704/9); filed in Japan 24 August 2001

A history tree for word prediction of partial parse trees is retrieved that covers words before a word to be predicted in a predetermined sentence. The shape of a partial parse tree is selected from the sentence structure of the history. A context tree for word prediction is retrieved and compared to the history tree, and nodes that match are identified. A word is predicted using a probability distribution appended to an identified node of the context tree.—DAP



7,487,083

43.72.Ne METHOD AND APPARATUS FOR DISCRIMINATING SPEECH FROM VOICE-BAND DATA IN A COMMUNICATION NETWORK

Peng Jie Zhang, assignor to Alcatel-Lucent USA Incorporated
3 February 2009 (Class 704/214); filed 13 July 2000

In order to reduce errors, periodicity characteristics, short-term autocorrelation coefficients, and power levels of input signal segments in a communication network are used to determine whether speech or high speed voice-band data (VBD), e.g., from modems or fax machines, are present. An assumption is made that changes from speech to VBD or vice versa during a communication event are unlikely. The result is used to select the type of signal processing to be performed.—DAP

7,487,095

43.72.Ne METHOD AND APPARATUS FOR MANAGING USER CONVERSATIONS

Jeffrey Hill and Yuri Zieman, assignors to Microsoft Corporation
3 February 2009 (Class 704/275); filed 2 September 2005

The subject of this patent is a natural language speech recognition system for use in a customer call center. The intent is to provide a fluid, natural conversation with a caller with the goal of learning the customer's problem and providing information in as much detail as necessary in order to resolve the customer's concerns with a minimum of annoyance and aggravation. Phrases spoken by the caller are recognized, and the words are used to consult a concept recognition process. If the system is thus able to construct an adequate information database so as to be able to proceed with the conversation, it will do so in the form of either voice or email. If not, all of the available information will be provided to a human operator so that the person can continue handling the call. The patent includes a moderate amount of detail on the operation of the concept recognition process.—DLR

7,453,040

43.75.Gh ACTIVE BRIDGE FOR STRINGED MUSICAL INSTRUMENTS

Stephen Gillette, Simi Valley, California
18 November 2008 (Class 84/723); filed 2 December 2005

By adding an active transducer underneath the pickup (as part of the bridge), an external input can sustain or damp a (metal) string vibration.—MK

7,482,518

43.75.Gh HIGH DENSITY SOUND ENHANCING COMPONENTS FOR STRINGED MUSICAL INSTRUMENTS

Robert DiSanto, assignor to Stone Tone Music, Incorporated
27 January 2009 (Class 84/291); filed 12 October 2005

As a higher density material for a solid body guitar, the inventor claims that using stone enhances, "focuses, retains and centralizes the instrument's core vibrations." Is this what Bob Dylan meant when he wrote that "everyone must get stoned"? (Rainy Day Women Nos. 12 and 35)—MK

7,476,794

43.75.Hi SOUND MODIFICATION SYSTEM

James H. May, Jr., assignor to Remo, Incorporated
13 January 2009 (Class 84/411 R); filed 12 September 2005

Remo, the well known drum manufacturer, proposes adding snaps to the top of the drum head. Then you can snap on dampers, springs, jangles, and whatnot. It is possible that these additions could add mass to the head but with enough force, anything is possible.—MK

7,485,791

43.75.Hi GOLDEN RATIO AIR VENT HOLES

Akito Takegawa, assignor to Pearl Musical Instrument Company
3 February 2009 (Class 84/411 R); filed 17 April 2007

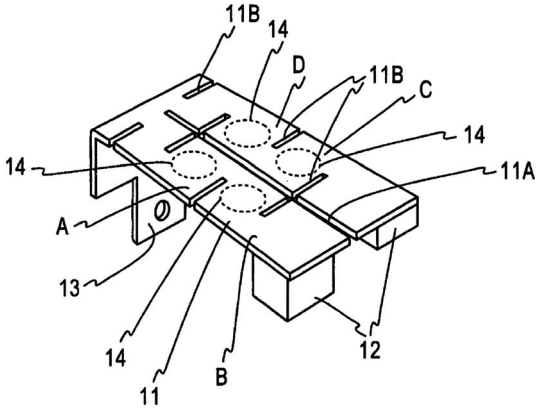
According to this patent, "the location defined by the Golden Ratio has been proven by the instant inventors to be the optimal location for the vent hole(s) to maximize the functional and tonal qualities of the drum." Golden ratio fans take note.—MK

7,488,887

43.75.Hi PERCUSSION-INSTRUMENT PICKUP AND ELECTRIC PERCUSSION INSTRUMENT

Yasuhiko Mori, assignor to Korg Incorporated
10 February 2009 (Class 84/743); filed in Japan 19 December 2005

There is not much new here except for the plate shown: here each cut in the bracket will produce a different response.—MK

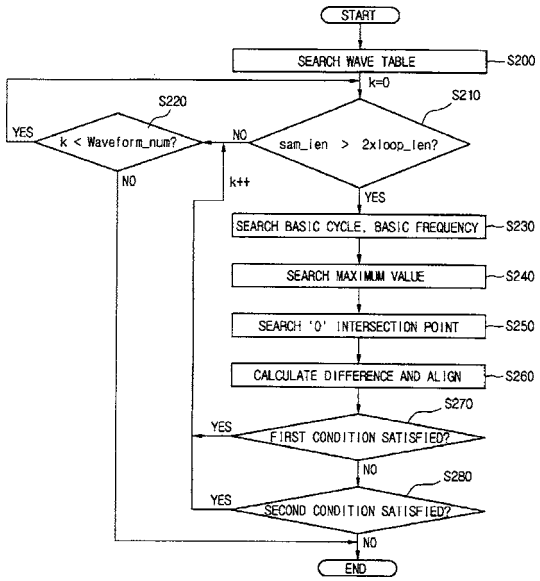


7,462,773

43.75.Wx METHOD OF SYNTHESIZING SOUND

Yong Chul Park *et al.*, assignor to LG Electronics Incorporated
9 December 2008 (Class 84/607); filed in Republic of Korea 15 December 2004

In a sampling synthesizer, the sustain tones are created by "looping" between "breakpoints" (the begin and end points of the sustain section in a waveform). These points are difficult to find. This proposal makes two assumptions: (a) There exists a maximum which is also the starting point of the waveform and (b) the waveform begins and ends at a zero crossing.—MK



7,465,868

43.75.Wx FRAMELESS MUSICAL KEYBOARD

Gerhard Lengeling, assignor to Apple Incorporated
16 December 2008 (Class 84/719); filed 8 June 2005

With another entry in the continuing saga of portable, separable keyboards, this inventor realizes that one octave looks like another on the piano keyboard. So, you could construct a multioctave instrument by adding pieces to the left and right of an instrument. Naturally they can be connected with a serial protocol like USB, too.—MK

7,491,879

43.75.Wx STORAGE MEDIUM HAVING MUSIC PLAYING PROGRAM STORED THEREIN AND MUSIC PLAYING APPARATUS THEREFOR

Mitsuhiro Hikino and Junya Osada, assignors to Nintendo Company Limited
17 February 2009 (Class 84/615); filed in Japan 25 April 2006

Behold a Wii patent: by using the acceleration vectors of the Wii baton, the performance parameters of the music playback can be altered. Specifically, the patent mentions tonality (major or minor keys), articulation (legato or staccato), and speed (beats per measure).—MK

7,481,770

43.80.Vj METHOD AND SYSTEM FOR TISSUE DIFFERENTIATION

Meir Botbol, assignor to DeepBreeze Limited
27 January 2009 (Class 600/481); filed 4 February 2004

Acoustic signals are obtained from locations on the surface of a body. Filtered versions of the signals are used to form images. Pixels from the images are divided into categories, to which are assigned a probability. Using the probabilities, a second image is formed.—RCW